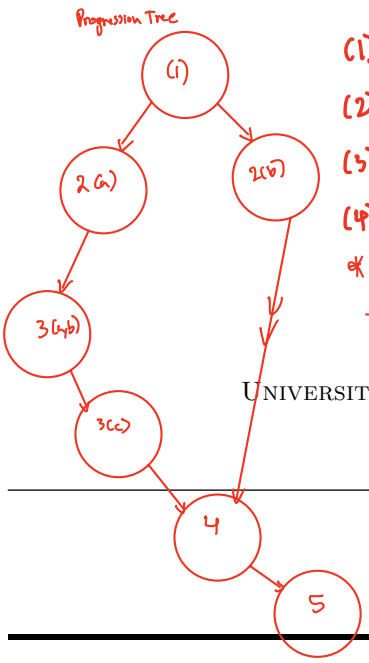

University of California, Los Angeles, CA
ECON 104, Fall 2023

Project 2

Due: November 17, 2023

Select a **time series** dataset which contains AT MOST two predictor variables and one response variable (y_t). We need to analyze and clean the data to ascertain a model which best fits the data. The suggested datasets for this project are:

- BenderleyZwick, AER
- PepperPrice, AER
- USGasB, AER
- TechChange
- OrangeCounty

Again, you are not confined to the aforementioned datasets. There are many publicly available timeseries datasets through Google DataSet Search, Kaggle, AER - base R datasets, etc.

1. **Exploratory Data Analysis**
   (a) Briefly discuss the question you are trying to answer.
   (b) Cite the dataset and give a summary of what the dataset is about
   (c) First check for completeness and consistency of the data (if there are NAs or missing observations, replace with the value of the previous observation; make a note of this)
   (d) Provide descriptive analyses of your variables. This should include the histogram with overlying density, boxplots, cross correlation. All figures/statistics must include comments.

2. **Data PreProcessing**

   (a) With `tsdisplay` or `ggtsdisplay`, for each variable, use its time series plot, ACF and PACF to comment on its stationarity (you can also decompose the time series; note if there is seasonality). To supplement this, use the appropriate Dickey-Fuller (unit root) test, to determine whether or not it is stationary. Note using its PACF what the suspected order might be.

   (b) If it is not stationary, determine the level of differencing to make our series stationary. We can use the `ndiffs` function which performs a unit-root test to determine this. After this, difference your data to ascertain a stationary time series. Re-do part a) for your differenced time series and comment on the time series plot, ACF and PACF. Recall that the time series models we've observed rely on stationarity.

3. **Feature Generation, Model Testing and Forecasting.**

   (a) Fit an AR(p) model to the data (using part 2(a), AIC or some built in R function)

   (b) Plot and comment on the ACF of the residuals of the model chosen in 3(a). If the model is properly fit, then we should see no autocorrelations in the residuals. Carry out a formal test for autocorrelation and comment on the results.

   (c) Using the appropriate predictors, fit an ARDL(p,q) model to the data and repeat step (b) in part 3.

4. Provide a brief summary of your findings and state which model performs better.

5. Suggest any limitations faced or improvements which could've been made to the model based on your findings, which should be supplemented with statistical tests(eg. degree of freedom restrictions, reverse causality).