UNIVERSITY OF PARIS 1 PANTHEON SORBONNE

MASTER 2 FINANCE TECHNOLOGY DATA

2022-2023

Asset Pricing

Empirical Application 3

Fundamental Analysis of a stock index

Group Members

- [redacted]
- [redacted]
- [redacted]
- Julien DOS REIS - FTD

Data

We propose the analysis of the UK stock index FTSE. As we consider a large index for an entire country, we use as fundamentals the UK GDP and the UK 10 Year Treasury Bond. As our data has quarterly frequency we decided to take a longer period of time in order to have enough observations (around 130). We analyze the time series over a 32 years period between 1984 to 2016, a period which has known different events which have impacted the UK economy. In fact, Real GDP in the UK has typically increased every year, although there have been three downturns in the economy since 1980. The choice of this time period is justified by our aim to analyze the evolution of the stock observed stock price compared to the fundamental (fair) price, while we consider the potential impact of the financial crisis.

The data source for the FTSE and for the Treasury Bond is FRED and for the UK GDP we import the data from datahub.io. We take the logarithm of both the Stock Price index and of the GDP, but not the Treasury Bond as it is already a rate. We can plot our time series.

Figure 1. FTSE Stock Price Index time series

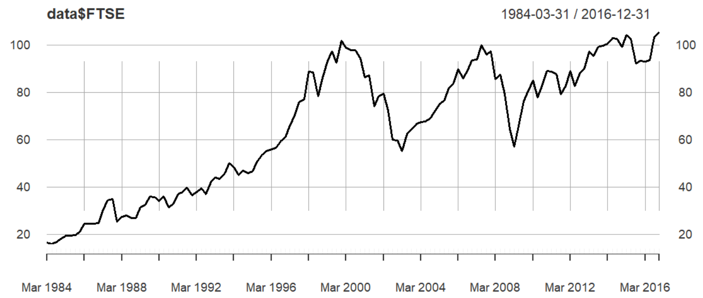


Figure 2. UK GDP time series

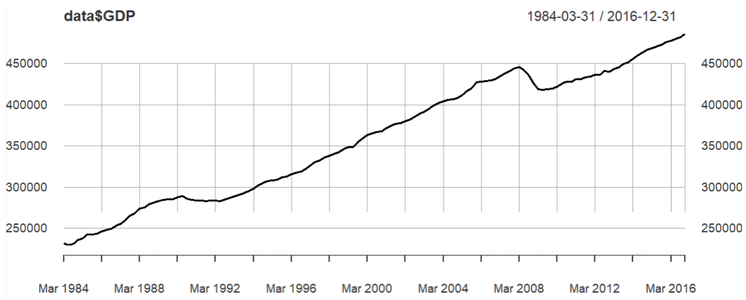
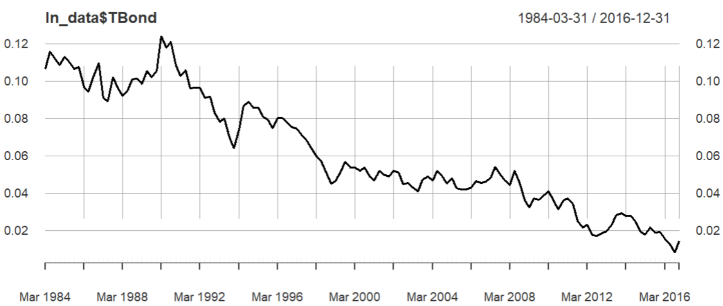


Figure 3. UK 10 year Treasury Bond



PART I: STATIONARITY

In order to check for stationarity of the time series we perform an Augmented Dickey Fuller test, using the function adf.test() in R, allowing us to test the following hypothesis:

H0: the series is not stationary

H1: the series is stationary

Figure 4. Augmented Dickey Fuller test results

```
> adf.test(ln_data$FTSE)

	Augmented Dickey-Fuller Test

data:  ln_data$FTSE
Dickey-Fuller = -2.3356, Lag order = 5, p-value = 0.4369
alternative hypothesis: stationary

> adf.test(ln_data$GDP)

	Augmented Dickey-Fuller Test

data:  ln_data$GDP
Dickey-Fuller = -1.8187, Lag order = 5, p-value = 0.6519
alternative hypothesis: stationary

> adf.test(ln_data$TBond)

	Augmented Dickey-Fuller Test

data:  ln_data$TBond
Dickey-Fuller = -3.092, Lag order = 5, p-value = 0.1224
alternative hypothesis: stationary
```

As you can see in Figure 4, the p-values of the test for the 3 time series Log FTSE, Log GDP and 10 Year Treasury Bond are above the 5% significance level. (0.4369, 0.6519 and 0.1224 respectively) Therefore we fail to reject the null hypothesis and we conclude that all 3 series are not stationary.

To render them stationary we need to calculate their differentiation, and test again for the unit root.

```
	Augmented Dickey-Fuller Test

data:  na.omit(diff_FTSE)
Dickey-Fuller = -4.5518, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary
```

```
	Augmented Dickey-Fuller Test

data:  diff_GDP
Dickey-Fuller = -3.9542, Lag order = 4, p-value = 0.01394
alternative hypothesis: stationary

> adf.test(na.omit(diff_tbond))

	Augmented Dickey-Fuller Test

data:  na.omit(diff_tbond)
Dickey-Fuller = -5.2259, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary
```

Now that we confirm that our 3 series are I(1), we can test for cointegration by applying the Johanson test.

PART II : VECM

```
####testing for cointegration
coint<-lm(ln_data$FTSE~ln_data$GDP+ln_data$TBond)
summary(coint)
adf.test(coint$residuals)
resur<-ur.df(coint$residu, type =  "none", lags = 1, selectlags = "BIC" )
summary(resur)
```

Here we run a linear regression between FTSE, GDP and Tbond then we extract the residuals to see whether they are stationary.

```
Residual standard error: 0.07479 on 112 degrees of freedom
Multiple R-squared:  0.08936,   Adjusted R-squared:  0.07309
F-statistic: 5.495 on 2 and 112 DF,  p-value: 0.005291


Value of test-statistic is: -3.0926

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

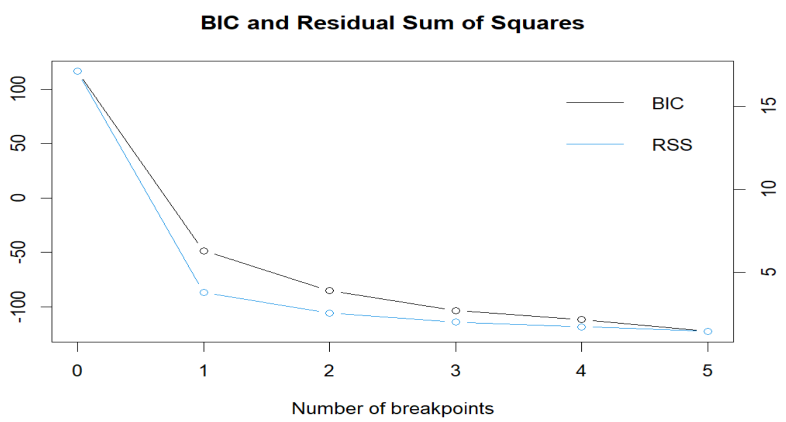
the t-stat -3.0926 < -1.95 we reject the null hypothesis, our residuals are I(0), so we confirm the existence of a cointegration relation between the 3 series.

Now, we are going to introduce the structural breaks, or what we call the exogenous factor in order to find the fair price P of our stock index FTSE.

Structural Breaks

We can identify the structural breakpoints of the Logarithm of the FTSE Stock Index time series, by applying the breakpoints() function in R. The summary of this function gives us the possible number of breakpoints (m) as well as their corresponding observation number and date. We have to choose their number based on the BIC criterion. We can see in **Figure 6** that the lowest BIC () is the one of m= 5. Therefore we look at the 5 breakpoints corresponding to observations number 19,37, 54, 71, 99.

Figure 6. Choosing Number of Structural Breakpoints Plot



Let's look at the first break point which correspond to the date of 30 september 1992 when we search on what has happened in this year and more precisely in this month we find the "Black Wednesday" and in which the government had suspended the UK's membership of the European Exchange Rate Mechanism following a wave of speculation against the Pound.

Now we are going to create a first dummy variable for this break (0 before the breakpoint and 1 after the breakpoint) in order to introduce it in our model and then we are going to test for the stationary of the residuals that we call Z, if we find that Z is I(0) we stop on this breakpoint and we confirm the existence of cointegration relation, otherwise we add a second break point (a second dummy variable) and so on.

first break : 30 september 1992

```
> test1 <- lm(ln_data$FTSE~ln_data$GDP + ln_data$TBond +
ln_data$Break1992)
> summary(test1)
```

```
> adf.test(test1$residuals)

        Augmented Dickey-Fuller Test

data:  test1$residuals
Dickey-Fuller = -2.9797, Lag order = 4, p-value
= 0.1699
alternative hypothesis: stationary
```

As we can see our p-value is greater than 0.05 so we fail to reject the null hypothesis that our residuals are not stationary. Now we move to the second breakpoint which is **31 march 1997**. When we check what happened in this quarter in the UK economy we find that in this period unemployment was continuing to fall and had reached over 1,800,000 – its lowest level since December 1990.

Let's now introduce our second dummy variable in our model and check the stationarity of the residuals again.

```
> test2 <- lm(ln_data$FTSE~ln_data$GDP + ln_data$TBond + ln_data$Break1992 + ln_data$Break1997)
```

```
        Augmented Dickey-Fuller Test

data:  test2$residuals
Dickey-Fuller = -3.1008, Lag order = 4, p-value
= 0.1196
alternative hypothesis: stationary
```

The ADF test indicates again that our residuals are not stationary therefore we will introduce the 3rd breakpoint, which is 2001-06-30, in this period we found that Blair's Labor party wins a second successive general election victory and apparently this party had a toxic effect on the UK economy.

Now we introduce the 3rd dummy variable in our model and we test again the stationarity of the residuals.

```
lm(formula = ln_data$FTSE ~ ln_data$GDP + ln_data$TBond + ln_data$Break1992 +
    ln_data$Break1997 + ln_data$Break2001)
```

```
        Augmented Dickey-Fuller Test

data:  test3$residuals
Dickey-Fuller = -4.963, Lag order = 4, p-value =
0.01
alternative hypothesis: stationary
```

The ADF test this time indicates that our Z is now stationary I(0), this leads us to say that there is a cointegration relation in our model which can allow us to use the VECM model, but before we need to know how many cointegration relations exist.

To sum up we created 3 dummy variables for each of the breakpoints (0 before the breakpoint and 1 after the breakpoint) in order to later use them in our models to obtain stationary residuals.

Figure 7. Dummy Variables based on structural breakpoints

	FTSE	GDP	TBond	Break1992	Break1997	Break2001
1995-06-30	3.928313	12.64165	0.081097	1	0	0
1995-09-30	3.981397	12.65144	0.079486	1	0	0
1995-12-31	4.013398	12.65430	0.074926	1	0	0
1996-03-31	4.026547	12.66334	0.080570	1	0	0
1996-06-30	4.036376	12.66649	0.080540	1	0	0
1996-09-30	4.086790	12.67324	0.078151	1	0	0
1996-12-31	4.114859	12.68262	0.075716	1	0	0
1997-03-31	4.189015	12.69697	0.074544	1	1	0
1997-06-30	4.255627	12.70763	0.071418	1	1	0
1997-09-30	4.330397	12.71357	0.068216	1	1	0
1997-12-31	4.345683	12.72576	0.063739	1	1	0

Figure 8. Model including the three structural break

```
lm(formula = ln_data$FTSE ~ ln_data$GDP + ln_data$TBond + ln_data$Break1992 +
    ln_data$Break1997 + ln_data$Break2001)

Residuals:
     Min       1Q   Median       3Q      Max
-0.32387 -0.06572  0.01392  0.07093  0.33538

Coefficients:
                   Estimate Std. Error t value
(Intercept)       -23.91765    3.00685  -7.954
ln_data$GDP         2.17966    0.23341   9.339
ln_data$TBond       0.64809    1.14309   0.567
ln_data$Break1992   0.24316    0.03998   6.082
ln_data$Break1997   0.28744    0.04330   6.638
ln_data$Break2001  -0.48439    0.04719 -10.265
                  Pr(>|t|)
(Intercept)       1.73e-12 ***
ln_data$GDP       1.28e-15 ***
ln_data$TBond        0.572
ln_data$Break1992 1.75e-08 ***
ln_data$Break1997 1.25e-09 ***
ln_data$Break2001  < 2e-16 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1026 on 110 degrees of freedom
Multiple R-squared:  0.9323,    Adjusted R-squared:  0.9293
```

we can clearly see that our breakpoints are significant . So if we find out that there is a cointegration relationships the regression we've just found will represent our long run relation

In order to know the rank of cointegration we are going to perform the Johansen test, but before this we need at first identify the the lag we are going to use in our test and for the VECM, so for this we use the function VARselect which helps us to know the best lag for our model, the results are as follows:

```
dset <- cbind(ln_data$GDP,ln_data$FTSE,ln_data$TBond)
lagselect <- VARselect(dset, lag.max = 15, type = "const")
lagselect$selection
jotest <- ca.jo(dset,type='trace',ecdet='const',K=2)
summary(jotest)
```

```
AIC(n)  HQ(n)  SC(n) FPE(n)
     2      2      2      2
```

all the criterias define 2 lags as the best lag for our model.

By comparing the statistics of the test to the critical values at r=0, we see that 296.95is higher than 31.52 at 5% respectively. Thus, we reject H0 (H0: there is no cointegration).

Same for r <=1 we reject H0, which means we have at least 1 cointegration relationships.

For r <=2, we don't reject H0 since 2.96<8.18 at 5%. And thus, we have at most 1 cointegration relationship.

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

Test type: trace statistic , with linear trend

Eigenvalues (lambda):
[1] 0.88959015 0.67928517 0.03305936

Values of teststatistic and critical values of test:

           test 10pct  5pct  1pct
r <= 2 |   2.96  6.50  8.18 11.65
r <= 1 | 103.03 15.66 17.95 23.52
r = 0  | 296.95 28.71 31.52 37.22
```

now we run our VECM:

```
Full sample size: 132   End sample size: 129
Number of variables: 3  Number of estimated slope parameters 24
AIC -3378.952   BIC -3304.597   SSR 0.68373
Cointegrating vector (estimated by 2OLS):
   FTSE        GDP     TBond
r1    1 -0.3794367 12.45269


               ECT                Intercept          FTSE -1           GDP -1
Equation FTSE  -0.0752(0.0319)*   0.0093(0.0098)     0.0427(0.0906)    2.0727(1.2984)
Equation GDP   -0.0040(0.0022).   0.0021(0.0007)**   0.0052(0.0063)    0.4366(0.0896)***
Equation TBond -0.0028(0.0021)    -0.0016(0.0006)*   0.0111(0.0060).   -0.0411(0.0859)
               TBond -1           FTSE -2            GDP -2            TBond -2
Equation FTSE  -1.8747(1.3662)    -0.0108(0.0914)    -1.4801(1.2465)   0.0105(1.3209)
Equation GDP   0.0207(0.0943)     0.0131(0.0063)*    0.1660(0.0860).   -0.0746(0.0912)
Equation TBond 0.0082(0.0904)     0.0030(0.0060)     0.1265(0.0825)    -0.2464(0.0874)**
```

From this table we can see that there is a negative relation between FTSE and GDP (-0.37) . We know that Strong GDP performance indicates a strong economy, which can embolden investors. More activity in the markets can lead to share price gains, which then raises the major indexes. If GDP falls, investors have less confidence in the economy which can slow their trading activity.
However, contrarianism is a real phenomenon in the market. Some picks are made in complete opposition to economic signals. Investors can also buy during economic dips with the hope of finding discounted stocks that will recover when the economy picks back up. So we think that the coefficient sign reflects more the functioning of the market in real life.Investor sentiment is complex and sometimes investors ignore GDP altogether.
Now for the error correction term which represents the lagged value of the residuals obtained from the cointegration regression of the dependent variable on the regressors it contains long run cointegrating relationship capture the long run information
the lambda in our case is equal to -0.07 is the speed of adjustment it measures the speed at which Y returns to equilibrium after changs in gdp and risk free, the value of FTSE to equilibrium path will take a longer time because the value ECT is quite small.
According to the P value the stock price isn't really impacted by GDP and the risk free.
However, for the other equations, we notice that there is no error correction mechanism.

```
#######SCRIPT#######
library(pdfetch)
library(forecast)
library(tseries)
library(strucchange)
library("readxl")
library('urca')
library('TTR')
library('quantmod')
library('tibble')
library('ggplot2')
library('dplyr')
library('forecast')

## load the data
#save the gdp series excel file as csv and copy an paste the path here instead of the one below
GDPExcelPath <- "C:/Users/shadjsahraou/Downloads/gdp-uk.csv"
```

```
#import the FTSE data from FRED
ftsedata <- (pdfetch_FRED("SPASTT01GBM661N")["1988-01-01::2016-12-31"])
data<- ftsedata[endpoints(ftsedata,'quarters')]
names(data)[1]<- "FTSE"

#add the UK GDP to the timeseries
gdpdata<- read.csv(GDPExcelPath)
gdpdata <- data.frame(gdpdata)
data$GDP <- gdpdata[133:248,2]

# calculate the log
ln_data <- log(data)

#add the UK Treasury Bond
tbonddata <- (pdfetch_FRED("INTGSBGBM193N")["1988-01-01::2016-12-31"])/100
ln_data$TBond <- tbonddata[endpoints(tbonddata,'quarters')]

#perform stationarity tests
acf(ln_data$TBond)
acf(ln_data$FTSE)
acf(ln_data$GDP)

adf.test(ln_data$TBond)
adf.test(ln_data$FTSE)
adf.test(ln_data$GDP)
#differenciation to be sure tht our series are I(1)
diff_GDP <- diff(ln_data$GDP)
diff_GDP<- na.omit(diff_GDP)
diff_tbond<-na.omit(diff(ln_data$TBond))
diff_FTSE<-na.omit(diff(ln_data$FTSE))
adf.test(na.omit(diff_FTSE))
adf.test(diff_GDP)
adf.test(na.omit(diff_tbond))

#our 3 series are I(1)
####testing for cointegration
coint<-lm(ln_data$FTSE~ln_data$GDP+ln_data$TBond)
```

```
summary(coint)
adf.test(coint$residuals)
resur<-ur.df(coint$residu, type =  "none", lags = 1, selectlags = "BIC" )
summary(resur)
### my first test to see whther is there cointegration and we found tht our series are cointegrated
####now we want to know the number of cointegration relation
#install.packages('vars')
dset <- cbind(ln_data$GDP,ln_data$FTSE,ln_data$TBond)
lagselect <- VARselect(dset, lag.max = 30, type = "none")
lagselect$selection
jotest=ca.jo(dset, type = 'trace', ecdet = 'none', K = 28,spec="longrun")
summary(jotest)
Ejotest <- ca.jo(dset, type = 'eigen', ecdet = 'none', K = 28)
summary(Ejotest)
#finding the breakpoints
bp.FTSE<- breakpoints(ln_data$FTSE ~ 1)
summary(bp.FTSE)
plot(ln_data$FTSE)

#the BIC chooses 3 breakpoints at observations 20, 50, 112
#we identify the exact dates
ln_data[19,1]
ln_data[37,1]
ln_data[54,1]
ln_data[71,1]
ln_data[99,1]

## the BIC chooses3 breakpoints; plot the graph with breakdates and their confidence intervals
plot(bp.FTSE)
plot(ln_data$FTSE)
lines(bp.FTSE)

## confidence intervals
ci.FTSE <- confint(bp.FTSE)
ci.FTSE
lines(ci.FTSE)
```

```
#we add the breakpoints as a dummy variable to the data set
ln_data$Break1992 <- ifelse(index(ln_data) >= "1992-09-30", 1, 0)
ln_data$Break1997 <- ifelse(index(ln_data) >= "1997-03-31", 1, 0)
ln_data$Break2001 <- ifelse(index(ln_data) >= "2001-06-30", 1, 0)

# we test the stationarity of the residuals for different models

#model including first breakpoint in 1992
#the p-value is higher than 5% so it is not stationary and we need to include the 2nd point
test1 <- lm(ln_data$FTSE~ln_data$GDP + ln_data$TBond + ln_data$Break1992)
summary(test1)

adf.test(test1$residuals)
acf(test1$residuals)

#including 1997 break
test2 <- lm(ln_data$FTSE~ln_data$GDP + ln_data$TBond + ln_data$Break1992 +
ln_data$Break1997)
summary(test2)

adf.test(test2$residuals)
acf(test2$residuals)
##not stationary
#including the 2001 break
test3 <-lm(ln_data$FTSE~ln_data$GDP + ln_data$TBond + ln_data$Break1992 +
ln_data$Break1997+ln_data$Break2001)
summary(test3)
adf.test(test3$residuals)
###residuals are now stationary

#cointegration test
library(vars)
dset <- cbind(ln_data$FTSE, ln_data$GDP, ln_data$TBond)
lagselect <- VARselect(dset, lag.max = 15, type = "none")
lagselect$selection
dset <- cbind(ln_data$FTSE,ln_data$GDP,ln_data$TBond)
lagselect <- VARselect(dset, lag.max = 15, type = "const")
```

```
lagselect$selection
jotest <- ca.jo(dset,type='trace',ecdet='const',K=2)
summary(jotest)
Ejotest <- ca.jo(dset, type = 'eigen', ecdet = 'trend', K = 2)
summary(Ejotest)
dset <- cbind(diff_FTSE, diff_GDP,diff_tbond)

###VECM
vecm <- VECM(dset,lag=2,r=1,estim="2OLS")
summary(vecm)
```