Master 2 Financial Economics – Université Paris 1 Panthéon-Sorbonne

Empirical Application 3

Asset pricing

Roland BOUILLOT
(Roland.Bouillot@etu.univ-paris1.fr)

Khalil JANBEK
(Khalil.Janbek@etu.univ-paris1.fr)

Mehdi LOUAFI
(Mehdi.Louafi@etu.univ-paris1.fr)

September 2021

EMPIRICAL APPLICATION 3

Question 1 – Consider the series of the (log)- prices for a stock index over a given period. Justify the choice of the period and of the frequency.

Of all the possible options, we have selected the S&P500 Stock Index (SP500) to conduct our empirical analysis, due to its representativeness of the state of the world's financial markets.

We aim at estimating the long-term fundamental value of the S&P500 index using the Present Value Model, expressed in its simplest form in the Gordon-Shapiro formula. We therefore rely on the three following series, over the longest period available to us, in order to capture their long-term joint dynamics and thereby to better perform our long-term fundamental analysis:

- S&P500: in levels, from 1985 Q1 to 2021Q2, on a quarterly basis (end of period);
- US real GDP: in levels, from 1985 Q1 to 2021Q2, on a quarterly basis (end of period);
- US 10-year Treasury yield: in %, from 1985 Q1 to 2021Q2, on a quarterly basis (end of period).

Our data are in quarterly frequency, in order to match the frequency of the US real GDP data.

Question 2 – Add the relevant series representative of the fundamentals. Check that all series are I(1). By referring to the Gordon-Shapiro model and its development presented in the course or derived from other references:
- Justify the existence of a long-term equation to describe the equilibrium price (estimation and test) with possible structural breaks
- Estimate the corresponding error-correcting equation

Our goal is to identify a "reliable fundamental value" of the S&P500 stock index, "standing for a long-run target value in Error-Correction Modelling of the dynamics of […] returns"[1] (Bruneau et al., 2010).

Such an identification is extremely useful from an investment perspective, for two reasons:

1- The fundamental value of the stock index provides a benchmark against which analyzing the present valuation of the index. Such a benchmark is useful for implementing contrarian strategies (i.e. "trading against prevailing market sentiments"[2]) or value investing strategies (i.e. "selecting currently undervalued stocks that you expect to increase in value in the future"[3]).
2- Gauging the fundamental value of the index enables us to identify whether price its developments are transitory shocks and a long-term change in the fundamental value.

---

[1] C. Bruneau ; C. Duval-Kieffer, J.P. Nicolai, *Managing funds in the US market: how to distinguish between transitory distortions and structural changes in the stock prices?*, The European Journal of Finance, 2000
[2] Corporate Finance Institute, on contrarian strategies
[3] Corporate Finance Institute, on value investing

For the value of a stock, the Gordon-Shapiro model "suggests two fundamentals: dividends and a discount rate factor, specified as a risk-free rate plus an ex-ante risk premium", as shown below:

Equation 1:

$$\boldsymbol{P_t} = \frac{\boldsymbol{D_t}}{\boldsymbol{R^e}_t - \boldsymbol{g^e}_t}$$

Where:

$\boldsymbol{P_t}$ **=** Price of the stock (index) at time t;
$\boldsymbol{D_t}$ **=** Dividends at distributed by the stock at time t;
$\boldsymbol{R^e}_t$ **=** the expected stock return;
$\boldsymbol{g^e}_t$ **=** the expected dividend growth rate.

To be more specific, we consider the expected stock return $\boldsymbol{R^e}_t$ as the sum of a risk-free return $\boldsymbol{r^F}_t$ and an ex-ante risk premium $\boldsymbol{\Pi^e}_t$:

Equation 2:

$$\boldsymbol{P_t} = \frac{\boldsymbol{D_t}}{\boldsymbol{r^F}_t + \boldsymbol{\Pi^e}_t - \boldsymbol{g^e}_t}$$

Taking the logarithm on both sides of the equation, we obtain:

Equation 3:

$$\boldsymbol{log(P_t)} = \boldsymbol{log(D_t)} - \boldsymbol{log(r^F{}_t + \Pi^e{}_t - g^e{}_t)}$$

For the estimation of the long-term equation describing the equilibrium price of the S&P500, we use:

- The logarithm of the real GDP of the USA as a proxy for dividends $\boldsymbol{log(D_t)}$ as it provides a good approximation of the activity of the firms.
- The 10-year US Treasury bond yields as a proxy for the risk-free rate $\boldsymbol{r^F}_t$**.**

However, as shown by the below graphs, the series do not appear to be stationary.

Chart 1

Logarithm of the S&P500

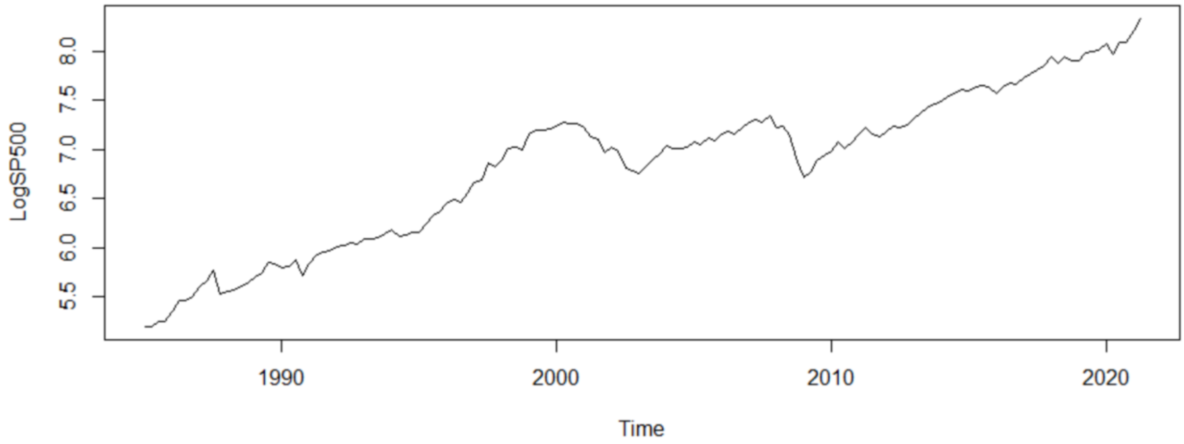


Chart 2

Logarithm of the US real GDP

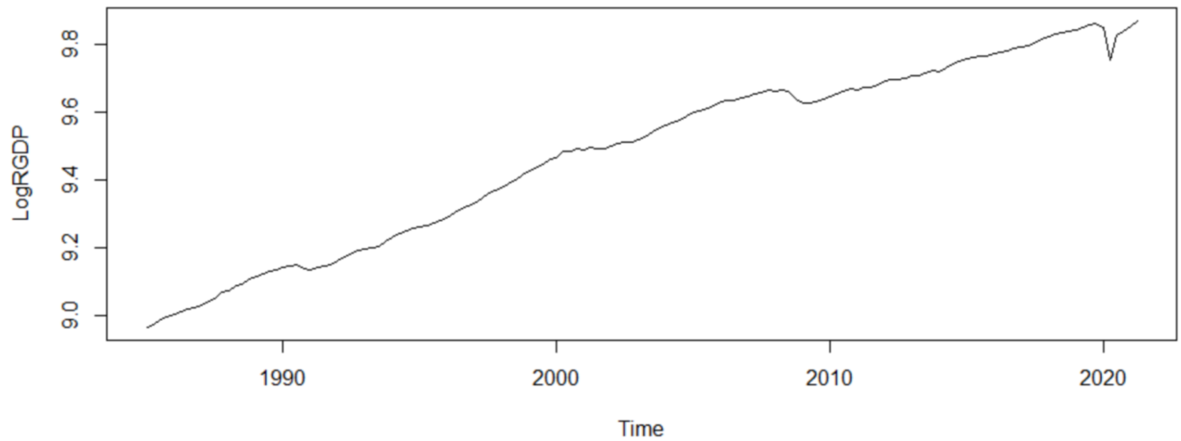
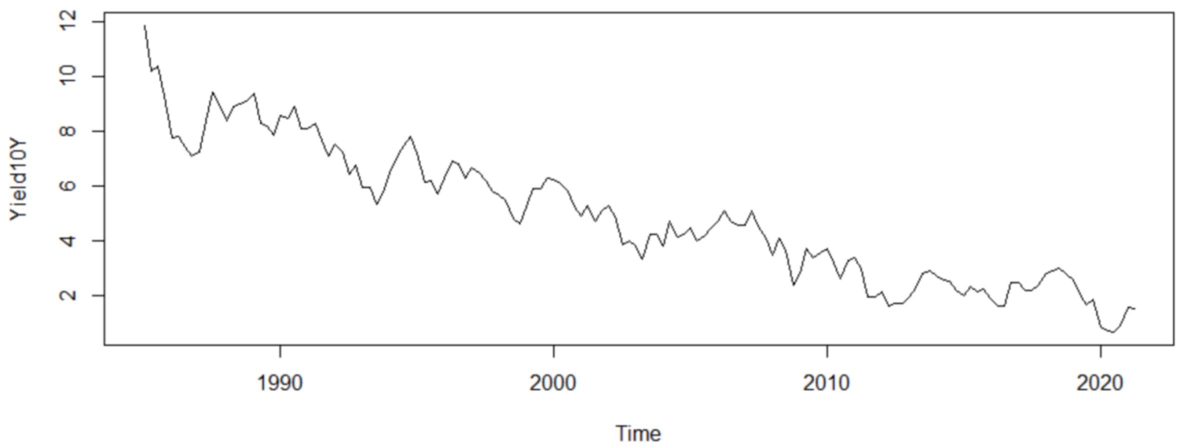


Chart 3

10-year US Treasury bond yield



2.1. Stationarity tests – proving the series are I(1):

(i) As confirmed by the below Augmented Dickey-Fuller test outputs, the series have a unit root, and therefore have no mean-reverting behavior.

Table 1

ADF test output for the logarithm of the S&P500

```
> adf.test(LogSP500)
Augmented Dickey-Fuller Test
alternative: stationary

Type 1: no drift no trend
     lag  ADF p.value
[1,]   0 3.67    0.99
[2,]   1 3.23    0.99
[3,]   2 2.82    0.99
[4,]   3 2.49    0.99
[5,]   4 2.40    0.99
Type 2: with drift no trend
     lag    ADF p.value
[1,]   0 -0.896   0.733
[2,]   1 -0.957   0.712
[3,]   2 -0.847   0.750
[4,]   3 -0.927   0.722
[5,]   4 -0.707   0.799
Type 3: with drift and trend
     lag   ADF p.value
[1,]   0 -1.90   0.617
[2,]   1 -2.07   0.545
[3,]   2 -2.13   0.519
[4,]   3 -2.39   0.413
[5,]   4 -2.20   0.488
----
Note: in fact, p.value = 0.01 means p.value <= 0.01
```

We fail to reject H0 in each of the 3 cases, meaning that we fail to reject the existence of a unit root.

Table 2

ADF test output for the logarithm of US real GDP

```
> adf.test(LogRGDP)
Augmented Dickey-Fuller Test
alternative: stationary

Type 1: no drift no trend
     lag  ADF p.value
[1,]   0 6.42    0.99
[2,]   1 6.38    0.99
[3,]   2 5.49    0.99
[4,]   3 4.86    0.99
[5,]   4 4.56    0.99
Type 2: with drift no trend
     lag   ADF p.value
[1,]   0 -1.79   0.412
[2,]   1 -1.98   0.336
[3,]   2 -1.87   0.377
[4,]   3 -1.85   0.386
[5,]   4 -2.06   0.304
Type 3: with drift and trend
     lag   ADF p.value
[1,]   0 -1.62   0.734
[2,]   1 -1.39   0.830
[3,]   2 -1.31   0.865
[4,]   3 -1.29   0.870
[5,]   4 -1.01   0.934
----
Note: in fact, p.value = 0.01 means p.value <= 0.01
```

We fail to reject H0 in each of the 3 cases, meaning that we fail to reject the existence of a unit root.

Table 3

ADF test output for the 10-year US Treasury bond yield

```
> adf.test(Yield10Y)
Augmented Dickey-Fuller Test
alternative: stationary

Type 1: no drift no trend
     lag   ADF p.value
[1,]   0 -2.64  0.0100
[2,]   1 -2.07  0.0394
[3,]   2 -2.31  0.0222
[4,]   3 -1.83  0.0676
[5,]   4 -1.73  0.0826
Type 2: with drift no trend
     lag   ADF p.value
[1,]   0 -2.53   0.124
[2,]   1 -1.84   0.389
[3,]   2 -1.99   0.333
[4,]   3 -1.51   0.517
[5,]   4 -1.05   0.678
Type 3: with drift and trend
     lag   ADF p.value
[1,]   0 -5.45    0.01
[2,]   1 -4.94    0.01
[3,]   2 -5.19    0.01
[4,]   3 -5.23    0.01
[5,]   4 -4.60    0.01
----
Note: in fact, p.value = 0.01 means p.value <= 0.01
```

We fail to reject H0 in the “with drift and no trend” case, meaning that we fail to reject the non-stationarity null hypothesis.

We confirm that the series has a drift with a complementary Augmented Dickey-Fuller test:

Table 4

ADF test output for a drift in the 10-year US Treasury bond yield

```
###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression drift


Call:
lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)

Residuals:
     Min       1Q   Median       3Q      Max
-1.25284 -0.34157 -0.00817  0.30696  1.27874

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.09749    0.09392   1.038   0.3010
z.lag.1     -0.03174    0.01721  -1.844   0.0673 .
z.diff.lag   0.02770    0.08016   0.345   0.7302
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4907 on 141 degrees of freedom
Multiple R-squared:  0.02433,   Adjusted R-squared:  0.01049
F-statistic: 1.758 on 2 and 141 DF,  p-value: 0.1761


Value of test-statistic is: -1.8437 2.6881

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.46 -2.88 -2.57
phi1  6.52  4.63  3.81
```

We fail to reject the null hypothesis of non-stationarity (as the test-statistic -1.84 is not negative enough to exceed the 10% significance level critical value of -2.57), and of the presence of a drift (the test statistic of 2.68 is smaller than the 10% significance level critical value of 3.81), meaning that we fail to reject the existence of a drift.

We therefore demonstrated that those series are not stationary.

(ii) We now show that more specifically, they are integrated of order 1.

Table 5

ADF test output for the first difference of the logarithm of the S&P500

```
> adf.test(Diff_SP500_TS)
Augmented Dickey-Fuller Test
alternative: stationary

Type 1: no drift no trend
     lag    ADF p.value
[1,]   0 -10.27    0.01
[2,]   1  -6.70    0.01
[3,]   2  -4.98    0.01
[4,]   3  -4.42    0.01
[5,]   4  -4.34    0.01
Type 2: with drift no trend
     lag    ADF p.value
[1,]   0 -11.17    0.01
[2,]   1  -7.51    0.01
[3,]   2  -5.76    0.01
[4,]   3  -5.16    0.01
[5,]   4  -5.10    0.01
Type 3: with drift and trend
     lag    ADF p.value
[1,]   0 -11.20    0.01
[2,]   1  -7.52    0.01
[3,]   2  -5.78    0.01
[4,]   3  -5.13    0.01
[5,]   4  -5.05    0.01
----
Note: in fact, p.value = 0.01 means p.value <= 0.01
```

We reject the null hypothesis of the presence of a unit root at the 1% significance level in each case, meaning that we reject the existence of a unit root for the first-differenced series. The log(S&P500) series therefore need to be differentiated once in order to be made stationary. It is therefore I(1).

Table 6

ADF test output for the first difference of the logarithm of the US real GDP

```
> adf.test(Diff_RGDP_TS)
Augmented Dickey-Fuller Test
alternative: stationary

Type 1: no drift no trend
     lag    ADF p.value
[1,]   0 -10.56  0.0100
[2,]   1  -6.22  0.0100
[3,]   2  -4.47  0.0100
[4,]   3  -3.31  0.0100
[5,]   4  -2.56  0.0116
Type 2: with drift no trend
     lag    ADF p.value
[1,]   0 -13.60    0.01
[2,]   1  -8.83    0.01
[3,]   2  -6.91    0.01
[4,]   3  -5.85    0.01
[5,]   4  -4.66    0.01
Type 3: with drift and trend
     lag    ADF p.value
[1,]   0 -13.82    0.01
[2,]   1  -9.04    0.01
[3,]   2  -7.14    0.01
[4,]   3  -6.17    0.01
[5,]   4  -5.05    0.01
----
Note: in fact, p.value = 0.01 means p.value <= 0.01
```

We reject the null hypothesis of the presence of a unit root at the 5% significance level in each case, meaning that we reject the existence of a unit root for the first-differenced series. The log(US real GDP) series therefore need to be differentiated once in order to be made stationary. It is therefore I(1).

Table 7

ADF test output for the first difference of US Treasury 10-year bond yield

```
> adf.test(Diff_RF_TS)
Augmented Dickey-Fuller Test
alternative: stationary

Type 1: no drift no trend
     lag    ADF p.value
[1,]   0 -11.91    0.01
[2,]   1  -8.95    0.01
[3,]   2  -6.91    0.01
[4,]   3  -7.09    0.01
[5,]   4  -6.85    0.01
Type 2: with drift no trend
     lag    ADF p.value
[1,]   0 -12.03    0.01
[2,]   1  -9.13    0.01
[3,]   2  -7.04    0.01
[4,]   3  -7.23    0.01
[5,]   4  -7.09    0.01
Type 3: with drift and trend
     lag    ADF p.value
[1,]   0 -12.00    0.01
[2,]   1  -9.14    0.01
[3,]   2  -7.01    0.01
[4,]   3  -7.19    0.01
[5,]   4  -7.06    0.01
----
Note: in fact, p.value = 0.01 means p.value <= 0.01
```

We reject the null hypothesis of the presence of a unit root at the 1% significance level in each case, meaning that we reject the existence of a unit root for the first-differenced series. The series therefore need to be differentiated once in order to be made stationary. It is therefore I(1).

As we proved that those series are I(1), we therefore aim at finding a relation of cointegration among those variables, according to the specification in Equation 3, deduced from the Gordon-Shapiro formula.

2.2. Testing for cointegration:

We start by examining whether there exists a linear combination of those three I(1) series that stationary (or I(0)) by running the following regression,

Equation 4:

$$\boldsymbol{log(P_t) = \ log(D_t) + log({r^F}_t) + \ Z_t}$$

where residuals $\boldsymbol{Z_t}$ must be proven stationary.

Table 8

Equation 4 estimation – regression output

```
> summary(lm1)

Call:
lm(formula = LogSP500 ~ LogRGDP + Yield10Y)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.52123 -0.13379 -0.02488  0.08235  0.47250 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -22.00393    1.92968 -11.403   <2e-16 ***
LogRGDP       3.03011    0.19322  15.682   <2e-16 ***
Yield10Y      0.02378    0.02100   1.132    0.259    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2041 on 143 degrees of freedom
Multiple R-squared:  0.9324,	Adjusted R-squared:  0.9314 
F-statistic:   986 on 2 and 143 DF,  p-value: < 2.2e-16
```

This regression's residuals are stationary, as shown below:

Table 9

ADF test for the residuals of equation 4's regression

```
###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression none


Call:
lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)

Residuals:
      Min        1Q    Median        3Q       Max
-0.273391 -0.025978  0.007201  0.044137  0.190439

Coefficients:
           Estimate Std. Error t value Pr(>|t|)
z.lag.1    -0.04668    0.02751  -1.697   0.0919 .
z.diff.lag  0.03709    0.08469   0.438   0.6621
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06491 on 142 degrees of freedom
Multiple R-squared:  0.02002,   Adjusted R-squared:  0.006213
F-statistic:  1.45 on 2 and 142 DF,  p-value: 0.238


Value of test-statistic is: -1.6967

Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62
```

With a value of -1.697, the t-statistic for "Z.lag.1" (which is the coefficient $\gamma = a - 1$ in front of the lag variable X_{t-1}) exceeds the critical value at the 10% level (of -1.62).

This therefore means that the coefficient $\gamma = a - 1$ is statistically different from 0, which implies that we can reject that the value of a = 1 at the 10% significance level.

We therefore reject the null hypothesis H0 of the existence of a unit root. The residuals of the regression are therefore stationary: we proved that there exists a stationary linear combination of I(1) series, meaning that the three series have a relationship of cointegration.

We confirm our result by testing the number of differences necessary to make the residuals stationary:

Table 10

Number of differences necessary to make residuals stationary

```
> ndiffs(resid)
[1] 0
```

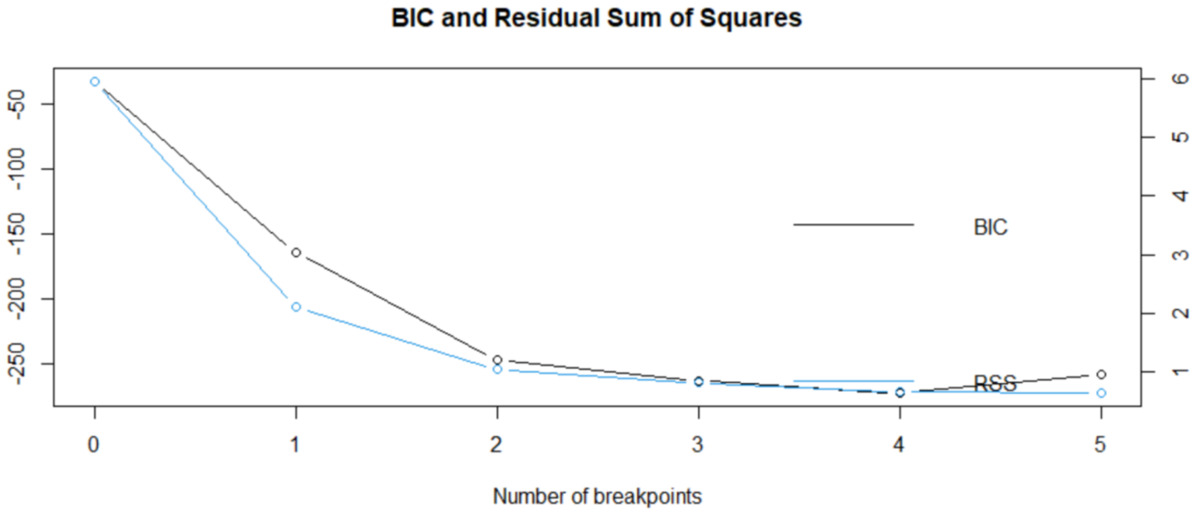
The residuals require no differentiation for being stationary, confirming their I(0) characteristic.

We now add dummy variables to this estimate of a long-term equation describing the S&P500 equilibrium price, denoted $\boldsymbol{DUM_t}$. Those dummy variables account for structural breaks in expectations (as to inflation, earnings growth, value added distribution, etc.) captured by the $\boldsymbol{\Pi^e{}_t - g^e{}_t}$ variable.

Those changes in expectations "may occur infrequently but have persistent effects on the dynamics of the stock price" (Bruneau et al., 2000). We therefore model $\boldsymbol{\Pi^e{}_t - g^e{}_t}$ as constant, allowing for some infrequent changes in level: the dummies take the values of 1 after a given structural break.

Chart 4

Optimal number of breakpoints



The optimal number of break points in the regression relationship is 4.

Table 11

Break dates

```
> breakpoints(Breaks)

	 Optimal 5-segment partition: 

Call:
breakpoints.breakpointsfull(obj = Breaks)

Breakpoints at observation number:
44 65 90 114 

Corresponding to breakdates:
1995(4) 2001(1) 2007(2) 2013(2) 
```

Those break points correspond to the following dates, where a structural break occurs in the regression relationship between our three I(1) series:

- Q4 1995;
- Q1 2001;
- Q2 2007;
- Q2 2013.

We include a dummy accounting for each in the below estimation of the long-term equilibrium equation – although this equation is unlikely to fit the data, it can be considered as the long-term target relationship.

Equation 5:

$$\boldsymbol{log(P_t) = \ log(D_t) - log\left(r^F{}_t + \Pi^e{}_t - g^e{}_t\right) + \ Z_t}$$

Where:

$\boldsymbol{Z_t}$ = the “stationary error-correcting variable” (Bruneau et al., 2000);

$\boldsymbol{\Pi^e{}_t - g^e{}_t}$ is accounted for with the structural breaks in expectations dummy $\boldsymbol{DUM_t}$ variables.

We therefore estimate the following equation and prove that Z is stationary.

Equation 6:

$$\boldsymbol{Z_t = \ log(P_t) - \ log(D_t) - log\left(r^F{}_t + \Pi^e{}_t - g^e{}_t\right)}$$

Table 12

Equation 5 – Regression output

```
Call:
lm(formula = LogSP500 ~ LogRGDP + Yield10Y + breakfactor(Breaks,
    breaks = 4))

Residuals:
     Min       1Q   Median       3Q      Max 
-0.27393 -0.06372 -0.02025  0.05106  0.35980 

Coefficients:
                                        Estimate Std. Error
(Intercept)                            -24.75088    1.56027
LogRGDP                                  3.36430    0.16524
Yield10Y                                -0.02035    0.01138
breakfactor(Breaks, breaks = 4)segment2  0.19197    0.04648
breakfactor(Breaks, breaks = 4)segment3 -0.31717    0.06582
breakfactor(Breaks, breaks = 4)segment4 -0.59364    0.07864
breakfactor(Breaks, breaks = 4)segment5 -0.36288    0.09488
                                        t value Pr(>|t|)    
(Intercept)                             -15.863  < 2e-16 ***
LogRGDP                                  20.360  < 2e-16 ***
Yield10Y                                 -1.789 0.075796 .  
breakfactor(Breaks, breaks = 4)segment2   4.130 6.23e-05 ***
breakfactor(Breaks, breaks = 4)segment3  -4.818 3.74e-06 ***
breakfactor(Breaks, breaks = 4)segment4  -7.549 5.25e-12 ***
breakfactor(Breaks, breaks = 4)segment5  -3.825 0.000197 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1057 on 139 degrees of freedom
Multiple R-squared:  0.9824,	Adjusted R-squared:  0.9816 
F-statistic:  1292 on 6 and 139 DF,  p-value: < 2.2e-16
```

This regression's residuals are stationary, as shown below:

Table 13

ADF test for the residuals of equation 5's regression

```
###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression none


Call:
lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)

Residuals:
     Min       1Q   Median       3Q      Max
-0.22996 -0.03020 -0.00078  0.03266  0.39609

Coefficients:
           Estimate Std. Error t value Pr(>|t|)
z.lag.1    -0.32158    0.06757  -4.759 4.73e-06 ***
z.diff.lag  0.10188    0.08543   1.193    0.235
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.07475 on 142 degrees of freedom
Multiple R-squared:  0.1414,    Adjusted R-squared:  0.1293
F-statistic: 11.69 on 2 and 142 DF,  p-value: 1.993e-05


Value of test-statistic is: -4.7593

Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62
```

With a value of -4.76, the t-statistic for "Z.lag.1" (which is the coefficient $\gamma = a - 1$ in front of the lag variable X_{t-1}) exceeds the critical value at the 1% level of significance (of -2.58).

This therefore means that the coefficient $\gamma = a - 1$ is statistically different from 0, which implies that we can reject that the value of a = 1 at the 1% significance level.

We therefore reject the null hypothesis H0 of the existence of a unit root. The residuals of the regression are therefore stationary. We therefore proved that there exists a stationary linear combination of I(1) series: the 3 series have a relationship of cointegration, with 4 structural breaks.

We confirm our result by testing the number of differences necessary to make the residuals stationary:

Table 14

Number of differences necessary to make residuals stationary

```
> ndiffs(resid2)
[1] 0
```

We therefore proved the existence of a relation of cointegration between the S&P500 index and its fundamentals, as we have demonstrated that there exists a linear combination of those I(1) series that is covariance and mean stationary, or I(0).
In other words, we have justified the existence of a long-term equation to describe the equilibrium price with four structural breaks.

We further confirm this result with a Trace test (**Table 15**) and then Maximum-eigenvalue test (**Table 16**), which includes our three I(1) series as well as our dummy variables as exogenous variables.

Table 15

Trace cointegration test output

Variables included: Log (SP500), Log (Real GDP), 10-year US Treasury yield, structural break dates dummy variables

```
> Trace <- ca.jo(LogData2, type = "trace", ecdet = "const", K = 2)
> summary(Trace)

######################
# Johansen-Procedure #
######################

Test type: trace statistic , without linear trend and constant in
 cointegration

Eigenvalues (lambda):
[1]  3.129395e-01  1.534029e-01  6.761981e-02  2.279559e-02
[5] -3.962326e-16

Values of teststatistic and critical values of test:

          test 10pct  5pct  1pct
r <= 3 |  3.32  7.52  9.24 12.97
r <= 2 | 13.40 17.85 19.96 24.60
r <= 1 | 37.38 32.00 34.91 41.07
r = 0  | 91.43 49.65 53.12 60.16
```

The test statistic (37.38) exceeds the critical values at the 5% significance level that there exists at most 1 relationship of cointegration (34.91) but fails to reject that there exist as most 2 relationships of cointegration (13.40 < 17.85).

We therefore conclude that there exist 2 relationships of cointegration in our system

Table 16

Maximum-eigenvalue cointegration test output

Variables included: Log (SP500), Log (Real GDP), 10-year US Treasury yield, structural break dates dummy variables

```
######################
# Johansen-Procedure #
######################

Test type: maximal eigenvalue statistic (lambda max) , without linear trend and constant
 in cointegration

Eigenvalues (lambda):
[1]  3.129395e-01  1.534029e-01  6.761981e-02  2.279559e-02 -3.962326e-16

Values of teststatistic and critical values of test:

          test 10pct  5pct  1pct
r <= 3 |  3.32  7.52  9.24 12.97
r <= 2 | 10.08 13.75 15.67 20.20
r <= 1 | 23.98 19.77 22.00 26.81
r = 0  | 54.05 25.56 28.14 33.24
```

The test statistic (23.98) exceeds the critical values at the 5% significance level that there exists at most 1 relationship of cointegration (22.00) but fails to reject that there exist as most 2 relationships of cointegration (10.08 < 13.75).

This result confirms that there exist 2 relationships of cointegration in our system.

2.3. Estimating the corresponding error-correction equation:

Consequently, we can deduce the error-correcting equation describing this long-term relationship by estimating a Vector Error-Correcting Model, with 8 lags:

Table 17

Optimal number of lags

```
> lagselect$selection
AIC(n)  HQ(n)  SC(n) FPE(n)
     8      6      6      8
```

We choose to integrate 8 lags in our model following the Akaike Information Criteria (AIC) and Final Prediction Error (FPE), as those "minimize the chance of under estimation while maximizing the chance of recovering the true lag length" (Liew, 2004)[4].

[4] V. Kim-Shen Liew, *Which lag selection criteria should we employ?*, January 2004, Economics Bulletin

Table 18

VECM model estimation output

Variables included: Log (SP500), Log (Real GDP), 10-year US Treasury yield, structural break dates dummy variables

8 lags, 2 relationships of cointegration

```
#############
###Model VECM
#############
Full sample size: 146   End sample size: 137
Number of variables: 3  Number of estimated slope parameters 81
AIC -2136.623   BIC -1894.264   SSR 22.59098
Cointegrating vector (estimated by ML):
        LogSP500 LogRGDP  Yield10Y      const
r1  1.000000e+00       0 0.3590560  -7.560727
r2 -1.387779e-17       1 0.1235555 -10.183080

                  ECT1                 ECT2
Equation LogSP500 0.0127(0.0189)       -0.3540(0.1169)**
Equation LogRGDP  0.0116(0.0031)***    -0.0792(0.0193)***
Equation Yield10Y -0.2564(0.1262)*     -1.2903(0.7791)
                  LogSP500 -1           LogRGDP -1
Equation LogSP500 0.0363(0.0979)        -0.1022(0.5738)
Equation LogRGDP  0.0022(0.0162)        -0.2811(0.0949)**
Equation Yield10Y -0.0049(0.6524)       -0.0523(3.8239)
                  Yield10Y -1          LogSP500 -2
Equation LogSP500 0.0401(0.0170)*      0.1129(0.0977)
Equation LogRGDP  0.0116(0.0028)***    0.0210(0.0162)
Equation Yield10Y 0.1197(0.1132)       -0.6063(0.6510)
                  LogRGDP -2            Yield10Y -2
Equation LogSP500 0.3641(0.5912)        0.0155(0.0176)
Equation LogRGDP  -0.1489(0.0977)       0.0031(0.0029)
Equation Yield10Y 3.5129(3.9400)        0.0182(0.1176)
                  LogSP500 -3          LogRGDP -3
Equation LogSP500 0.1050(0.1011)       -0.6668(0.6113)
Equation LogRGDP  0.0260(0.0167)       -0.2191(0.1011)*
Equation Yield10Y 0.5385(0.6741)       -2.3571(4.0741)
                  Yield10Y -3          LogSP500 -4
Equation LogSP500 0.0128(0.0168)       -0.0738(0.1013)
Equation LogRGDP  0.0059(0.0028)*      -0.0119(0.0168)
Equation Yield10Y 0.0422(0.1117)       0.4654(0.6752)
                  LogRGDP -4            Yield10Y -4
Equation LogSP500 -1.3395(0.7276).      0.0540(0.0160)***
Equation LogRGDP  -0.2105(0.1203).      0.0073(0.0026)**
Equation Yield10Y -2.7771(4.8494)       0.0306(0.1063)
                  LogSP500 -5           LogRGDP -5
Equation LogSP500 -0.0556(0.1007)       0.7093(1.4540)
Equation LogRGDP  0.0087(0.0166)        0.1281(0.2404)
Equation Yield10Y -0.2066(0.6709)       -2.6844(9.6901)
                  Yield10Y -5           LogSP500 -6
Equation LogSP500 0.0377(0.0157)*       0.0935(0.0994)
Equation LogRGDP  0.0037(0.0026)        0.0021(0.0164)
Equation Yield10Y -0.0420(0.1046)       1.0795(0.6627)
                  LogRGDP -6           Yield10Y -6
Equation LogSP500 0.2431(1.4441)       0.0080(0.0157)
Equation LogRGDP  0.1498(0.2388)       0.0040(0.0026)
Equation Yield10Y 17.9700(9.6244).     -0.0608(0.1045)
                  LogSP500 -7           LogRGDP -7
Equation LogSP500 -0.1241(0.1003)       1.5224(1.3893)
Equation LogRGDP  -0.0258(0.0166)       -0.0683(0.2297)
Equation Yield10Y -1.0310(0.6682)       17.2823(9.2592).
                  Yield10Y -7           LogSP500 -8
Equation LogSP500 0.0043(0.0152)        -0.0055(0.1019)
Equation LogRGDP  0.0010(0.0025)        0.0166(0.0169)
Equation Yield10Y -0.0471(0.1015)       -1.3844(0.6792)*
                  LogRGDP -8            Yield10Y -8
Equation LogSP500 -0.8703(1.3381)       0.0310(0.0141)*
Equation LogRGDP  -0.2657(0.2212)       0.0016(0.0023)
Equation Yield10Y 0.0729(8.9176)        -0.0221(0.0940)
                  exo_1
Equation LogSP500 -0.0063(0.0045)
Equation LogRGDP  -0.0023(0.0007)**
Equation Yield10Y -0.0247(0.0297)
```

Armed with this estimation of the S&P500 long-run target value, we can therefore use this reliable fundamental value of the S&P500 stock index to forecast future S&P500 price levels.

Chart 5

S&P500 price levels forecasting (3 years ahead)