Citibeats Home Assignment: French Population

Miguel Conner

Our goal was to analyze the trends in the population of France, using various economic indicators. Up until 2021, the population has apparently increased at a fairly linear rate. We cleaned the data and converted all time series to an annual frequency. Our approach was to first attempt to model the observed data with a univariate approach: an ARIMA model. Then, we tried a multivariate Vector Auto Regression (VAR) approach in which we consider our economic data sets to aid our prediction.

For our ARIMA model we found the best model through a combination of ACF, PACF plots, and comparison of BIC to find `ARIMA(1,3,1)`. We find that using a moving average variable that lags by one and an auto-regressive variable that lags by one is sufficient to achieve reasonably close estimate. Confidence intervals were also computed for our predictions.

Our VAR model, also informed by ACF and PACF plots as well as other graphs, is given below:

$$\hat{p}_t = 1.65 p_{t-1} + 0.65 m_{t-1} - 0.65 p_{t-2} + 3165.04 c_{t-2} - 1034143.59 \tag{1}$$

where $\hat{p}$ is population prediction, p is population, m the net migration, and c the monthly consumer confidence. The $t-1$ and $t-2$ indicate that the variable lags by one or two years, respectively. Thus, our analysis finds that the migration and consumer confidence of the previous one or two years data sets are statistically significant influences in predicting the general population.

In both cases, our models predict fairly similar conclusions that are not entirely obvious by looking at the population graph. They both predict an eventual decline in population over the next two decades, though one the VAR model makes a more aggressive prediction than the ARIMA model.

Further directions and improvements are noted in the `ipython` notebook provided in the repo.