

Problem Set 7: Forecasting

due Monday April 27, 2020

Instructions: Submissions are individual, R code must be readable, commented and attached at the end of your problem set. Plots, tables and other outputs should be given in the answers or at the end of the problem set.

Problem 1: Forecasting Unemployment

The data set consists of weekly initial jobless claims¹ aggregated to monthly data and the monthly unemployment rate in the US.² Jobless claims are very valuable for policymaking because they are released more often and earlier than unemployment numbers, as a result they are often used to predict the current level of unemployment during recessions. **Make sure to load the `dynlm` package before opening the dataset.**

- i. Initial claims number measure how many individuals filed for unemployment in a given week/month while unemployment numbers measures the total numbers who are unemployed at a given point in time. We will thus try to predict *changes in unemployment* using the claims number rather than the level of unemployment itself. Explain why this is a good strategy using the notion of flow and stock variables from macroeconomics?
- ii. We will consider ADL models of the form:

$$\Delta U_t = \beta_0 + \beta_1 \Delta U_{t-1} + \cdots + \beta_p \Delta U_{t-p} + \delta_1 X_t + \cdots + \delta_q X_{t-q} + e_t,$$

where $\Delta U_t = U_t - U_{t-1}$ is the change in the unemployment rate from $t-1$ to t and X_t are the jobless claims. Estimate the ADL(p, q) model with $p = 0$ and $q = 1, 2, 3$ as well as $p = 1$ and $q = 1, 2, 3$. Report the OLS estimates.

Hint: to estimate the ADL model described above, you can use the following command from the `dynlm` package: `dynlm(diff(Unempl) ~ L(diff(Unempl), 0:p) + L(Claims, 0:q), data = data)`

¹<https://fred.stlouisfed.org/series/ICSA>

²<https://fred.stlouisfed.org/series/UNRATE>

- iii. Using the `window` command, create two subsets of data. The first consists of observations from January 1990 to December 2005 and the second will consist of observations from January 2006 to March 2020. To create the first subset, you could use `data1 = window(data, start = "Jan 1990", end = "Dec 2005")`.

Estimate the regression models in ii. on the first subset of data and compute the forecasted values on the second subset. The forecasts are computed manually using `forecast1 = model1$coef[1] + model1$coef[2]*lag(data2$Claims,0)` for an ADL(0,1) model for instance.

- iv. Compute the RMSFE for the specifications you estimated above and choose the best model for out-of-sample forecasting.
- v. Using the preferred model from the previous question, predict the change in the unemployment rate in April 2020 and, using the March 2020 number, the level of unemployment in April 2020.³

³The jobless claims number for April 2020 are computed using the first two week of data.