Assignment_3_RutgerGeelen_WashantvanDam

Rutger Geelen (13510517) and Washant van Dam (13510053)

December 8th, 2020

Preparation

Clear workspace and load the required packages:

Q1

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 DAnnual (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(P/CPI), logarithm of "real" index, i.e., corrected for changes in consumer prices
- lrd = ln(D/CPI), logarithm of real dividends
- lpd = ln(P/D), logarithm of price-dividend ratio
- ret = (Pt + Dt - Pt-1)/Pt-1, annual return on the index, including dividends
- TSpr = Rl-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.

Read data:

```
MyQ1Data <- read.csv("https://raw.githubusercontent.com/rutgerg/econometrics_assigment_3/master/SP500.cs
MyQ1Data = na.omit(MyQ1Data)
```

Q1A

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 test.

If all roots are greater than 1 in absolute value, the AR(p) series is stationary. If at least one root equals 1, the AR(p) is said to have a unit root and thus has a stochastic trend.

We use the ADF test for a unit auto-regressive root to test the hypothesis H0:d=0 (stochastic trend) against the one-sided alternative H1:d<0 (stationarity) using the usual OLS t-statistic.

General specification: deltaPt = b0 + (a * t) + d * Pt-1 + g1 * deltaPt-1 + …. + gp-1 * deltaPt-p+1 + ut

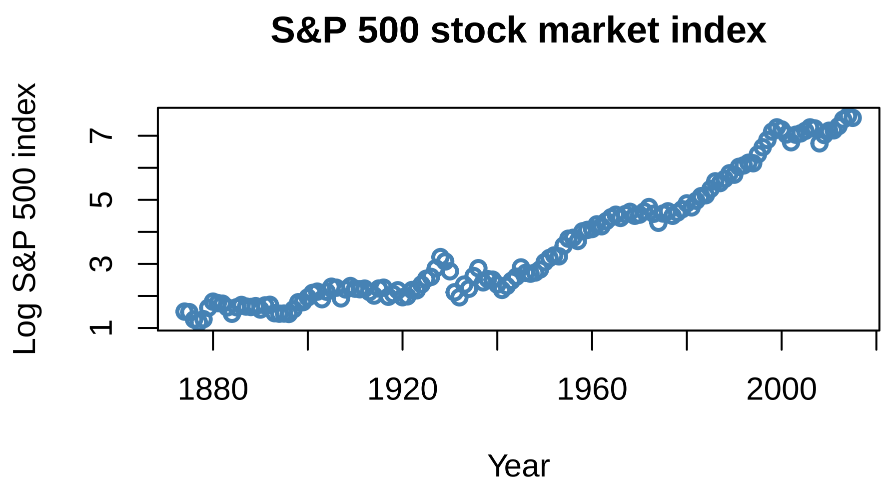
b0 is intercept (a * t) is linear time trend

d = B1 + .. + Bp - 1

H0: d = 0 (unit root) H1: d < 0 (stationarity)

Intercept only specification only if there is not long term growth in the series (eg interest, inflation, unemployment). Since this is index data we expect intercept and linear time trend. To check we first plot the data:

```
plot(MyQ1Data$X,log(MyQ1Data$P),
     col = "steelblue",
     lwd = 2,
     xlab = 'Year',
     ylab = 'Log S&P 500 index',
     main = "S&P 500 stock market index")
```



Since we she a stochastic upward trend we choose the specification including intercept and time trend.

Let's do the Augmented Dickey-Fuller (ADF) test with maximum 5 lags:

```
tsdata1a <- ts(MyQ1Data$P,MyQ1Data$X)
adf1a <- ur.df(tsdata1a, type = "trend", lags = 5, selectlags = "BIC")
summary(adf1a)
```

```
##
## ###############################################
## # Augmented Dickey-Fuller Test Unit Root Test #
## ###############################################
##
## Test regression trend
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -452.10  -12.62   -2.84    6.96  264.37
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -15.36025   15.21563  -1.010  0.31464
## z.lag.1       0.05012    0.02440   2.054  0.04202 *
## tt            0.36472    0.22830   1.598  0.11260
## z.diff.lag1   0.08607    0.08899   0.967  0.33528
## z.diff.lag2  -0.28056    0.08830  -3.177  0.00186 **
```

```
## z.diff.lag3  -0.10939    0.09229  -1.185  0.23812
## z.diff.lag4  -0.14654    0.08990  -1.630  0.10555
## z.diff.lag5  -0.42753    0.09094  -4.701 6.59e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 73.86 on 128 degrees of freedom
## Multiple R-squared:  0.2661, Adjusted R-squared:  0.226
## F-statistic:  6.63 on 7 and 128 DF,  p-value: 1.08e-06
##
##
## Value of test-statistic is: 2.0539 8.1619 9.0781
##
## Critical values for test statistics:
##       1pct  5pct 10pct
## tau3 -3.99 -3.43 -3.13
## phi2  6.22  4.75  4.07
## phi3  8.43  6.49  5.47
```

We need to recalculate the p-value since we can not use the p-value from the linear regression above which assumes normal distribution and 2 sided test.

```
pval1a <- punitroot(adf1a@teststat[1], N=Inf, trend='ct', statistic='t')
cat("p-value of ADF test:", pval1a)
```

```
## p-value of ADF test: 0.9999995
```

Since Padf-value is 0.99 we do not reject H0 so there is non stationary data and have unit root.

Q1B

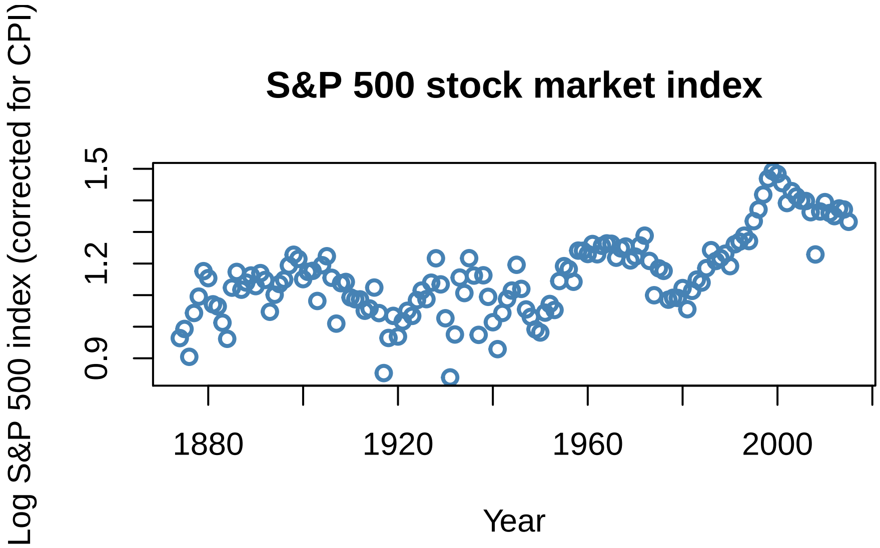
Particular financial theories imply that log (real) stock prices and log (real) dividends should be cointegrated, with a coefficient of 1. Explain that, under the assumption that both prices and dividends have a unit root, this hypothesis can be tested with a unit root test on the variable lpd, and report 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.)

Cointegration happens when two or more series share the same stochastic trend. In this case that makes sense since stockprices rise (cp) with rising dividends because it assumes higher future cashflows. Cointegration says that if both series are non-stationary then difference can be stable and stationary and H0 should be rejected in favor of H1.

Coefficient of 1 assumes thèta = 1

To check we first plot the data:

```
plot(MyQ1Data$X,log(MyQ1Data$lpd),
     col = "steelblue",
     lwd = 2,
     xlab = 'Year',
     ylab = 'Log S&P 500 index (corrected for CPI)',
     main = "S&P 500 stock market index")
```



It looks like a positive trend; not stationary.

Let's test:

```
tsdata1b <- ts(MyQ1Data$lpd,MyQ1Data$X)
adf1b <- ur.df(tsdata1b, type = "trend", lags = 5, selectlags = "BIC")
summary(adf1b)
```

```
##
## ###############################################
## # Augmented Dickey-Fuller Test Unit Root Test #
## ###############################################
##
## Test regression trend
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.62235 -0.11534  0.00943  0.11861  0.46315
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.3987696  0.1548224   2.576  0.01111 *
## z.lag.1     -0.1493394  0.0552352  -2.704  0.00777 **
## tt           0.0012557  0.0005625   2.232  0.02728 *
## z.diff.lag1 -0.0596104  0.0850817  -0.701  0.48478
## z.diff.lag2 -0.2728431  0.0833521  -3.273  0.00136 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1903 on 131 degrees of freedom
## Multiple R-squared:  0.1751, Adjusted R-squared:   0.15
## F-statistic: 6.954 on 4 and 131 DF,  p-value: 4.158e-05
##
##
```

```
## Value of test-statistic is: -2.7037 2.6161 3.8103
##
## Critical values for test statistics:
##       1pct  5pct 10pct
## tau3 -3.99 -3.43 -3.13
## phi2  6.22  4.75  4.07
## phi3  8.43  6.49  5.47
```

Again we need to recalculate the p-value since we can not use the p-value from the linear regression above which assumes normal distribution and 2 sided test.

```
pval1b <- punitroot(adf1b@teststat[1], N=Inf, trend='ct', statistic='t')
cat("p-value of ADF test:", pval1b)
```

```
## p-value of ADF test: 0.2349874
```

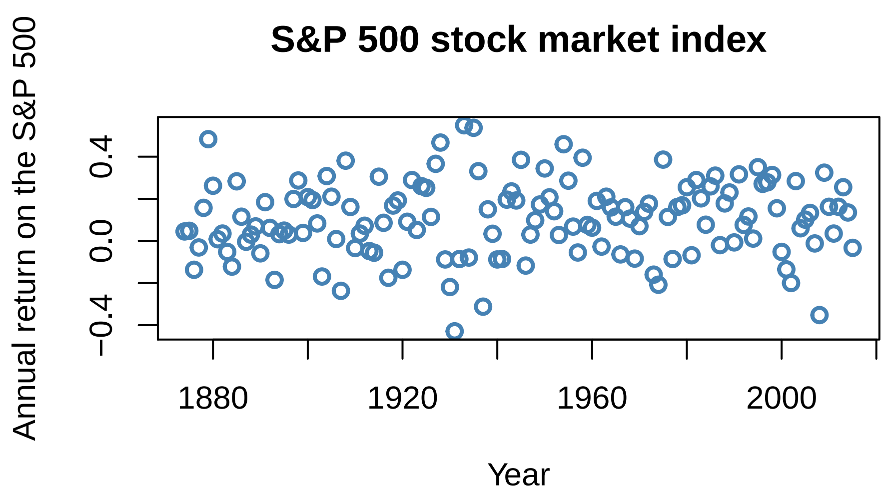
Since Padf-value is 0.23 we do not reject H0 so there is non-stationary data and have unit root. There is not enough evidence to support cointegration.

Q1C

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).

First we plot the data:

```
plot(MyQ1Data$X,MyQ1Data$ret,
     col = "steelblue",
     lwd = 2,
     xlab = 'Year',
     ylab = 'Annual return on the S&P 500',
     main = "S&P 500 stock market index")
```



The outcome suggest ret is constant and stationary; there is no trend.

Let's test AR(2):

```
aq1c  <- lm(ret ~ ret_1 + ret_2, data=MyQ1Data)
linearHypothesis(aq1c, c("ret_1=0","ret_2=0"), vcov = vcovHC(aq1c,"HC1"))
```

```
## Linear hypothesis test
##
```

```
## Hypothesis:
## ret_1 = 0
## ret_2 = 0
##
## Model 1: restricted model
## Model 2: ret ~ ret_1 + ret_2
##
## Note: Coefficient covariance matrix supplied.
##
##   Res.Df Df      F Pr(>F)
## 1    141
## 2    139  2 2.3845 0.0959 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

F statistic is significant at 5% level so we don't reject H0 that ret_1 and ret_2 are jointly 0.

Q1D

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.

The Granger causality test does not test whether X actually causes Y but whether the included lags are informative in terms of predicting Y.

F-test for H0 : d1 = . . . = dq = 0 (non-causality) in Yt =b0 +B1 * Yt-1 +. . .+ bp * Yt-p + d1 * Xt-1 +. . .+ dq * Xt-q + ut.

ADL(3,3) model and testing for joint predictiveness.

```r
aq1d  <- lm(ret ~ ret_1 + ret_2 + ret_3 + TSpr_1 + TSpr_2 + TSpr_3, data=MyQ1Data)
linearHypothesis(aq1d, c("ret_1=0","ret_2=0","ret_3=0","TSpr_1=0","TSpr_2=0","TSpr_3=0"), vcov = vcovHC
```

```
## Linear hypothesis test
##
## Hypothesis:
## ret_1 = 0
## ret_2 = 0
## ret_3 = 0
## TSpr_1 = 0
## TSpr_2 = 0
## TSpr_3 = 0
##
## Model 1: restricted model
## Model 2: ret ~ ret_1 + ret_2 + ret_3 + TSpr_1 + TSpr_2 + TSpr_3
##
## Note: Coefficient covariance matrix supplied.
##
##   Res.Df Df      F   Pr(>F)
## 1    141
## 2    135  6 3.3533 0.004117 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Conclusion: p-value is small and F statistic high significance so we do not reject H0. Hence the variables have predictive power.

Q1E

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, focusing in particular on the difference in predictability of the two time series.

We calculate the appropriate log order using BIC

BIC(K) = ln(SSR(K)/T) + K x ln(T)/T

```
# BIC calculation below used from https://www.econometrics-with-r.org/14-6-llsuic.html
BIC <- function(model) {

  ssr <- sum(model$residuals^2)
  t <- length(model$residuals)
  npar <- length(model$coef)

  return(
    round(c("p" = npar - 1,
          "BIC" = log(ssr/t) + npar * log(t)/t,
          "R2" = summary(model)$r.squared), 4)
  )
}

order <- 1:12

BICs <- sapply(order, function(x)
        "AR" = BIC(dynlm(ts(MyQ1Data$ret) ~ L(ts(MyQ1Data$ret), 1:x) + L(ts(MyQ1Data$TSpr), 1:x))))
BICs
```

```
##        [,1]    [,2]    [,3]    [,4]    [,5]    [,6]    [,7]    [,8]    [,9]
## p    2.0000  4.0000  6.0000  8.0000 10.0000 12.0000 14.0000 16.0000 18.0000
## BIC -3.3330 -3.3277 -3.2804 -3.2322 -3.1697 -3.1192 -3.0535 -2.9708 -2.8869
## R2   0.0025  0.0721  0.0895  0.1143  0.1300  0.1290  0.1388  0.1379  0.1376
##       [,10]   [,11]   [,12]
## p   20.0000 22.0000 24.0000
## BIC -2.8314 -2.7620 -2.6887
## R2   0.1585  0.1628  0.1686
```

```
BICs[, which.min(BICs[2, ])]
```

```
##       p     BIC      R2
##  2.0000 -3.3330  0.0025
```

p with minimum BIC: 2

The resulting VAR(2) model:

```
VAR_EQ1 <- dynlm(ts(MyQ1Data$ret) ~ L(ts(MyQ1Data$ret), 1:2) + L(ts(MyQ1Data$TSpr), 1:2))
VAR_EQ2 <- dynlm(ts(MyQ1Data$TSpr) ~ L(ts(MyQ1Data$ret), 1:2) + L(ts(MyQ1Data$TSpr), 1:2))
coeftest(VAR_EQ1, vcov. = sandwich)
```

```
##
## t test of coefficients:
##
##                              Estimate Std. Error t value  Pr(>|t|)
## (Intercept)                  0.124210   0.021656  5.7356 6.094e-08 ***
## L(ts(MyQ1Data$ret), 1:2)1    0.062844   0.099124  0.6340  0.527157
## L(ts(MyQ1Data$ret), 1:2)2   -0.253928   0.096572 -2.6294  0.009546 **
```

```
## L(ts(MyQ1Data$TSpr), 1:2)1 -0.011712   0.012350 -0.9483  0.344669
## L(ts(MyQ1Data$TSpr), 1:2)2  0.026401   0.011644  2.2673  0.024960 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
coeftest(VAR_EQ2, vcov. = sandwich)
```

```
##
## t test of coefficients:
##
##                              Estimate Std. Error t value  Pr(>|t|)
## (Intercept)                   0.29437    0.12359  2.3819  0.018621 *
## L(ts(MyQ1Data$ret), 1:2)1  -0.68283    0.48890 -1.3967  0.164807
## L(ts(MyQ1Data$ret), 1:2)2  -1.89253    0.53361 -3.5467  0.000537 ***
## L(ts(MyQ1Data$TSpr), 1:2)1  0.58380    0.12623  4.6250 8.674e-06 ***
## L(ts(MyQ1Data$TSpr), 1:2)2  0.13103    0.11709  1.1191  0.265093
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ret t = 0.12 + 0.06 * ret t-1 - 0.25 * ret t-2 - 0.01 * Tspr t-1 + 0.02 * Tspr t-2

Tspr t = 0.29 - 0.68 * ret t-1 - 1.89 * ret t-2 + 0.58 * Tspr t-1 + 0.13 * Tspr t-2

Running the VAR function gives the same estimates:

```r
VAR_data <- cbind("ret" = MyQ1Data$ret, "TSpr" = MyQ1Data$TSpr)
VAR_est <- VAR(y = VAR_data, p=2, type="const")
summary(VAR_est)
```

```
##
## VAR Estimation Results:
## =========================
## Endogenous variables: ret, TSpr
## Deterministic variables: const
## Sample size: 140
## Log Likelihood: -168.814
## Roots of the characteristic polynomial:
## 0.6817 0.4423 0.4423 0.1252
## Call:
## VAR(y = VAR_data, p = 2, type = "const")
##
##
## Estimation results for equation ret:
## ====================================
## ret = ret.l1 + TSpr.l1 + ret.l2 + TSpr.l2 + const
##
##         Estimate Std. Error t value Pr(>|t|)
## ret.l1   0.06284    0.08696   0.723  0.47112
## TSpr.l1 -0.01171    0.01289  -0.909  0.36518
## ret.l2  -0.25393    0.08676  -2.927  0.00402 **
## TSpr.l2  0.02640    0.01299   2.033  0.04405 *
## const    0.12421    0.01936   6.416 2.19e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 0.1766 on 135 degrees of freedom
```

```
## Multiple R-Squared: 0.07213, Adjusted R-squared: 0.04463
## F-statistic: 2.623 on 4 and 135 DF,  p-value: 0.0375
##
##
## Estimation results for equation TSpr:
## =====================================
## TSpr = ret.l1 + TSpr.l1 + ret.l2 + TSpr.l2 + const
##
##         Estimate Std. Error t value Pr(>|t|)
## ret.l1  -0.68283    0.58244  -1.172  0.24312
## TSpr.l1  0.58380    0.08634   6.762 3.75e-10 ***
## ret.l2  -1.89253    0.58109  -3.257  0.00142 **
## TSpr.l2  0.13103    0.08700   1.506  0.13435
## const    0.29437    0.12967   2.270  0.02479 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 1.183 on 135 degrees of freedom
## Multiple R-Squared: 0.4481,  Adjusted R-squared: 0.4317
## F-statistic:  27.4 on 4 and 135 DF,  p-value: < 2.2e-16
##
##
##
## Covariance matrix of residuals:
##          ret    TSpr
## ret  0.03119 0.05016
## TSpr 0.05016 1.39908
##
## Correlation matrix of residuals:
##         ret   TSpr
## ret  1.0000 0.2401
## TSpr 0.2401 1.0000
```

ret: The negative -0.25 with a 99% CI for ret t-2 is peculiar since we would not expect a strong negative relation between ret now and 2 year ago.

Tspr: The positive relation between Tspr now and past makes sense. A negative relation with ret would also make sense since higher interest spread might lure investors away from investing in the index and in stead invest in LT interest related products like bonds.