Some of existing method of Pair trading

"Pairs trading is one of Wall Street's quantitative methods of speculation which dates back to the mid-1980s (Vidyamurthy, 2004[i]). In its most common form, pairs trading involves forming a portfolio of two related stocks whose relative pricing is away from its "equilibrium" state. By going long on the relatively undervalued stock and short on the relatively overvalued stock, a profit may be made by unwinding the position upon convergence of the spread, or the measure of relative mispricing"[ii]. Pair trading is easiest to understand when considering equities can be use with any assets futures, options and currency. Especially currency are more deeply influenced by macroeconomic events than other security type and, as such, require a degree of awareness.

There is many method of pairs trading. I'll mention about 2:

a) Distance trading,
b) Cointegration model.

The distance method is used in Gatev[iii] et al (1999)and Nath [iv](2003) for empirical testing whereas the cointegration method is detailed in Vidyamurthy (2004).

In example below I use Close price but better way is to use VWAP Volume Weighted Average Price.

Distance trading

In distance trading we measure distance between two (or more) moving together assets. We can do this in different ways. Often in assets selecting process one use squared distance between normalized prices. (Equation ([1]). One could also use a maximum correlation criteria, but the results are quite similar.

$$D = \sum_{i=1}^{n}(P_{ai}-P_{bi})^2$$

[1]

Where P_{ai} and P_{bi} are normalized assets prices.

Prices should be normalized because use of original prices (without normalization) would be a problem for the case of minimum squared distance rule since two assets can move together but still have a high squared distance between them. The transformation employed is the normalization of the price series based on its mean and standard deviation.

$$P_{it} = \frac{P_{it} - E(P_{it})}{\sigma_i}$$

$$[\,2\,]$$

Where P_{it} is normalized price of asset i at time t, $E(P_{it})$ is just the expectation of P_{it}, in this case the average, and σ is the standard deviation.
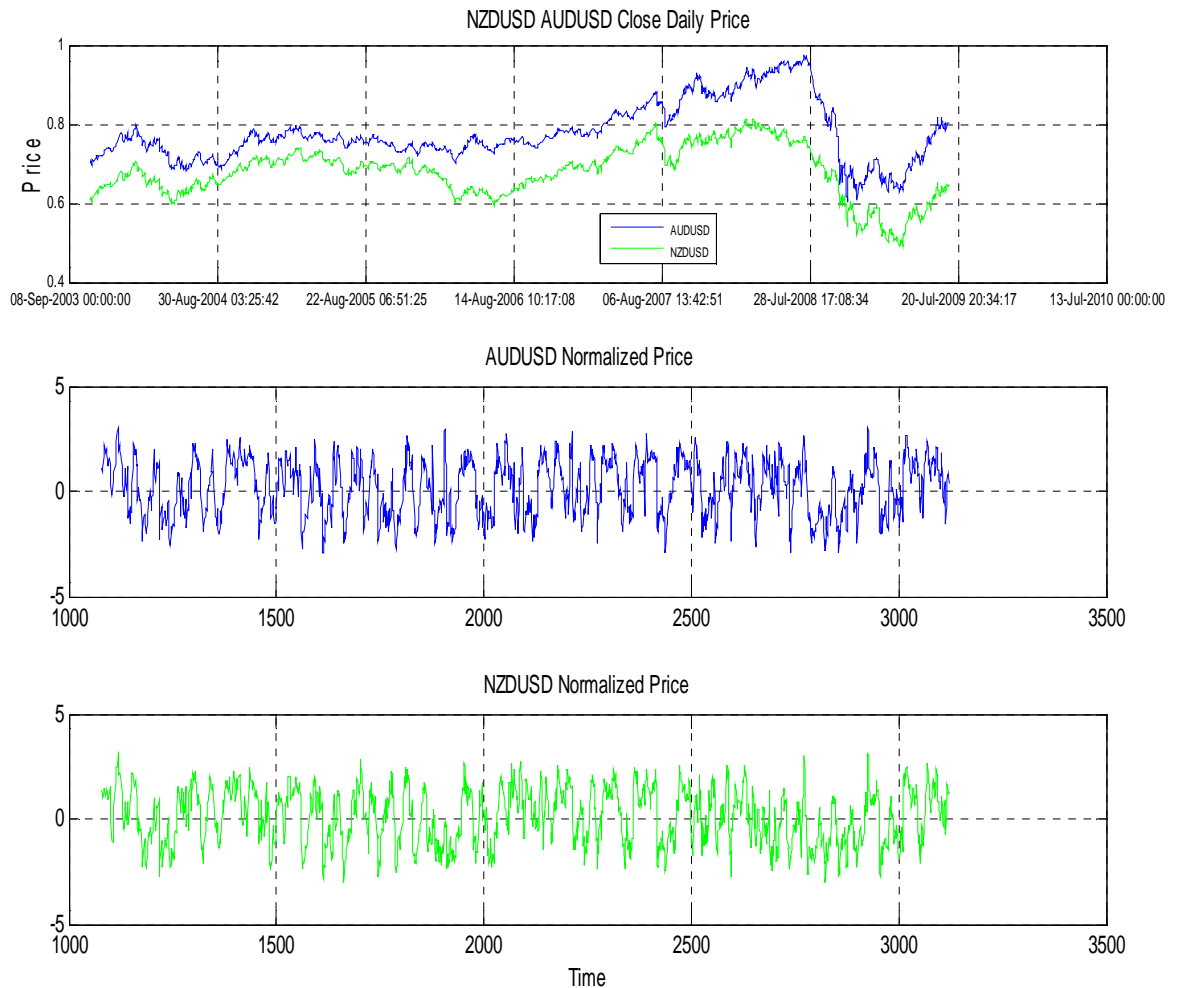

Figure 1 Example of normalized prices

We than could use difference between this normalized prices as trade trigger . I'll write about this a little later.
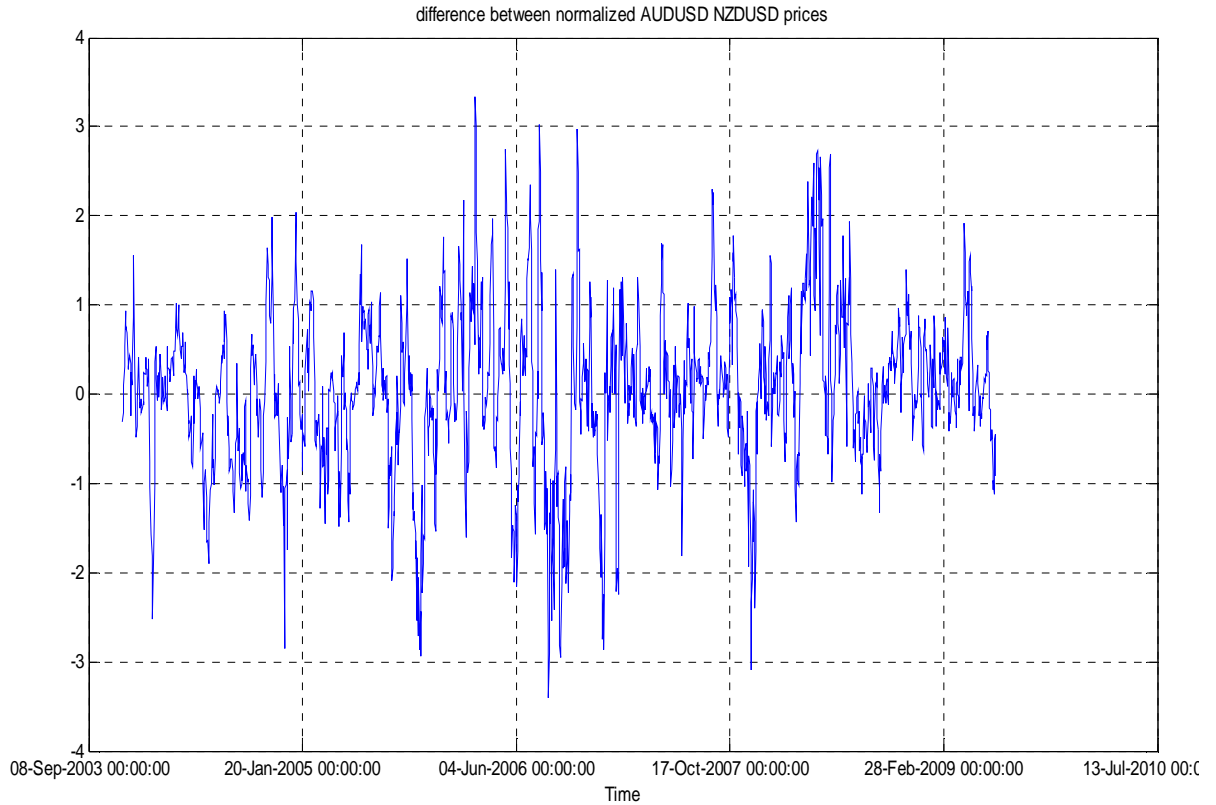
Figure 2. Difference between normalized pair

As we can see there is mean reversing bias we could use to trade.

Some traders use normalized difference between prices another measure difference between pairs (spread between market prices) and normalize that difference (if we selected pair earlier). We can use the same equation [2] but P_{it} is now d_{it} – difference between prices.

$$d_{it} = \frac{d - E(d_{it})}{\sigma_i}$$
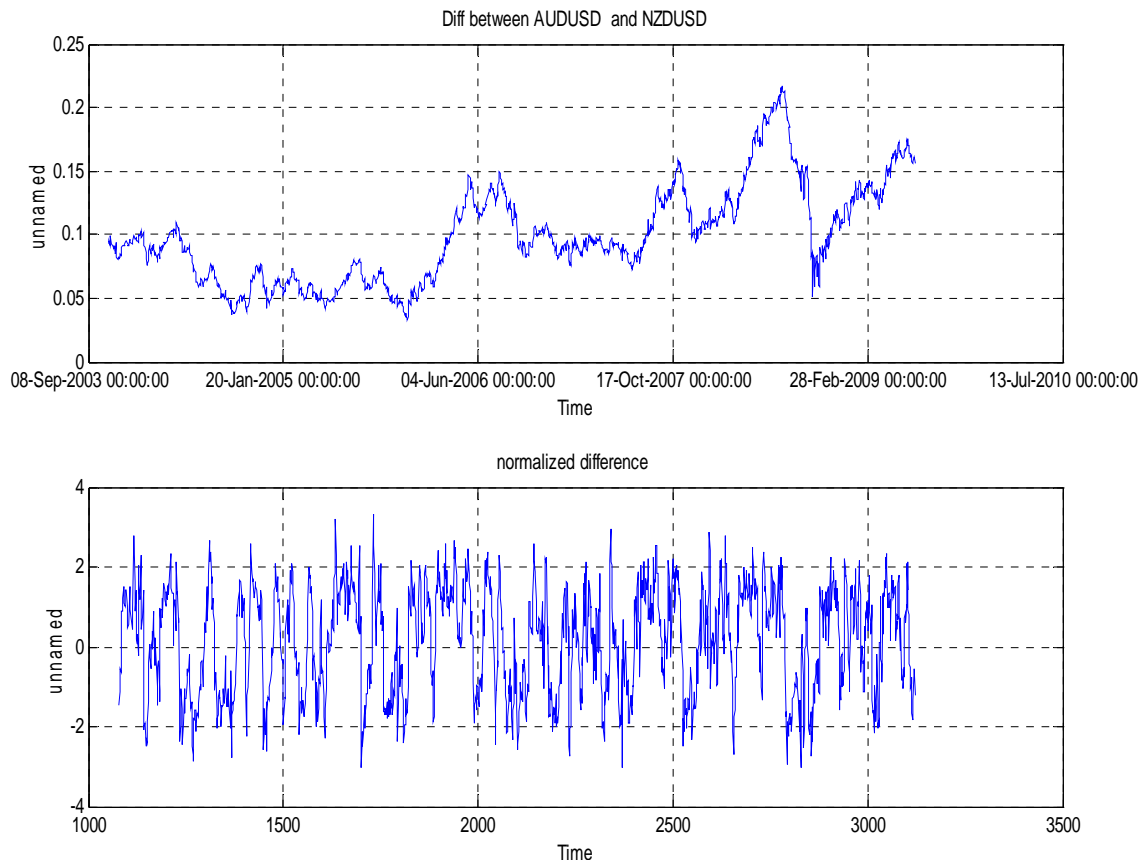
[3]

Where $d_{it}= P_{ait}-P_{bit}$

Figure 3. Normalized difference

One could set different test window length and mean / std.dev length ,figures above are made for mean length=20.

When we know something about price and spread behavior in test period we could try to set some trading rules.

We should trade (buy or sell synthetic asset – buy one asset and sell another) when distance is above / below some level. How we set this levels is important. Levels could be set in different terms e.g. as standard deviation or percentile terms. In standard deviation terms we set it on some multiplier of standard deviation of test period (I mean deviation from difference between normalized prices or normalized difference of prices as describe above). In percentile terms we use distribution of our difference and calculate percentile ." A percentile (or centile) is the value of a variable below which a certain percent of observations fall. So the 20th percentile is the value (or score) below which 20 percent of the observations may be found." [v]
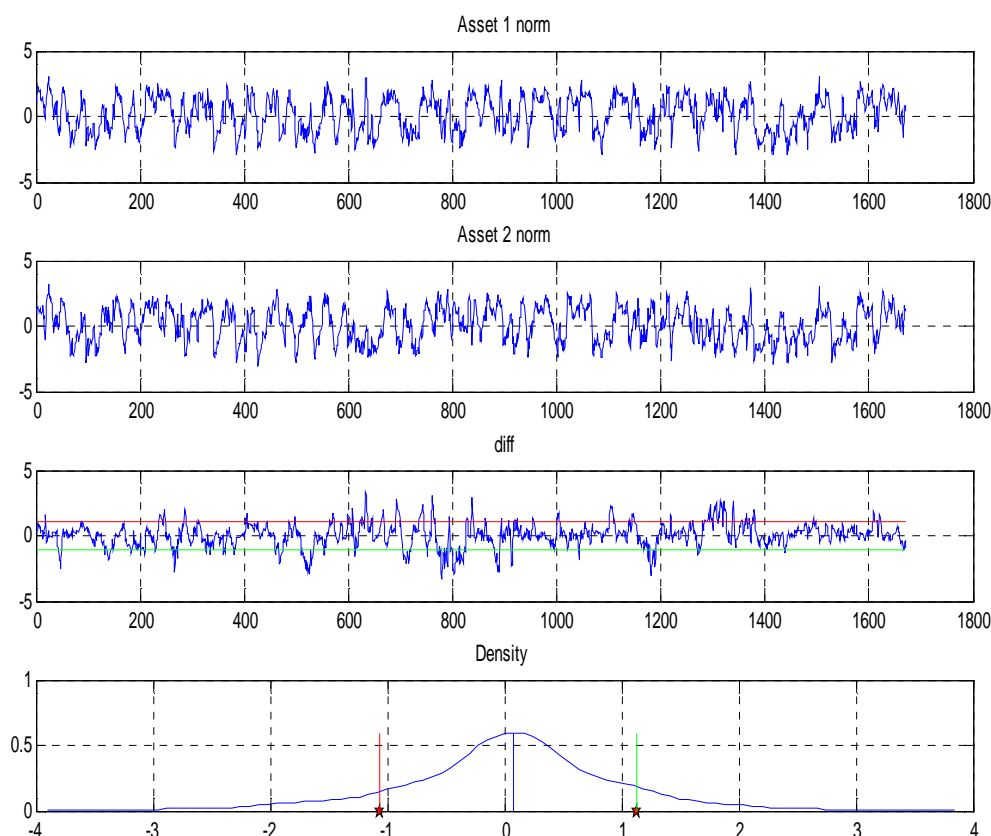
Figure 4. Percentile Triggers.

In the figures above red and green lines are 10% and 90% percentile, blue line is median. So we could trade when our spread is above/ below this value. We sell first asset and buy second when diff is above red line and buy first and sell second when diff is below green lines. We could reverse position at opposite level (when we enter on red then close and open on green and vice versa) but it could be too aggressive (for me at least). I usually close trade when spread return to it's mean – on zero line.

But what percentile or standard deviation multiplier should we use to maximize profit ? First we should check if spread is tradable in the meaning of bid/ask spread and transaction costs. Potential profit should be greater than costs (as always in trading - pretty simple ;).

When we use small multiplier of std. dev or big percentile we'll have more trade but smaller, when we use bigger multiplier or smaller percentile we'll have less trade but bigger. So we should calculate optimal levels.

It could be done by counting trade level crossing (how many trade) and calculate profit as proportional to that level (bigger levels = bigger profits on trade) .On fig. below we see count of level crossing as a function of standard deviation.
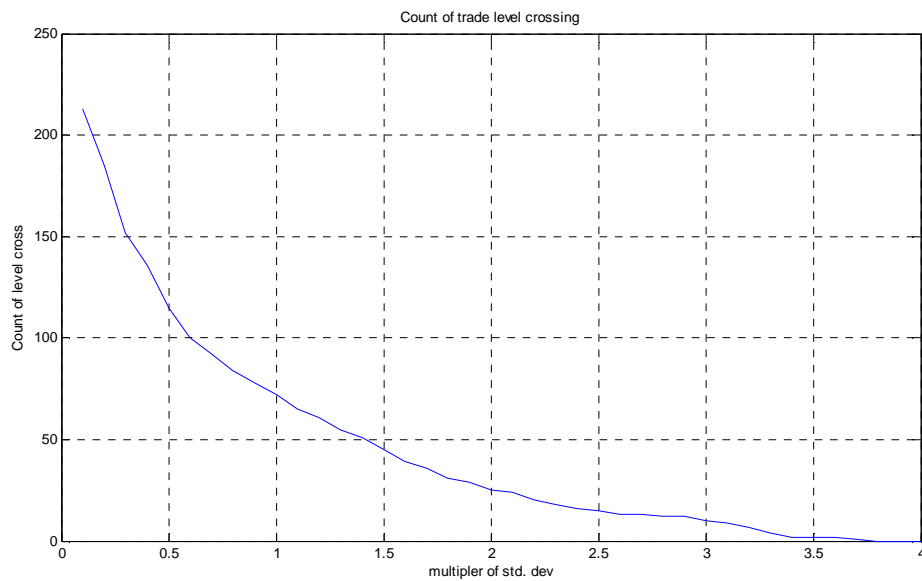
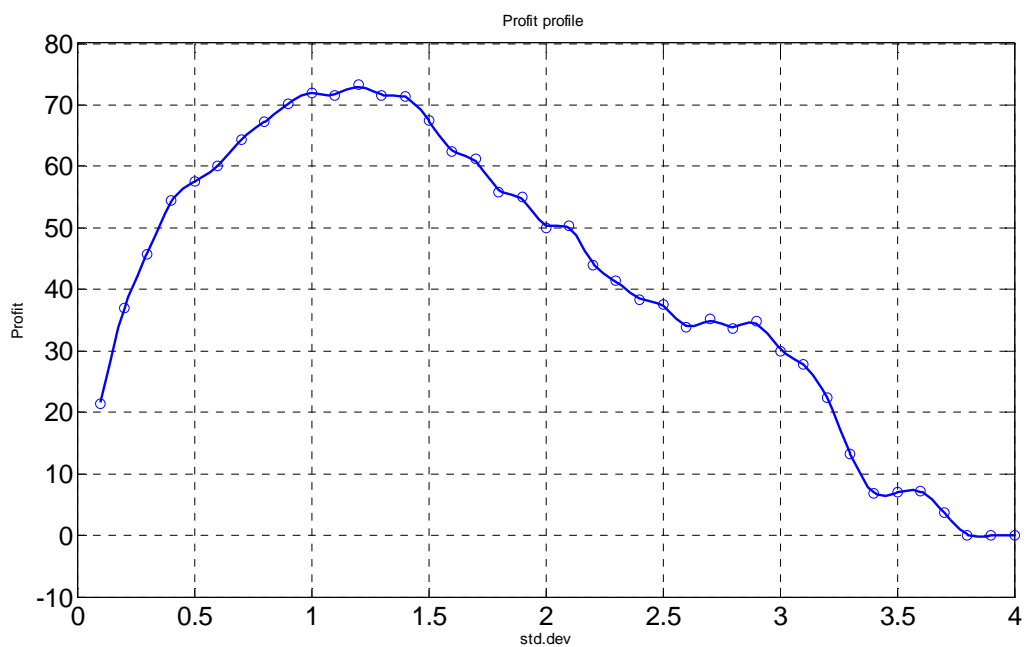So we could estimate optimal level by multiply level crossing count by that level

Figure 6. Profit profile

As we see max profit we get about 1.2*σ.   To calculate real profit we should use market prices and real costs, but max profit should be about 1.2* σ (for this example).

In overall, the distance approach purely exploits the statistical relationship of a pair, at a price level. As the approach is economic model-free, it has the advantage of not being exposed to model mis-specification and mis-estimation.

Cointegration

Cointegration method is little more complicated and one need some statistical background to understand it fully, but it could be use without understanding all theory behind it.

Cointegration is an econometric property of time series variables. If two or more series are themselves non-stationary, but a linear combination of them is stationary, then the series are said to be cointegrated [v]. Stationary means (in our example) that probability distribution does not change when shifted in time. Stationary time series are mean reversing – as we need under pair trading.

"The steps involved are as follows:

1. Identify stock pairs that could potentially be cointegrated. This process scan 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." Vidyamurthy (2004) [i]

The model that is most commonly assumed for stock price movement is called a log-normal process; that is, the logarithm of the stock price is assumed to exhibit a random walk. This have some statistical implications (to be found in statistic books).

So in first step we calculate log of prices and create two series $\{\log(P^A_t\}$ and $\{\log(P^B_t\}$.

Then we try to find linear correlation between this series in mathematical sense:

$$\log(P^A_t) - \gamma \log(P^B_t) = \mu + \varepsilon_t \qquad [\,4\,]$$

left side is linear combination of log time series, right hand side represents the residual series and is expressed as sum of two components μ is equilibrium value and ε_t is time series with zero mean. When series is mean reverting we expect that it will oscillate about the

equilibrium. So we can say that spread will oscillate about its mean μ as described by ε_t time series.

We have a model, now we should estimate this relationship (that's mean γ cointegration coefficient and μ equilibrium value).

It could be done in different ways, we will use linear regression. So we use some software e.g. Excel (see description of my file below) and calculate this coefficient and residuals.

For our AUDUSD NZDUSD (90 days from about March 16' 2009) example it could look like this:

γ=1.0481 μ=0.2525 R^2=0.9306

Rsquere say us about how good this linear relationship describe log price difference behavior. There are of course also some statistic like t or f statistic.

So we have model that describe linear combination between AUDUSD and NZDUSD

$$\log(P_t^A) - 1.0481 \log(P_t^B) = 0.2525 + \varepsilon_t$$

Where P_t^A is price of AUDNZD and P_t^B is price of NZDUSD at time t.

Now we can calculate residual series ε_t.

$$\varepsilon_t = \log(P_t^A) - \gamma \log(P_t^B) - \mu \quad [\,5\,]$$

so:

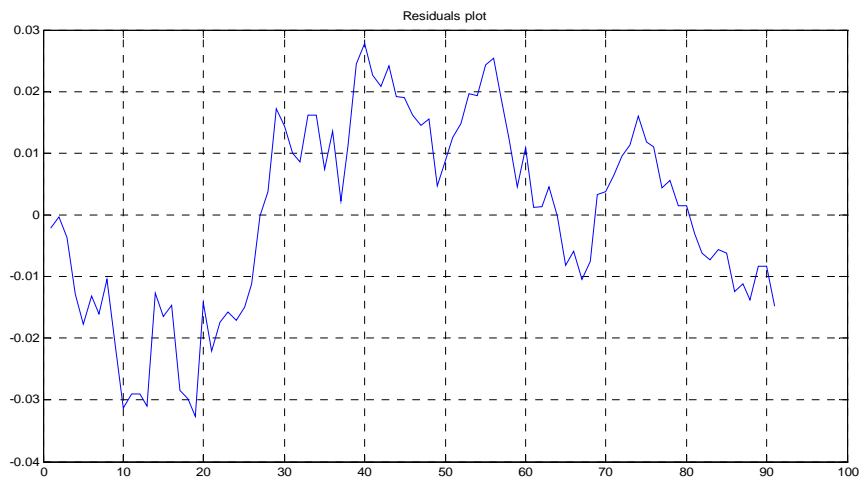$$\varepsilon_t = \log(P_t^A) - 1.0481 \log(P_t^B) - 0.2525$$

Figure 7. Residuals series plot.

Now we should perform some test if this series are stationary e.g. Dickey-Fuller or Philips-Perron test. Our series passed it (or if we don't know how, we just don't do it, if we believe that our series is mean reversing).

We could use this residual series in similar way as we used earlier series of difference . We could set some trade levels and calculate how many times series cross zero line (mean value of spread) and profit.
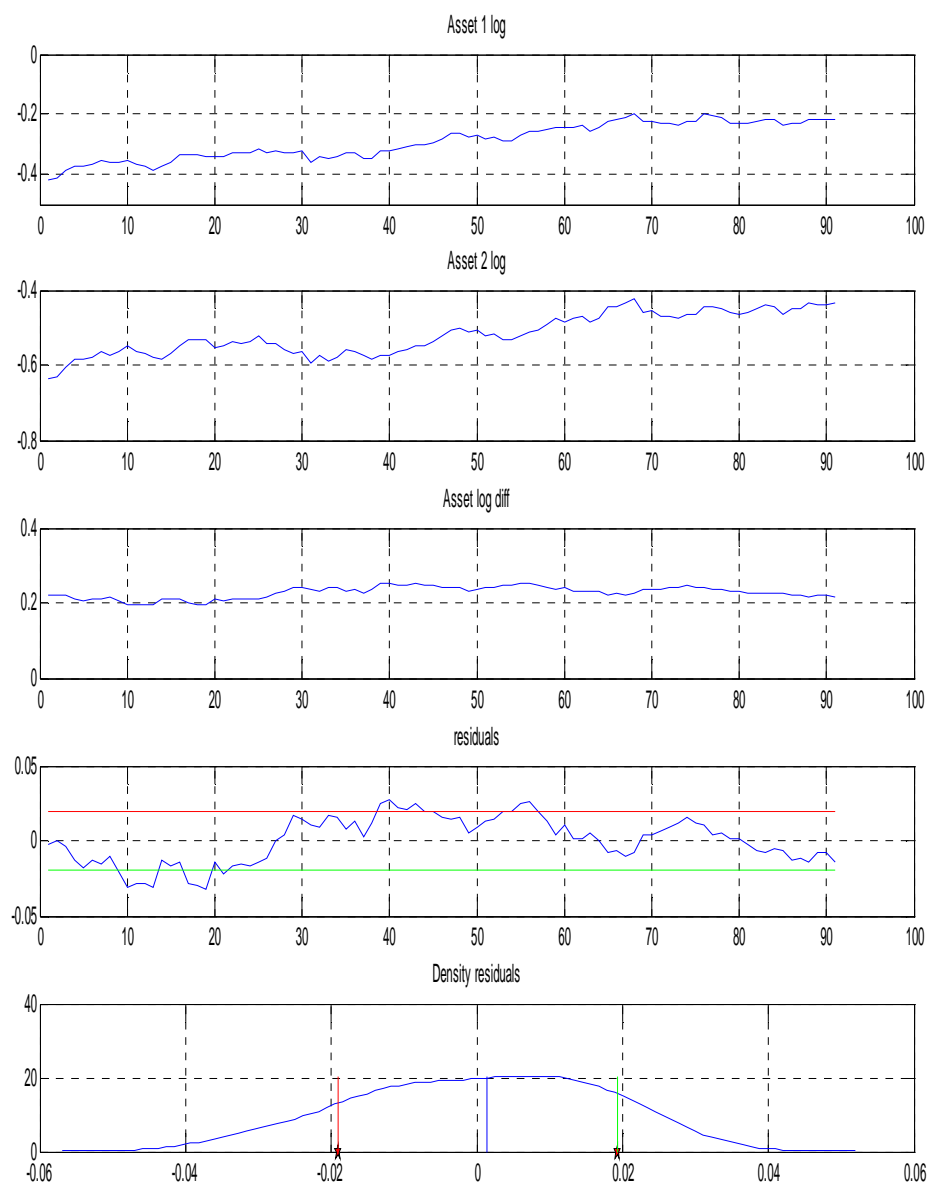

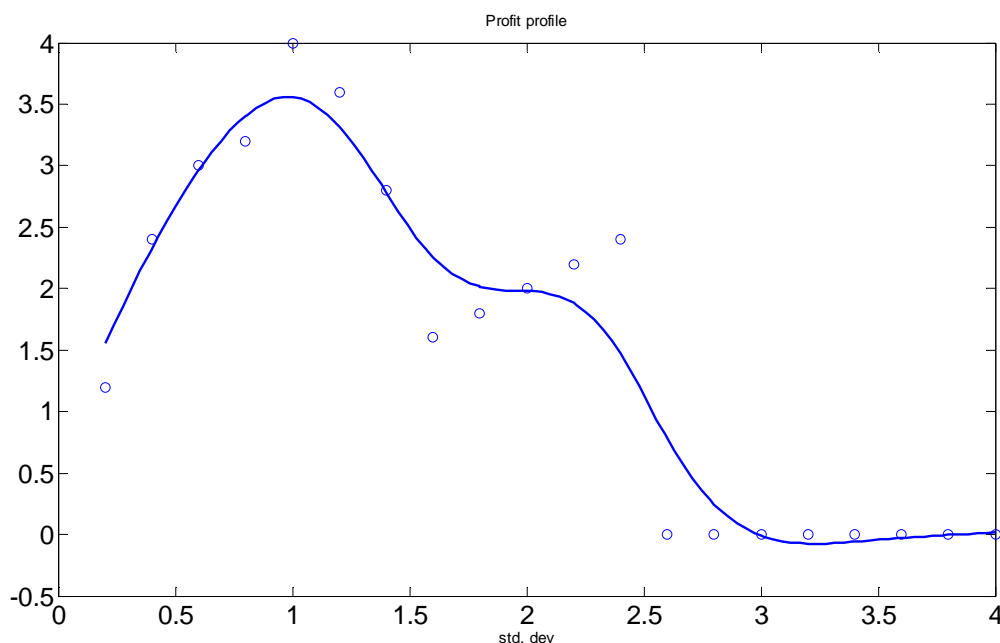**Figure 8. Residuals plot and levels**

Figure 9. Profit profile

In this example max profit is about 0.98*σ.

So, we first calculate regression, than we set trade levels and trade crossing it by residuals. Residuals we calculate using equation [5].

Parameters of this models always change in time. As longer we are in trade it's more risky. There is phenomena called "mean drift" - mean are drifting with time in markets and it's not good for us because we trade reverse to this mean.

Excel file

Excel file contains some spreadsheets one can use to test pair and find level to trade. In first and second are pair data (in example I use Close prices of USDCHF and EURUSD). In next "CALC 1" are formulas for normalized difference calculating. In "CALC 2" are formulas for difference of normalized prices and in "Regr" are cointegration model.

So You could use it to estimate model and trade it.

In real trading system I use trading rules as above and some another gadgets like trailing stops (profit trailing) and emergency stop loss.

Size of both legs of trade should be calculated using market neutrality and money management rules.

Enjoy.

Andrzej (Andrew) Endler

andrzej_endler@post.pl

[i] Vidyamurthy, G. (2004) Pairs Trading, Quantitative Methods and Analysis, John Wiley & Sons, Canada.

[ii] Binh Do _ Robert Faff † Kais Hamza (2006) "A New Approach to Modeling and Estimation for Pairs Trading"

[iii] Gatev, E., G., Goetzmann, W. and Rouwenhorst, K. (1999) "Pairs Trading: Performance of a Relative Value Arbitrage Rule", Unpublished Working Paper, Yale School of Management.

[iv] Nath, P. (2003) "High Frequency Pairs Trading with U.S Treasury Securities: Risks and Rewards for Hedge Funds", Working Paper, London Business School.
[v] Wikipedia