Assignment 1

General Comments

- Please write a self-explanatory report. Explain your answers using economic and/or econometric reasoning and any equations (or figures) needed to make your point. Calculations: Provide final results and also show how you achieved your results using intermediate steps.
- The report may not exceed 5 pages.
- Front page, bibliography and appendices are excluded from the page count. There is no character count. Instead incorporate the format specifications provided in the "Formal Requirements for Standard Page."
- Tables and graphs you refer to should be in the main text of your report. Every number in any of your tables needs to be explained in your text. Do NOT paste raw output from Stata or other software directly into the main text of your report (except for figures). You can use Stata or other software to obtain results and figures.
- Upload your report in pdf-format before the deadline.
- It is not necessary to hand-in programming code. If you want to provide your code, include it in an appendix and refer to it only if needed. Make sure everything, including the code itself, is contained in a single pdf file.
- If you find help in books, papers, or on the web, remember to cite these sources in your report. Please only cite publicly available sources (do not cite the lecture slides).
- Collaboration among different groups is strictly forbidden and assignments will be checked for excessive similarity.
- Throughout the assignment use a 5% significance level (if not indicated otherwise). The normal distribution is denoted as $N(\mu, \sigma^2)$

Part 1: Simulations

Trends in Time Series

Simulate the following model for different values of the parameters (α, β, ϕ) specified below.

$$y_t = \alpha + \beta t + \phi y_{t-1} + \epsilon_t, \qquad \epsilon_t \sim IIDN(0,1)$$

- Set the seed to 1232020 at the beginning of your simulations. Assume a starting value $y_0 = 0$ where necessary and obtain 500 observations for four different sets of parameters:

1. $(\alpha, \beta, \phi) = (1, 0, 0.8)$
2. $(\alpha, \beta, \phi) = (1, 0.1, 0.8)$
3. $(\alpha, \beta, \phi) = (0, 0, 1)$
4. $(\alpha, \beta, \phi) = (0.1, 0, 1)$

- Provide four separate line-plots of the four data series for the first 50 observations only.
- Provide four separate line-plots of the four data series for all 500 observations.
- How can you characterize the four time series? Are they stationary? Do they have a drift? Do they have a trend?
- How can the characteristics be assessed from the 50 periods plots? From the 500 periods plots?
- Explain how you could make the four data series stationary (in case they are not stationary).

The Dangers of Non-Stationarity

- Obtain 150 observations from two independent variables x and y (with $x_0 = 0$ and $y_0 = 0$):

$$x_t = 1.004 x_{t-1} + u_t, \qquad y_t = 1.04 y_{t-1} + v_t$$

  where you simulate u_t and v_t as independent Gaussian (=normally distributed) white noise processes with $\mu = 0$ and $\sigma = 0.01$. (You can look up the steps you have to take for generating variables from a random distribution in Problem Set 1, Exercise 2. Remember that you can call the previous value of a variable as variable[_n-1]).

- Are x and y stationary? Why / why not? Which time series processes do x and y come from?
- `regress` y on x to estimate the OLS regression model:

$$y_t = \alpha + \beta x_t + \epsilon_t$$

- `generate` a variable $R2$ and store the regression's R^2 into the variable $R2$ using the `replace` command. You can either calculate R^2 by hand as

$$R^2 = \hat{\beta}^2 \hat{\sigma_x}^2 / \hat{\sigma_y}^2$$

  or obtain the value from the results Stata stores for the most recent regression.
  ($\rightarrow$ help regress $\rightarrow$ scroll down to Stored results)

- Program a `forvalues` loop to repeat these tasks 150 times in total and every time store the results for R^2 in the variable $R2$.
- Repeat the whole exercise for the variables xs and ys:

$$xs_t = 0.96 x_{t-1} + u_t, \qquad ys_t = 0.58 y_{t-1} + v_t$$

  Remember to store the results for R^2 into another variable $Rs2$.

- Obtain two `histograms` for $R2$ and $Rs2$. What is your interpretation? What is the danger of non-stationarity?
- What is the probability that you observe $R2 > 0.8$ and $Rs2 > 0.8$?
- What happens if $\phi >> 1$? What happens if $t = 1000$?

Part 2: - Dividend-Price Ratio Regressions

Test whether it is possible to predict stock returns out-of-sample. Use the data set `dp_data` (obtained from Amit Goyal's webpage). For the years 1950 to 2018, the annual data contains: S&P 500 index level `sp500index`, and the S&P 500 index returns from the Center for Research in Security Prices (CRSP) `CRSP_SPvw_ret`. Returns are the log-returns on the S&P 500 index, including dividends. Dividends paid on the S&P 500 index are named `div`.

1. Calculate the dividend price ratio as:

$$(D/P)_t = ln(D_t) - ln(P_t)$$

   Conduct a regression of returns on the one-period lagged dividend price ratios using the first 20 years of observations. Use the estimated coefficients and the dividend price ratio of year 20 to predict the return of year 21. Then move the estimation window one year forward and repeat; always use the past 20 years of observations for parameter estimation, i.e. the estimation is "rolling".
   Based on your rolling regressions:

   (a) Compute the time-series of the rolling OLS coefficients. What is your average beta estimate for the coefficient on the lagged dividend price ratio? Plot the time series of beta estimates.

   (b) Compute the time-series of predicted returns ($\hat{r_t}$) using only data that is available in the respective period. Example: if the calibration period is 1963-1982, then use the dividend-price-ratio during that time to calculate the predicted stock return (for year 1983). How many predictions do you obtain? Provide mean and variance of the predicted returns.

   (c) Provide a time-series line graph of predicted returns, the average return, and realized returns beginning in the year 1971.

2. Compute the "out-of-sample" R^2_{OOS}, which is defined as

$$R^2_{OOS} = 1 - \frac{\frac{1}{T}\sum(r_t - \hat{r_t})^2}{\frac{1}{T}\sum(r_t - \frac{1}{T}\sum r_t)^2}$$

   On how many residuals is the R^2_{OOS} based? Provide the value you obtain for R^2_{OOS}.

3. Obtain the in-sample R^2 from the equation:

$$ret_{t+1} = a + b(D/P)_t + \epsilon_{t+1}$$

   Estimate the equation over the entire sample period. Provide the R^2 of the regression. What is there difference between this and the former estimation procedure? Compare the in-sample R^2 to the out-of-sample R^2_{OOS}. Was it possible to make money with the dividend-price ratio as a predictor?