FIN/ECON 250 Final Project

My project is to explore the cointegration relationship between United states HPI and California HPI, which are quarterly data from the first season in 1975 to the end of 2019, and further built a Vector Error Correction Model (VECM) to do a simple forecasting. Apart from that, I also compared the forecasting performance among VECM model, Var model, ARIMAX model and ARIMA model to see how well VECM can support me to finish the task.

First, I converted the two raw series to seasonality adjusted ones. Next, I drew them in the same plot to observe their patterns. As shown below, them appear non-linear and moving together.

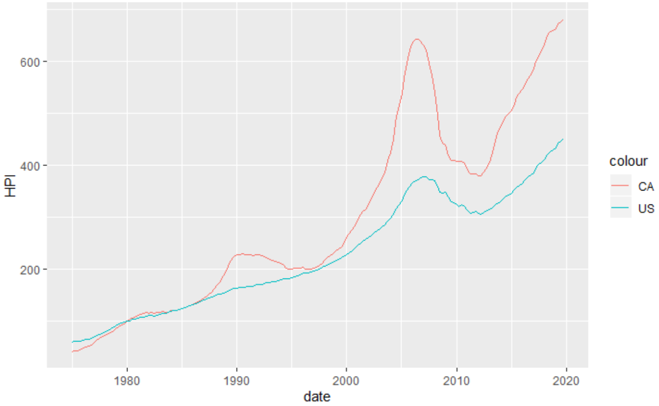

Figure 1: Raw sereis plots of US HPI and CA HPI

I guess them might be conintegrated. Therefore, I took the cointegration test on US HPI and CA HPI. The outcome shows that both these two series cannot reject the null hypothesis of DF test (type = “trend”, since two series have obvious trend pattern), but the residuals of lm(US HPI ~ CA HPI) passes the DF test (-4.4614 < -1.95). Consequently, I concluded that the two series are exactly cointegrated.

In the next step, I built the VECM model based on the cointegration. Since the two series seems exponential, I took Logarithms of both them and the series became more linear.

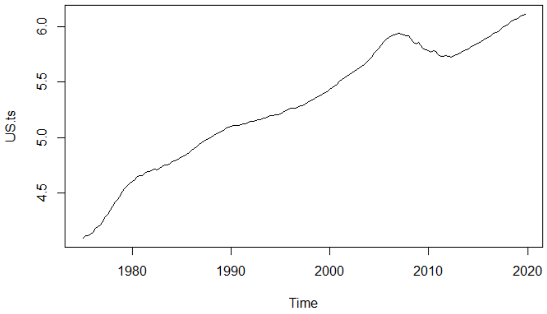

Figure 2: Logarithmic US HPI

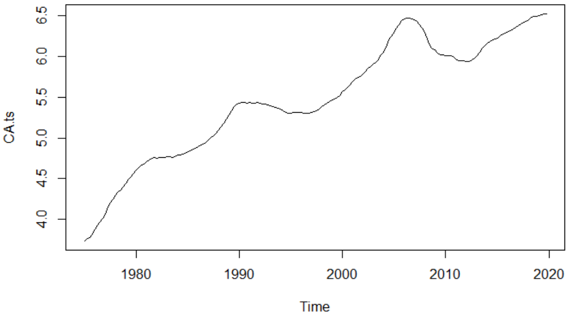

Figure 3: Logarithmic CA HPI

Due to the nature of HPI series, I differentiated these series twice. The two transformed series fluctuated around zero, which means they should be stationary. The DF test results also shows that these two differentiated series are exactly stationary.

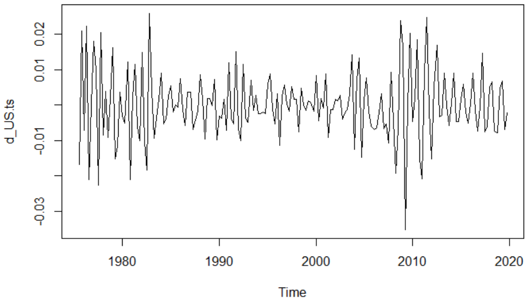

Figure 4: Twice differentiated and Logarithmic US HPI

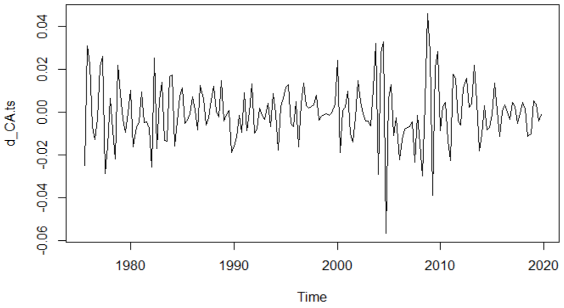

Figure 5: Twice differentiated and Logarithmic US HPI

Then, I plotted the cross-correlation of them, which indicates that there are spikes at lag -0.5 and 0.5. My data is quarterly. So, 0.5 lags mean 0.5 * 4(quarters) = 2 quarters. Therefore, I decided to include 2 lags when running VECM and Var.

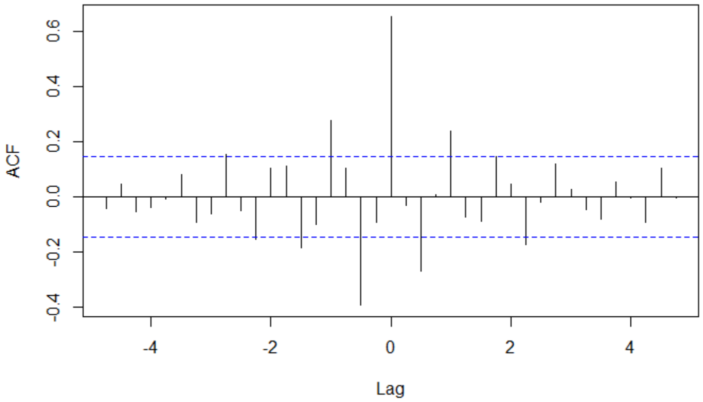

Figure 6: Cross-Correlation (covariance) plot

Finally, I ran the VECM, Var, ARIMA and ARIMAX; and calculated the RMSE of each model by transforming the predicted series with logarithmic and differentiated scale back to the series with original scale. VECM generated RMSE of 4.036, which is the model with lowest RMSE among VECM, Var, ARIMAX and ARIMA.

Model	RMSE
VECM	4.036
Var	5.064
ARIMA	4.309
ARIMAX	5.994

The next two graphs are the comparison of the performance of these models in validating period.

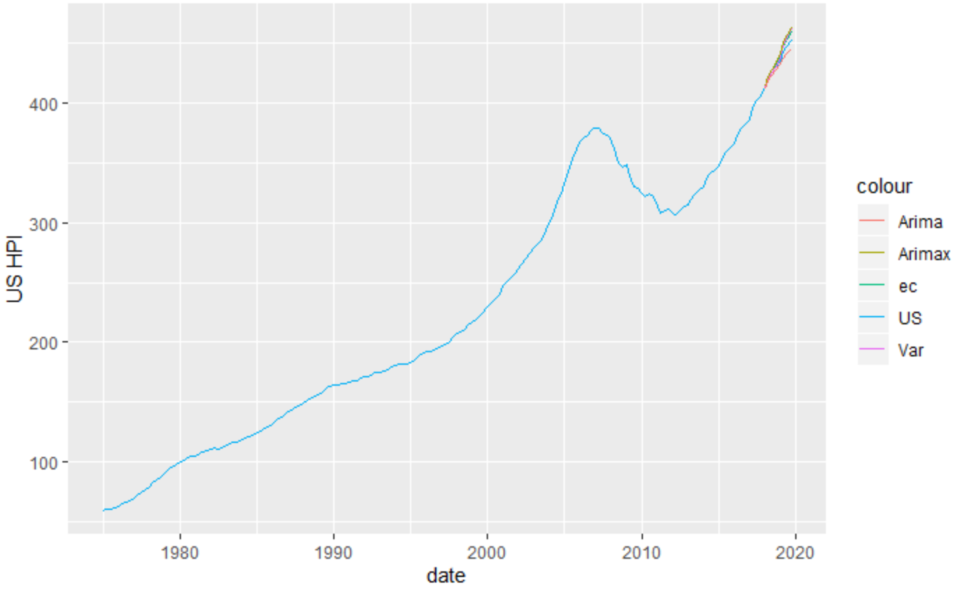

Figure 7: 8-steps ahead forecast

(US: the real value of US HPI in validating period)

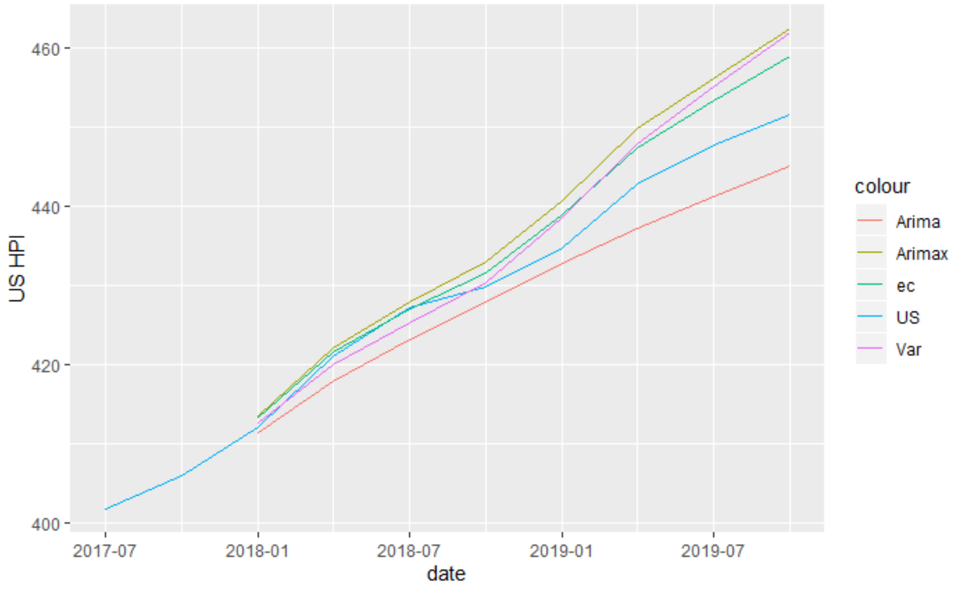

Figure 8: Zoom in the forecasting results

We can easily find that these four forecasting models all have good forecast performance, which are extremely close to the real series in validating period. However, since VECM has lowest RMSE, I came to my final conclusion that VECM is the best model. The Granger test result also supports the whole analysis: The F statistic are statics is statistically significant, which means that CA HPI does have some statistical leading properties to US HPI.

Data Source

[1] https://fred.stlouisfed.org/series/USSTHPI(All-Transactions House Price Index for the United States)

[2] https://fred.stlouisfed.org/series/CASTHPI(All-Transactions House Price Index for California)