UCLA Anderson School of Management

Time Series Analysis and Forecasting

Instructor: William Yu 12/2/2018

Individual Project: Developing the City Economic Benchmark Index

Introduction:

It is usually more difficult to predict in the context of time series than in cross-sectional data. One source for this difficulty is the trend and cyclical components of time series as opposed to the static nature of cross-sectional data.

In a structural regression model, typical predictors/variables might not be sufficient to capture/predict all the change of dependent variable, in particular for a very dynamic environment with different kinds of trend and cycles.

For instance, a new strategy or a new marketing campaign of a company might have different sales results in Dallas and Detroit. If product sales in Dallas increased but in Detroit they decreased, our data analytics might suggest this strategy/marketing is neutral at best. In fact, to determine whether it is effective, we need to control for the different economic dynamics of these two cities: Dallas was booming while Detroit was not doing well in the past several years. Regardless of specific corporate factor, your sales increase in Dallas and decrease in Detroit in general are partly driven by local economic growth.

What we learned before about the mixed model (both structural and reduce-form) would be a solution to control for the trend and cyclical components of Dallas and Detroit by adding, e.g. AR(1) – a lag term of the dependent variable.

Nevertheless, sometimes it might be challenge to run a mixed model. In this project, we are going to develop a so called City Economic Benchmark Index (CEBI) in order to control/predict the trend/cyclical component for a specific city during a period of time. That is, you can put the CEBI as one of the predictors in your structural model to predict the dependent variable.

Data:

The two most important variables to represent local economic growth and therefore market demand for your products and services are employment/jobs and wages. The data (ces_data.xlxs) is from Current Employment Survey at Bureau of Labor Statistics for 60 largest metros in the U.S. from January 2010 to October 2018. I have downloaded the data for you but you need to convert the series ID to the actual metro names. The details are in the description tab.

If you are interested in exploring and importing data through API, I highly encourage you to do so. See the detail here: https://www.bls.gov/developers/home.htm and https://www.bls.gov/developers/api_sample_code.htm

Task:
Submit your R script or notebook/markdown or Jupyter with some EDAs and result visualization. It will be better to submit some word/pdf files to accompany your R script.

(1) Use the sample period of 2018 (January 2018 to October 2018) as the test-set to validate a best and simple time series model for employment and wages. For instance, a model (across all 60 metros) to give you the smallest RMSE in the test set. You can try all the models we have learned in the class. Note: there should be seasonal component in the data.
(2) After you identify the model, use it to forecast all employment and wages for 60 metros up to the end of 2019. Shows top three and bottom three metros for employment and wages growth from November 2018 to November 2019.
(3) *Remove the seasonal component of the data.* Use some simple decomposition method we learned in the class to remove the seasonal patterns of the data.
(4) *Impute the missing values.* In the wage data, some metros have missing values. Use some simple imputation method to get the full sample.
(5) *Convert the series from its original data to an index.* The reason to do this is because the index will make comparison across metro and time easier and more intuitively. That is all the series in each metro will start as 100 in January 2010. And the series will have the exactly same monthly growth rate as its original data.
(6) *Combine both employment and wage into one index.* One simple way is to take an average of both employment and wage index. Note: depending on the nature of dependent variable, such as company sales, the optimal weight of employment and wage could be reallocated. For example: for an expensive restaurant, the wage composition might be necessary. For an inexpensive fast food restaurant, e.g. McDonald, wage composition might not be important.
(7) *Some Exploratory Test.* Test if CEBI is a significant predictor for some other variables. For instance, in Assignment 2, we analyze Zillow home price index. Run a simple regression to see if CEBI could explain median home prices across the metro. E.g. Y = Median home price percentage change from 1/2010 to 10/2018 for each metro. X = CEPI percentage change from 1/2010 to 10/2018 for each metro.
(8) *Convert the monthly index to weekly and daily index* from January 2010 to December 2019 (including your forecast). The reason is to match your dependent variables, are mostly likely to be daily or weekly.