# MBA Big Data Econometrics 2020–2021

## Assignment 3

The due date for handing in the assignment is Tuesday December 8, 23:59h.

1. We analyse annual data on the S&P 500 stock market index and related time series, with observations over the period 1871–2015. The data are given in `SP500.csv`, and the variables are defined as follows:

   | | |
   |---|---|
   | `P` | S&P 500 index (value-weighted average of stock prices of 500 important US companies) |
   | `D` | Annual (value-weighted average) dividend per share on the stocks in the index |
   | `Rs` | Short-term (one-year) US interest rate |
   | `Rl` | Long-term (10-year) US interest rate |
   | `CPI` | Consumper price index |
   | `lrp` | $= \ln(\mathrm{P}/\mathrm{CPI})$, logarithm of "real" index, i.e., corrected for changes in consumer prices |
   | `lrd` | $= \ln(\mathrm{D}/\mathrm{CPI})$, logarithm of real dividends |
   | `lpd` | $= \ln(\mathrm{P}/\mathrm{D})$, logarithm of price-dividend ratio |
   | `ret` | $= (\mathrm{P}_t + \mathrm{D}_t - \mathrm{P}_{t-1})/\mathrm{P}_{t-1}$, annual return on the index, including dividends |
   | `TSpr` | $= \mathrm{Rl} - \mathrm{Rs}$, difference between long- and short-term interest rate (term spread) |

   The file also contains up to 3 lags of the (lower case) variables, indicated by the extension "`_j`", $j = 1, 2, 3$.

   (a) Test for a unit root in `lrp`, `lrd`, `Rs` and `Rl`. Motivate your choice between either a constant only, or a constant and a linear trend in the test regression. Report and interpret the outcome of the tests.

   (b) Particular financial theories imply that log (real) stock prices and log (real) dividends should be cointegrated, with a coefficient of $1$. Explain that, under the assumption that both prices and dividends have a unit root, this hypothesis can be tested with a unit root test on the variable `lpd`, and report and interpret the outcome of this test.
   (The assumption of a unit root in both series may not be supported by your answer to (a); this would have to be included in your discussion.)

   (c) We now wish to investigate if the annual returns on the S&P 500 index can be forecasted. First, plot the autocorrelation function of `ret`, and interpret the outcome. Next, estimate an AR(2) model for `ret` and test if the lagged returns have zero coefficients (jointly).

   (d) Estimate an ADL model with `ret` as the dependent variable, and with three lags of `ret` and the term spread `TSpr` as explanatory variables. (You may assume that `TSpr` is stationary.) Carry out a Granger-causality test to see if the term spread Granger-causes returns. Interpret the outcome.

   (e) Select an appropriate lag order $p$ (motivate your choice), and estimate the resulting VAR($p$) model for `ret` and `TSpr` jointly. Report and interpret the outcomes, focussing in particular on the difference in predictability of the two time series.