# What the hell statistical arbitrage is?

Statistical arbitrage is the mispricing of any given security according to their expected value, base on the mathematical analysis of its historic valuations. Statistical arbitrage is often involved with pairs trading. A statistical arbitrage pairs trading position consists in a long position on one security and a short position on another security. In order to calculate the arbitrage opportunities, a formula have to be better than a simple correlation. Cointegration is this type of mathematical formula we want.

***In this article, we will take the EURUSD/GBPJPY for the calculations of spread between 5/1/2012 and 5/1/2013 with a period set later below.***

This article is not exhaustive about pairs trading and statistical arbitrage, but I want this one just to a short presentation. To go forward, you need to backtest several strategies and find good parameters for your models.

## 1. Definitions : cointegration and statistical arbitrage

### 1.1 Cointegration

**Two time series are cointegrated if the two series have the same stochastic drift.** I don't entre in much detail here, because of all mathematical components of cointegration and in general, it is far above the understanding of the average person. But still, I will give some sources to go further.

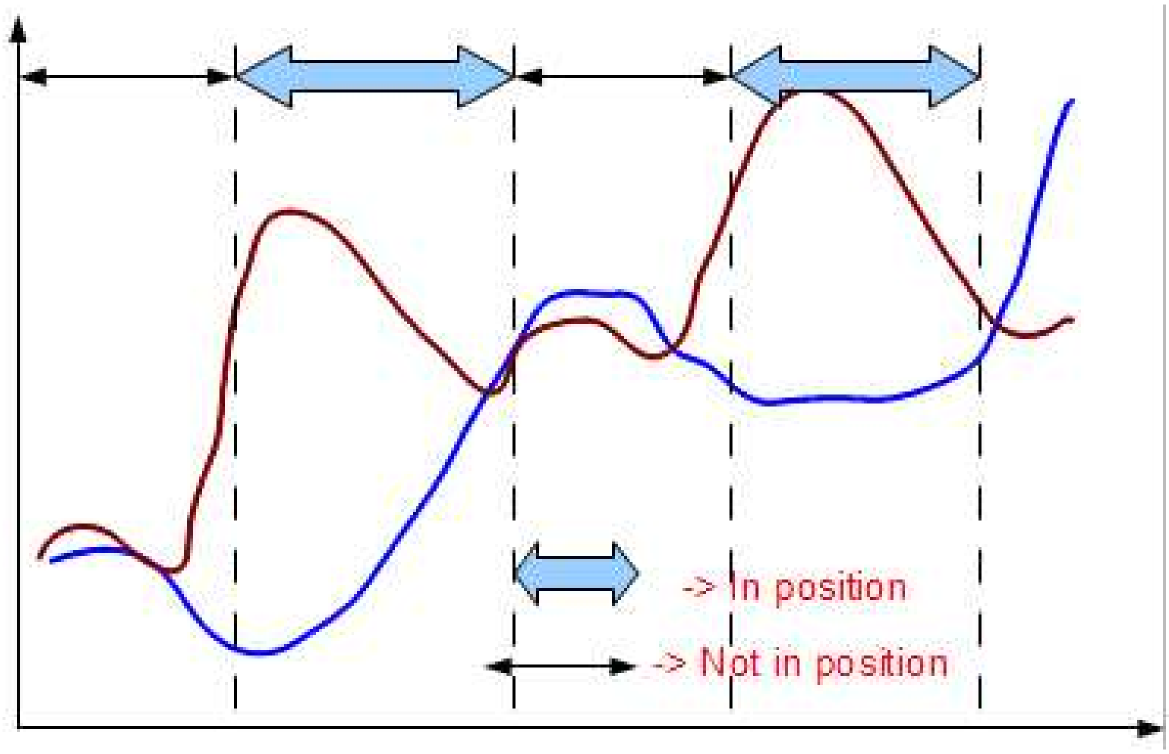

**Figure 1: Timing for a pairs trading position**

In the figure 1, the graphic gives a example of timing for a pairs trading position. We can say the cointegration measures the difference between two securities which are a cointegration relationship on the long run.

The main difference between correlation and cointegration is such as correlation implies a same movement between two securities, whereas cointegration implies a equilibrium relation.

To describe what statistical arbitrage is, let take a quote from "Statistical arbitrage, Algorithmic Trading Insights and Techniques" by Andrew Pole.

"

*Mean reversion in prices, as in much of human activity, is a powerful and fundamental force, driving systems and markets to homeostatic relationships. Starting in the early 1980s, statistical arbitrage was a formal and successful attempt to model this behavior in the pursuit of profit. Understanding the arithmetic of statistical arbitrage (sometimes abbreviated as stat. arb.) is a cornerstone to understanding the development of what has come to be known as complex financial engineering and risk modeling.*

...

*Statistical arbitrage describes the phenomena, the driving forces generating those phenomena, the patterns of dynamics development of exploitable opportunities, and models for exploitation of the basic reversion to the mean in securities prices. It also offers a good* *deal more, from hints at more sophisticated models to valuable practical advice on model building and performance monitoring - advice far beyond statistical arbitrage.*

"

I couldn't sum up a best definition of statistical arbitrage myself. The kind of statistical arbitrage, we are going to see, is on stock pairs and forex pairs. When I talk about pairs, it is pairs trading which is composed by two securities.

## 1.2 Statistical arbitrage and trading rules

We need to set up some trading rules, in order to trade the arbitrage opportunities. The trading rules are in fact quite simple.

The trading rules are the following:

- Open position when the ratio hits the 2 standard deviation band for two consecutive times.
- Close position when the ratio hits the mean.

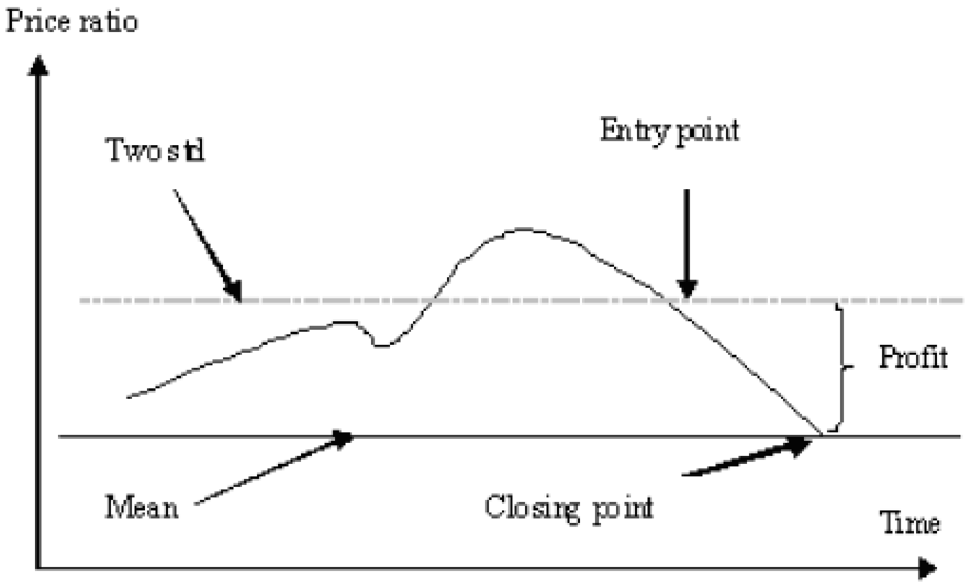

Figure 2: Pairs trading rules [1]

## 1.3 Exploitation of spread

Let's do an **ordinary least squares (OLS)** to calculate the spread for the trading distance models in the next part.

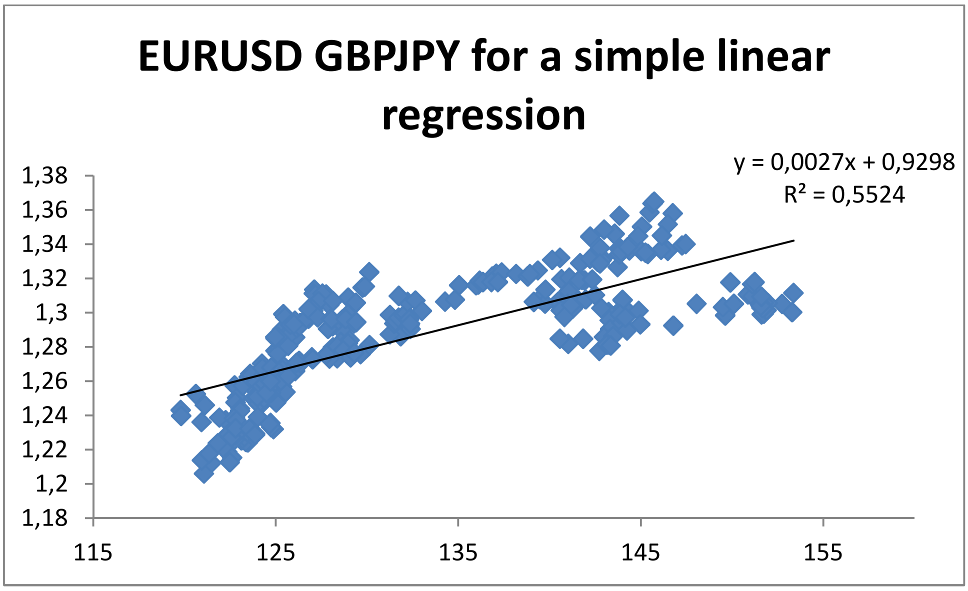

Figure 3: Simple linear regresion to find the beta ratio

Here are the equations to draw the spread between two securities.

$$Spread = A - n\,B \;(1)$$

$$0 = A - n\,B \;\rightarrow n = \frac{A}{B} \;(2)$$

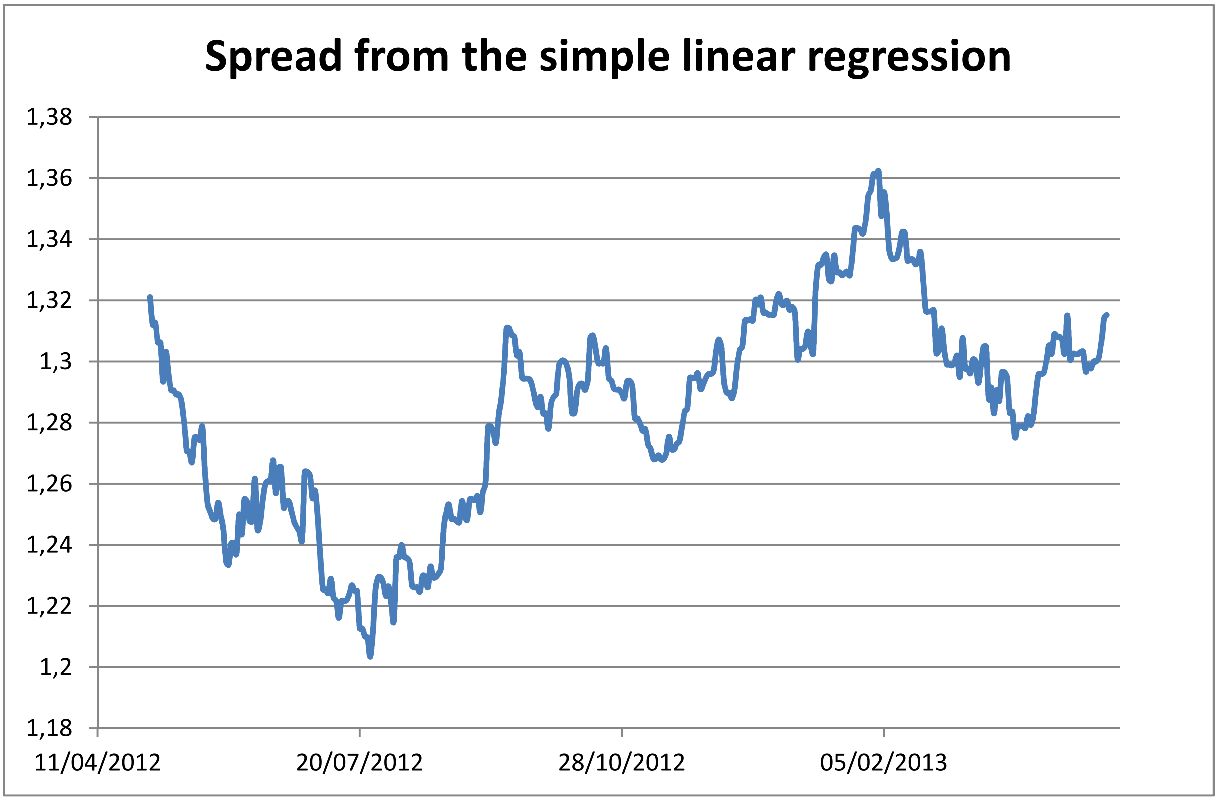

Figure 4: Spread from OLS between 1/5/2012 and 1/5/2013

The figure 4 shows the spread series which are obtained by the following equation:

$$Spread = EURUSD\ rate - 0.0027\ GBPJPY\ rate$$

The spread serie is not cointegrated and stationary, but it is a good example to calculate the spread in different ways as in the next chapter.

## 2. Basic models for statistical arbitrage

The spread is the place to go in order to see and profit from arbitrage opportunities. Several models exist in the literature, and we are going to see only the basic models. There exists many method of pairs trading, to keep it simple I will talk about 3 methods:

1. Distance trading model
2. Cointegration model
3. Differencing model

### 2.1 Distance trading model

With the trading distance model, it is possible to measure the distance between two assets with different methods of calculations.

#### 2.1.1 Distance between normalized prices

With this method, the normalization of prices is necessary (equation 1). The reason why prices are normalized in order to avoid

$$\overline{P_{\imath t}} = \frac{P_{it} - E(P_{it})}{\sigma_i} \quad (3)$$

The variables in the equation 1 are as follows:

- $\overline{P_{\imath t}}$ is the normalized price of asset i at time t.
- $P_{it}$ is the price of asset i at time t.
- $E(P_{it})$ is the average of normalized price of $P_{it}$ on window length of 20 for example.
- $\sigma_i$ is the standard deviation of normalized price of asset i.

Here comes the squared distance between normalized prices.

$$D = P_{ai} - P_{bi}\ (4)$$

Where,

- $D$ is the distance between normalized prices,
- $P_{ai}\ et\ P_{bi}$ are the normalized prices of securities A and B at time t.

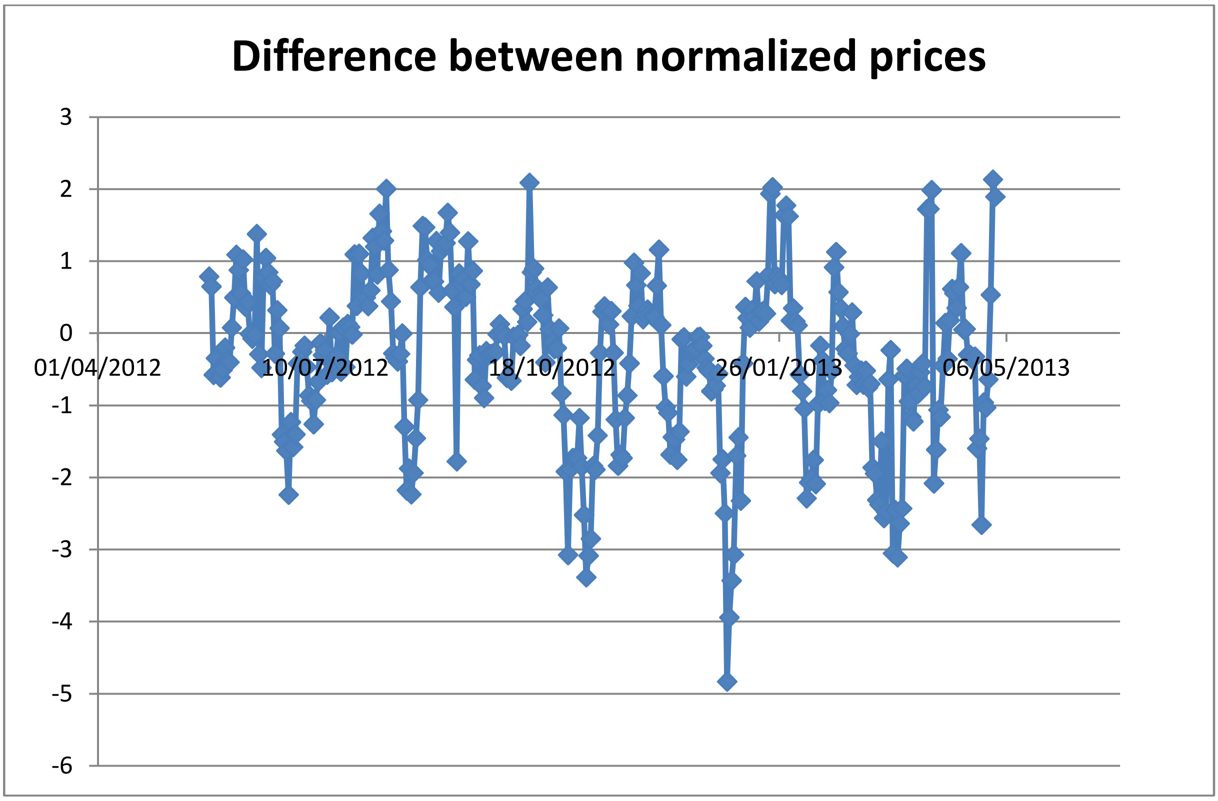

Figure 5: Difference between normalized prices for a EURUSD GBPJPY spread between 1/5/2012 and 1/5/2013

### 2.1.2 Normalized difference between prices

In another paper "Some of existing method of Pair trading" by Andrew Endler [3], several trading distance formulas exist to calculate the distance between two securities. Endler says:

*"Some traders use normalized difference between prices another measure difference between pairs and normalize that difference."*

There is not one good method to calculate the trading distance method. Every and each method has the little steps in order to make their signals, the spread.

$$\overline{d_{\iota t}} = \frac{d - E(d_{it})}{\sigma_i} \quad (5)$$

Where,

- $\overline{d_{\iota t}}$ is the normalized difference between prices of pair trading i at time t,
- $d_{it}$ is the difference between prices of pair trading i at time t,
- $d$ is the difference between prices,
- $E(d_{it})$ is the mathematical average of the difference between prices of pair trading i at time t,
- $\sigma_i$ is the standard deviation of pair trading i.

All these variables are made for a window of 20 periodes. In order to understanding better, you could see the video at the beginning of that article.

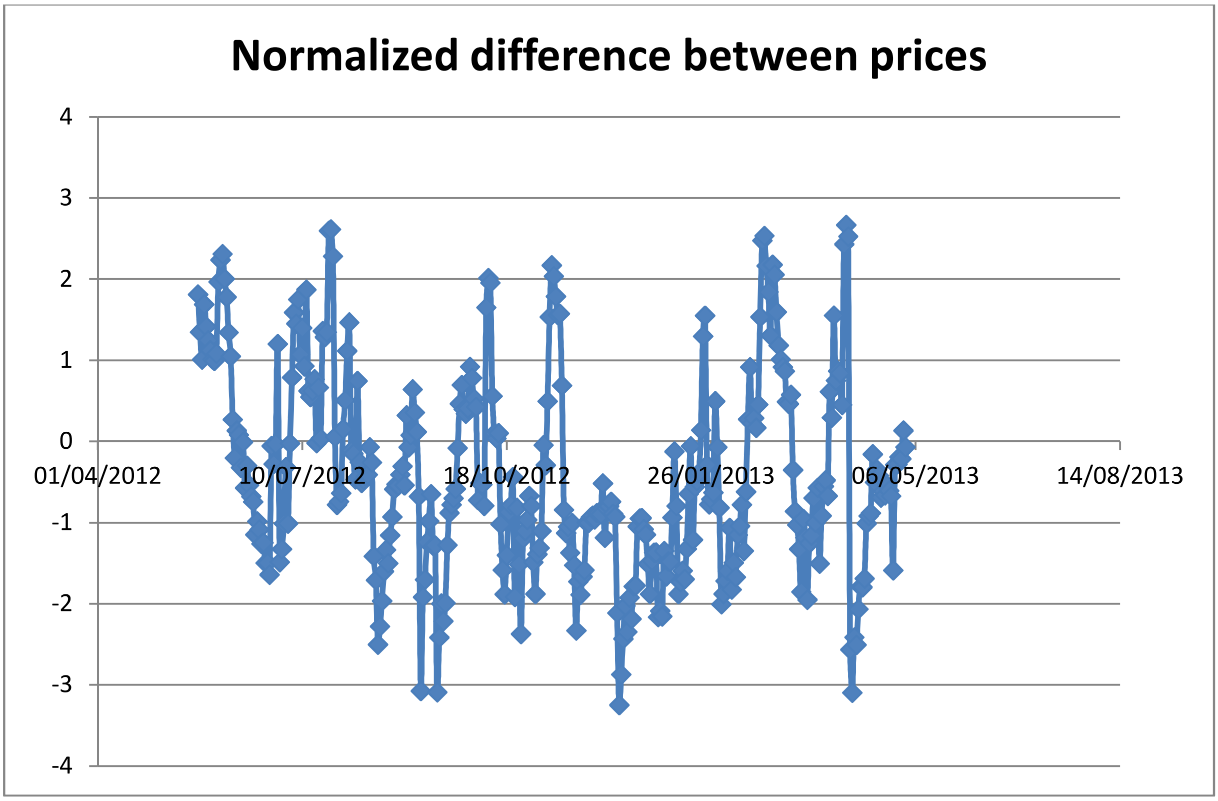

Figure 6: Normalized difference between prices for a EURUSD/GBPJPY spread between 1/5/2012 and 1/5/2013

## 2.2 Cointegration model

Cointegration model as its name said is based on the cointegration between two assets. This cointegration could be calculated by the Engle-Granger test, the Dickey-Fuller test, the Phillips-Perron test or the Johansen test.

We are not going to see in detail the cointegration calculations, but you can see them on the next page by clicking on the link at the end of this page.

An econometric property of time series variables is what cointegration is. Anyway, we want the trading steps for the cointegration model as in the book of Vidyamurthy [2], they are as follows:

"

*The steps involved are as follows:*

1. *Identify stock pairs that could potentially be cointegrated. This process can be based on the stock fundamentals or alternately on a pure statistical approach based on historical data. Our preferred approach is to make the stock pair guesses using fundamental information.*
2. *Once the potential pairs are identified, we verify the proposed hypothesis that the stock pairs are indeed cointegrated based on statistical evidence from historical data. This involves determining the cointegration coefficient and examining the spread time series to ensure that it is stationary and mean reverting.*
3. *We then examine the cointegrated pairs to determine the delta. A feasible delta that can be traded on will be substantially greater than the slippage encountered due to the bid-ask spreads in the stocks. We also indicate methods to compute holding periods.*

"

The cointegration model formula is as follows:

$$\log(P_t^A) - \gamma \log(P_t^B) = \mu + \varepsilon_t \quad (6)$$

Where the variables are as follows:

- $P_t^A$ and $P_t^B$ are the prices of securities A and B at time t,
- $\mu$ is the equilibrium value,
- $\varepsilon_t$ is the residual.

The right hand site of the equation (6) are the residual series, which are composed by two components, $\mu$ is the equilibrium value and $\varepsilon_t$ is the residual.

The steps to obtain the figures 7 to 9 are the following:

1. Drawing **log EURUSD =f( log GBPJPY ) (figure 7)** in order to do a OLS and find the cointegration coefficient.
2. Drawing $\log\left(EURUSD_t^A\right) - \gamma \log\left(GBPJPY_t^B\right) = \mu + \varepsilon_t$ (figure 8) to find out the equilibrium value $\boldsymbol{\mu}$ ***.***
3. Drawing $\varepsilon_t = \log(EURUSD_t^A) - \gamma \log(GBPJPY_t^B) - \mu$ (figure 9) to dawn the residual $\varepsilon_t$.

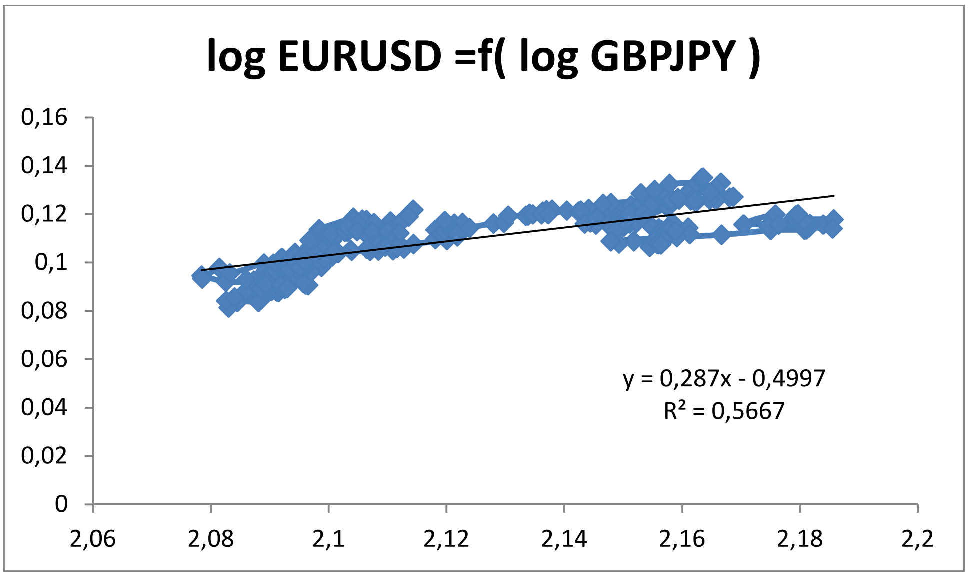

Figure 7: Log EURUSD in function of log GPBJPY to find the cointegration coefficient γ

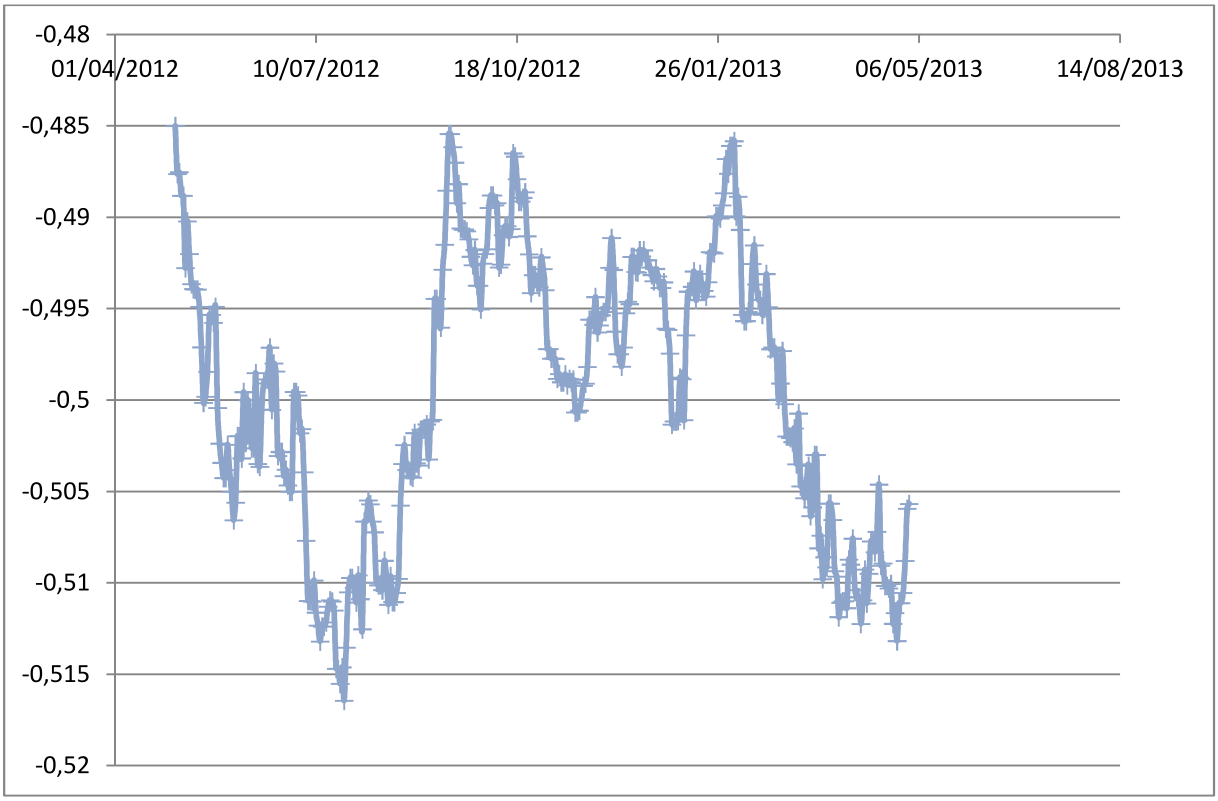

Figure 8: $\log(P_t^A) - \gamma \log(P_t^B)$ curve to find out the equilibrium value

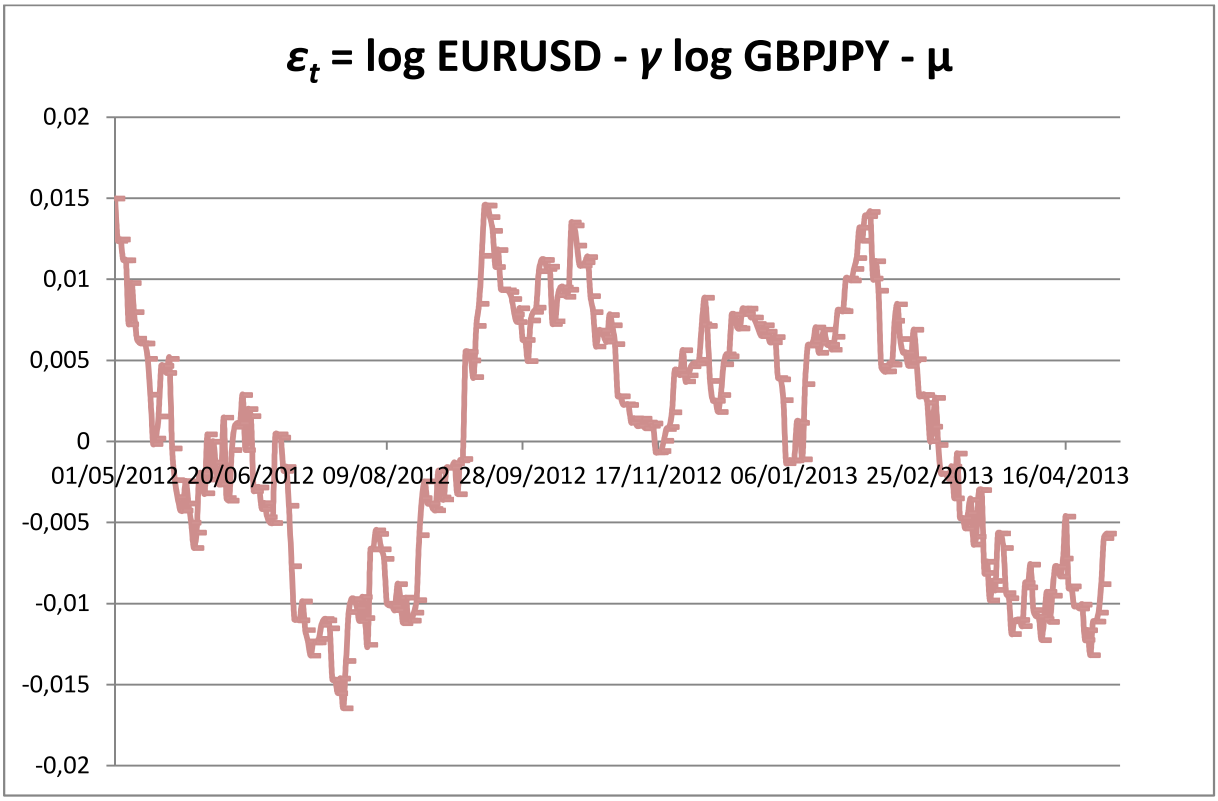

Figure 9: Curve to draw the residual $\varepsilon_t$ to trade with

### 2.3 Differencing model

It is possible to differencing a time series until infinity, but it is important and necessary to do this on a cointegrated time series. If the spread or time series is cointegrated, so the spread is more likely to be mean reverting. That implies the quality of pair trading and the stationary behavior of this one. In fact, if the spread is not stationary, we could make a simple differencing and get a spread which would be cointegrated and stationary.

It is not always true, by the way. Just keeping in mind, we can apply as much as models we want, because models can be created in an infinite quantity.

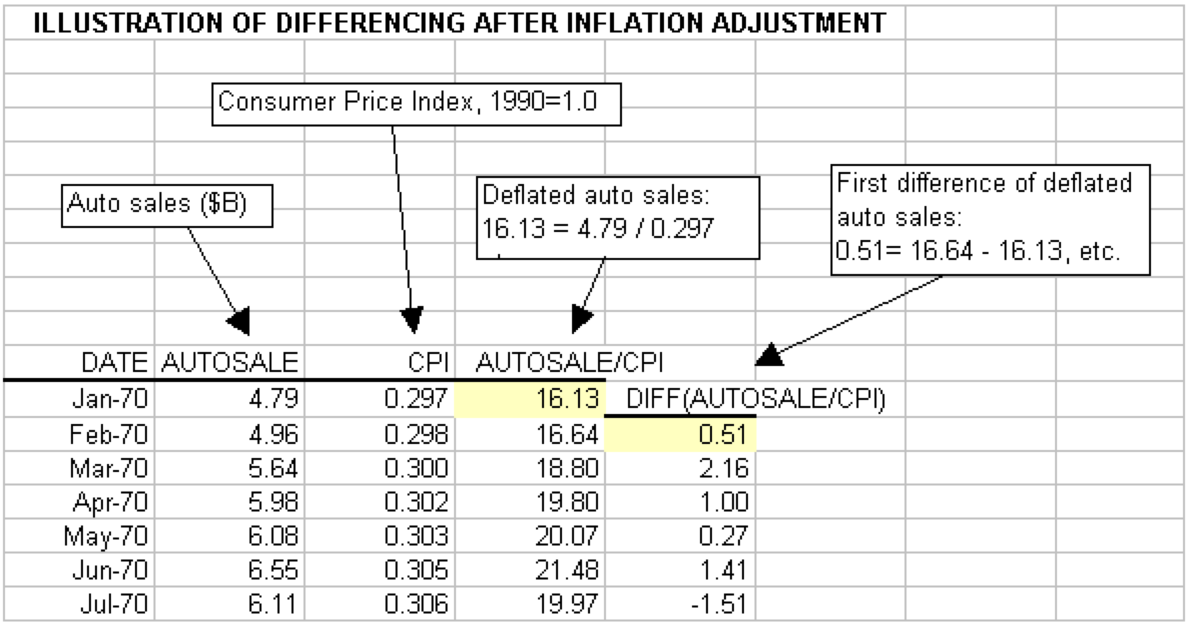

ILLUSTRATION OF DIFFERENCING AFTER INFLATION ADJUSTMENT

Consumer Price Index, 1990=1.0

Auto sales ($B)

Deflated auto sales:
16.13 = 4.79 / 0.297

First difference of deflated auto sales:
0.51= 16.64 - 16.13, etc.

| DATE | AUTOSALE | CPI | AUTOSALE/CPI | DIFF(AUTOSALE/CPI) |
|---|---|---|---|---|
| Jan-70 | 4.79 | 0.297 | 16.13 | |
| Feb-70 | 4.96 | 0.298 | 16.64 | 0.51 |
| Mar-70 | 5.64 | 0.300 | 18.80 | 2.16 |
| Apr-70 | 5.98 | 0.302 | 19.80 | 1.00 |
| May-70 | 6.08 | 0.303 | 20.07 | 0.27 |
| Jun-70 | 6.55 | 0.305 | 21.48 | 1.41 |
| Jul-70 | 6.11 | 0.306 | 19.97 | -1.51 |

Figure 10: Differencing of cointegrated time series AUTOSALE CPI [4]

So, the formula to differencing is really simple as follows:

$$y_t - y_{t-1} = diffirencied\ value\ (7)$$

Where, the variables are:

- $y_t$ and $y_{t-1}$ are the value of a security at time t and t - 1 respectively.

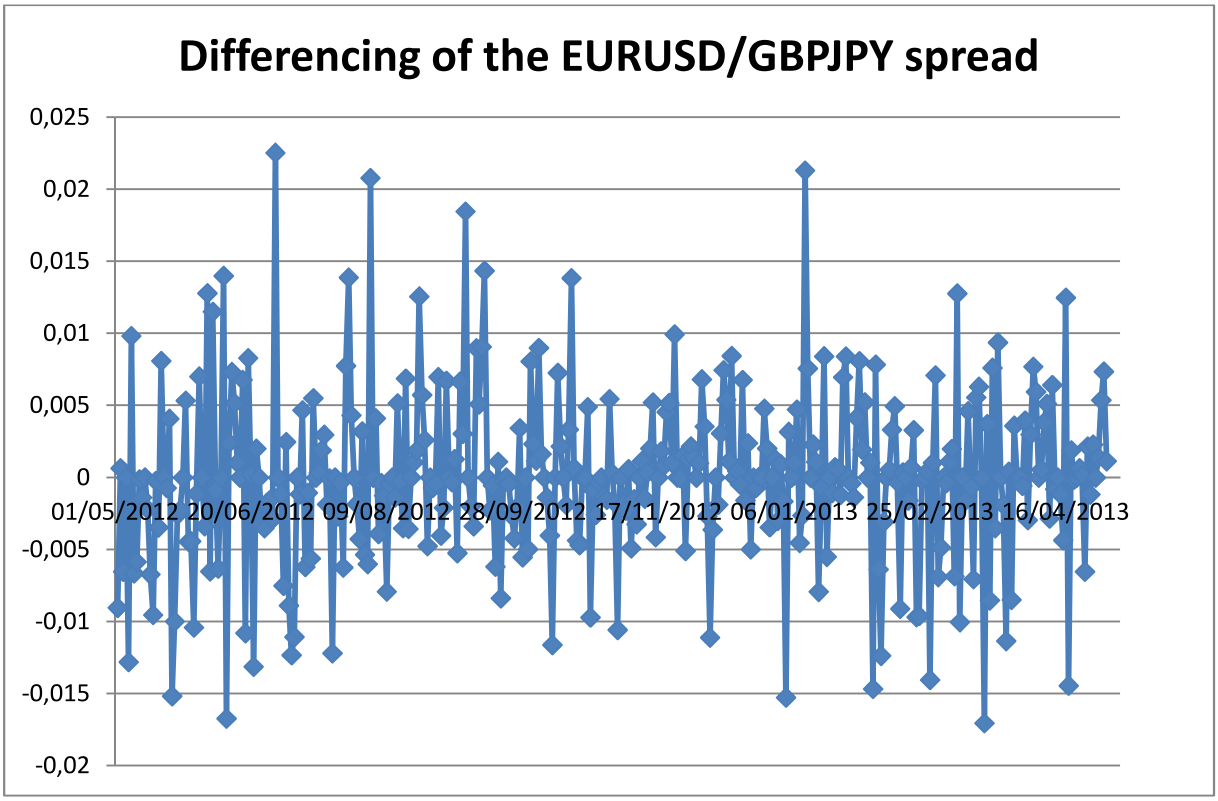

Figure 11: Differencing the EURUSD/GBPJPY spread between 1/5/2012 and 1/5/2013

## Conclusion

Each and every method to build spread series has its advantages and its disadvantages for a particular pair trading. So, it is necessary to backtest with good datas in order to find the right model for the right pair trading securities.

These models give a statistical arbitrage in fact, they are not true arbitrage opportunities by definition but who cares.

The main issue is the calculations of cointegration for hundreds or even thousands of pairs trading combinations. If we should be doing that with a simple spreadsheet in Excel, we could shoot ourselves! :)

*Samuel Caan from http://gamblingandinvesting.com*

## Bibliography

[1] *Pairs Trading, Convergence Trading, Cointegration*, Daniel Herlemont, YATS, 2004

[2] *Pairs Trading, Quantitative Methods and Analysis*, G. Vidyamurthy, 2004, John Wiley &Sons, Canada

[3] *Some of existing method of Pair trading*, Andrzej Endler, 2004

[4] http://people.duke.edu/~rnau/411diff.htm