EMET3007/8012 Computer Lab 3

Ruitian Lang

August 16, 2023

Part I

In this part, we examine the properties of an OLS estimator. The data generating process is the sum of an AR(2) process with polynomial $1 - .7z - .3z^2$ and a linear trend ct, where c is constant:

$$y_t = y_{t-1} - .3(y_{t-1} - y_{t-2}) + ct + \epsilon_t,$$

where $\{\epsilon_t\}$ is white noise. In this part, we will use the value $c = .2$. For a sample generated by this process, we estimate the following model using OLS with particular interest in the coefficient a_1:

$$y_t = a_1 y_{t-1} + \tilde{a}_2(y_{t-1} - y_{t-2}) + \mu + \tilde{c}t + \epsilon_t.$$

Of course, the true value of a_1 is 1.

Task 1

Try to determine whether the OLS estimator of a_1 is consistent and asymptotically normal. A full mathematical proof can be accepted as an answer, in which case the hurdle task is to show your proof to your tutor, who may quiz you on some random details of your proof. If you decide to complete this task by simulation, the hurdle is to show the histogram and Q-Q plot of the estimator to your tutor.

Task 2

Use simulation to determine the validity of the ADF test in the presence of the linear trend. We generate $N = 10000$ sample paths of length $T = 400$ and run the ADF test on each sample path. Use your UID as the random seed. We reject the unit root if the p-value reported by the ADF test is below .05. How many times do you expect unit root to be rejected? Run the simulation and report your result.

Part II

In this part, we try to forecast the quarterly earnings per share (EPS) of Johnson & Johnson's in 1980 using data from 1960 to 1979. The data file can be downloaded from Marco Peixeiro's Github; it is in the "data" folder and called "jj.csv".

Through the entire task, the training set is data up to the end of 1979, or the first 80 observations, and the test set is 1980, or the last four observations. In particular, you should use the training set only for Task 1.

Task 1

Determine the order of integration d using the training set. If the original time series is stationary, then $d = 0$. Otherwise, you keep differencing the series until you can reject unit root at the 5% level. The integration order d is the number of times you need to difference the series before the resultant time series is stationary.

Task 2

Throughout the part, we only forecast one quarter ahead. For example, we only forecast the EPS of 1980Q1 initially. Then we use the known value of 1980Q1 EPS to forecast for 1980Q2, and so on.

We compare the MSE produced by three forecasting models.

1. Use y_{t-1} as the forecast of y_t.

2. Use $(y_{t-1} + y_{t-2} + y_{t-3} + y_{t-4})/4$ as the forecast of y_t.

3. Use the ARIMA(4,d,4) model, where d is the integration order you determined in the previous task. You do no need to implement the estimation of the model yourself. The class statsmodels.tsa.arima.model.ARIMA can be used to fit the model and produce forecast. Read the documentation of the class to learn how. You do not have to re-estimate the model after each quarter of 1980, but it is OK if you do re-estimate.

For the purpose of this project, you do not need to implement "rolling forecast", as you have only four forecasting steps and you can perform them one by one. However, you must implement the first two forecasting models in Python code; you are NOT allowed to use your calculator to produce the forecasts and hard-code them.

Plot the actual EPS and your forecasts in the same plot. You can either make a separate plot for each forecasting models or show the results of all three models in the same plot.

Hurdle: show plot(s) of your first two forecasts to your tutor.