PROCEDURE TO DETECT MEAN REVERSION IN (STOCK) PRICES

Aretina David-Pearson

University "Politehnica" Bucharest, Faculty of Applied Sciences, Romania, e-mail: aretina_david@yahoo.co.uk

Abstract: *A method helping in detecting and extracting the mean reversion trend of stock prices is described hereby, starting from the arithmetic Ornstein-Uhlenbeck mean-reversion model.*

Keywords: *mean reversion, stocks, trend, strategic allocation*

1. INTRODUCTION

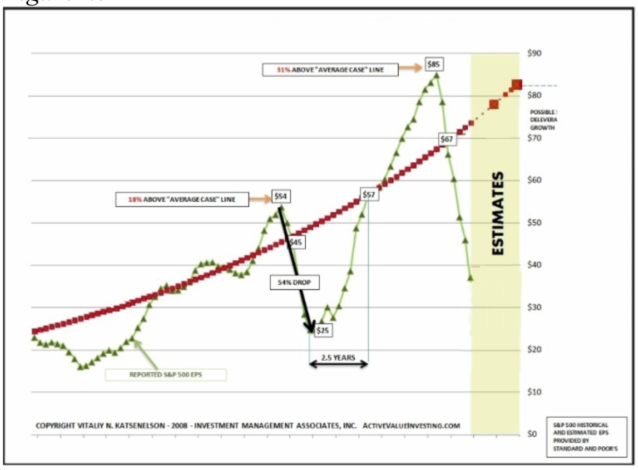
There have been a large number of studies (see among others *Poterba and Summers* [1], or *Spierdijk, Bikker and van den Hoek* [2]) trying to infer if stock prices exhibit generally mean reversion or not. The present paper is not trying to answer such a question, but rather to provide a practical, automisable approach in order to identify the stocks which have a stable trend, on which on can bet.

From practical point of view, it is important to get the general trend of stock prices, especially when deciding the strategic, long time, asset allocation. However, it is not obvious which stocks exhibit mean reversion and how to calculate it, while the noise might play an important role.

As one can see from *Figure1*, the mean-reversion property is not a striking one, and most of the stocks have such jumps that this algorithm does not guarantee always finding the mean to which the stock reverts.

Example of stock (index) exhibiting mean-reversion

Figure no 1



As in most of the analysis, one should consider only data points where no significant financial events showing up, such that the overall financial system is not perturbed. In addition, the historical time window should be at least of the order of the time period on which the mean reversion is inferred for.

The second section of this paper describes the mathematical formalism used in deciding on the parameters signaling a mean-reverting trend. The relationships that they should satisfy is inferred and described in this part.

The third section goes through three possible methods to validate the stability of the inferred values of the parameters, with some visual examples on how distributions ideally should look like.

2. PARAMETERS ESTIMATION

One starts from the arithmetic Ornstein-Uhlenbeck mean reversion model as developed by Schwartz [3], with m being the mean to which the reversion is done and h being the speed of the reversion:

$$dx = h(m-x)*dt+s*dz \quad (1)$$

Equation(1), describing the variation of the logarithm of the stock price P, $dx = \frac{dP}{P} = d(\ln P)$ is a continuous time version of the first-order autoregressive process, AR(1) in discrete time (see *Dixit and Pindyk* [4]), as dx represents the limiting case (Dt tends to zero) of the AR(1) process:

$$x_t - x_{t-1} = m\,(1 - e^{-h\,Dt}) + (e^{-h\,Dt} - 1)\,x_{t-1} + e_t \quad (2)$$

Hereby e_t is the noise, normally distributed with mean zero and standard deviation s_e, and:

$$s_e^{\,2} = \frac{[1-\exp(-2h)]}{2h} s^2 \quad (3)$$

Considering the Taylor expansion of *Equation(2)*, for the limit case *Dt->0*, *Equation(4)* is obtained:

$$\begin{aligned} x_t - x_{t-1} &\sim m(1-(1-h\,dt)) + ((1-h\,dt)-1)\,x_{t-1} + e_t = \\ &= mh\,dt - hx_{t-1}dt + e_t \\ &= h(m-x_{t-1})dt + e_t \quad (4) \end{aligned}$$

In order to estimate the parameters of mean-reversion, one runs the regression:

$$x_t - x_{t-1} = a + b\,x_{t-1} + e_t \quad (5)$$

The two parameters, m and h, can be calibrated as follows:

$$m = -a/b\,; \quad (6)$$

$$h = -\ln(1+b); \quad (7)$$

From *Equation(3)* and *Equation(7)*

$$s = s_e * \sqrt{\frac{2*\ln(1+b)}{(1+b)^2 - 1}} \quad (8)$$

where s_e is the standard deviation from the regression, calculated from the residuals distribution e_t.

The unit in the above equation is percentage by time unit, and of course the time unit is the same time-series unit (if using monthly time-series, it is % per month, etc.). If one uses monthly data and wants to obtain annual values for the parameters, one has multiply the value of h, obtained in the equation above, by 12 while multipling the value of *s* obtained *Equation(8)* by the square-root of 12.

The S&P 500 index is an excelent toy-data model, especially given its clear mean reverting trend for the analysed period. In real life, especially for shorter time periods, the trend is mostly not visually observable.

Starting from *Equation(4)*, one obtains the next formula for the fitted data, by considering $dt=1$:

$$x_t = m(1-(1-h)^{(t-1)}) + (1-h)^{(t-1)} * x_1 \quad (9)$$

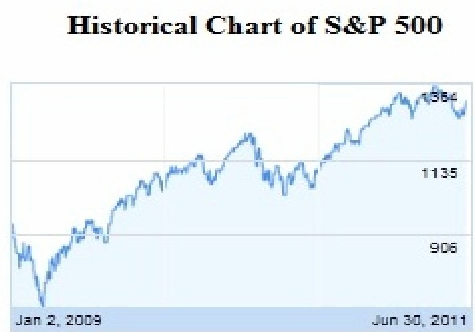
3. MODEL VALIDATION

As one can always find a solution for the set of calibration parameters, one should decide when such values make sense, and how well they describe the actual behaviour.

In an exemplifying analysis, one has used the S&P 500 daily closing prices, from *01.01.2009* to *30.06.2011* (one and a half years). The actual prices S_t, presented in the historical chart from *Figure2,* have been translated to the logarithmic scale, $x_t=ln(S_t)$.

Historical chart of S&P 500, for the considered period

Figure no 2



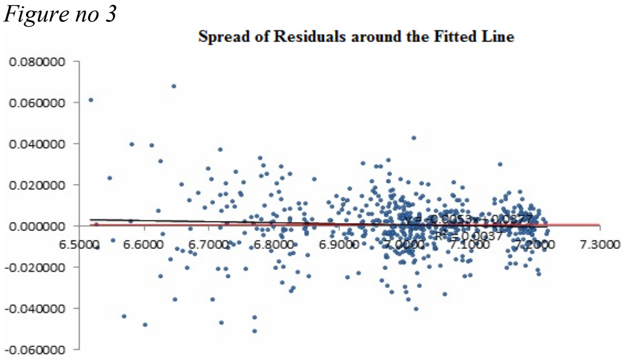
Below there are three possible complementary methods one can validate the final results with: goodness of fit/distribution of residuals, cross-validation and the Jackknife method/bootstrapping.

3.1. Goodness of fit/Distribution of the residuals

An ideal regression should look like in *Figure3*, where the blue data points are distributed on both sides of the red fitted line*Figure 3: Almost ideal distribution of data for linear regression.*

Spread distribution of data points around the fitted line

Figure no 3



On the other side, the residuals should be distributed more or less normally, as *Equation(2)* assumes.

Histogram of residuals, bell shaped, with a standard deviation s_e=0.012923, a kurtosis of 3.28 and a skewness of -0.58

Figure no 4

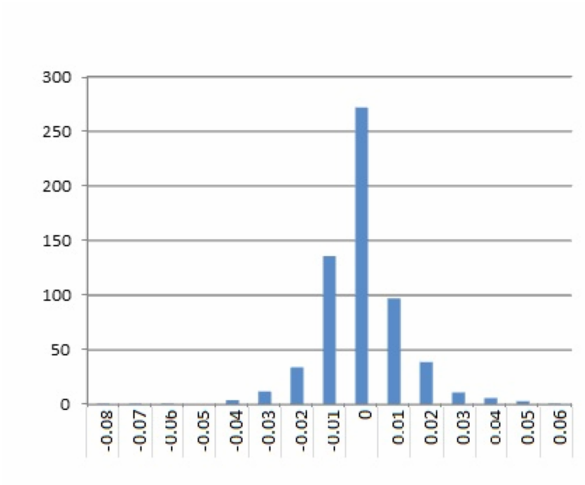


Figure4, containing the histogram respresentation of the distribution showed in *Figure3,* and its fitted parameters indicate a distribution closed to a perfect Gaussian one (which would be characterized by an expected kurtosis value of 3 and an expected skewness of 0.00).

The fitted regression parameters were as it follows:

$a=1.710 * 10^{-5}$

$b=1.065*10^{-4}$

They lead to a standard error *s* giving good confidence in the results:

$m=-0.1605$

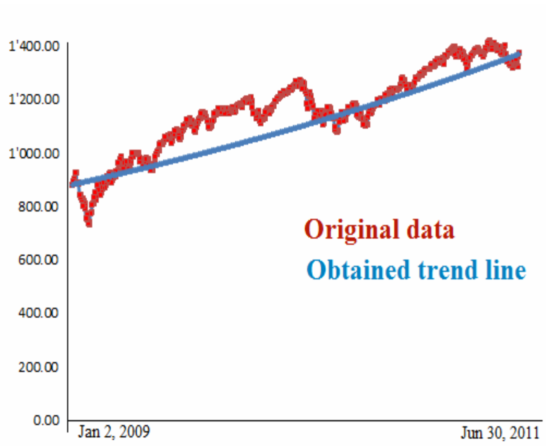
$h=-1.065*10^{-4}$

$s=0.012922$

Making use of *Equation(9)*, one obtains the results from *Figure5:*

The fitted data, in blue, do show the increasing mean-reverting trends line of the original data, in red

Figure no 5



3.2. Cross-validation

The cross-validation excludes one observation at a time, when estimating regression coefficients, and then uses these coefficients to predict the excluded data point.

This procedure is repeated for all data points.

At the very end, the estimated values can be compared with the ones obtained making use of the full distribution. There are no methods for testing statistical significance with cross-validation.

When cross-validation is going to be used as a mean-reverting feature selection, to predict the future values, it is good to remember that it might be possible to over-fit the crossed validated statistic and end up with a model that performes under expectations. As the effects of optimised cross-validation statistic can be a too optimistic performance estimate (see for example *Ambroise and McLachland* [5])

However, averaging the quality of the predictions across the validation sets yields an overall measure of prediction accuracy.

3.3. Data-Resampling: Jackknife & Bootstrap

As in the previous *Subsection2.2,* within the Jacknife method one excludes one observation at a time, when estimating regression coefficients. In the case when the observation *i* is excluded, one gets *Equation(9)*:

$$x_t - x_{t-1} = a_i + b_i x_{t-1} \quad (9)$$

From this new set of replicates of the original statistic, an estimate for the bias and an estimate for the standard error of *a*and *b* parameters can be calculated. The Jackknife method is an useful approach when the dispersion of the distribution is wide or when extreme values are present in the data set. For an overview of this method one could use *Yu* [6].

Alternatively, especially when the data set is of about twenty points, one can make use of the bootstrapping method. The standard bootstrapping method, applied to, say twenty events, uniformly generates a random number between 1 and 20, as many times as the enriched sample has to count (three hundred events is among the most used numbers). By arbitrarily ordering the twenty events, from 1 to 20, every time the number 3 shows up, the third event is added to the enriched sample. When the number of events is much lower, say about five events, bootstrapping is still possible, but making use of low-sample specific, dedicated techniques. Detailed algorithms for bootstraping regression models can be found, for example, in *Davison and Hinkley* [7].

4. CONCLUSIONS

The method presented here gives an additional, practical tool to the portfolio risk managers in order to decide which financial instruments their portfolios should contain.

The method is intuitive, simple, and its results are easy to interpret, as well as being sufficiently uncomplicated to allow automation of the whole procedure, including the validation tests.

5. REFERENCES

[1] James Poterba and Lawrence Summers (1987). *Mean Reversion in Stock Prices: Evidence and Implications*. National Bureau of Economic Research Working Paper Series, Cambridge - MA, Working Paper No. 2343.

[2] Laura Spierdijk, Jacob Antoon Bikker and Pieter van den Hoek (2010). *Mean Reversion in Stock Markets: An Empirical Analysisof the 20^{th} Century*. DNB Working Papers from Netherlands Central Bank, Research Department, No. 247/April 2010.

[3] Eduardo Schwartz (1997). *The StochasticBehaviour of Commodity Prices:Implications for Valuation and Hedging*. The Journal of Finance, Vol 52, No. 3, 923-973.

[4] Avinash Dixit and Robert Pindyck (1994). *Investment Under Uncertainty*. Princeton University Press, ISBN-10: 0-691-03410-9, 76.

[5] Christophe Ambroise and Geoffrey McLachlan (2002) *selection bias in gene extraction on the basis ogf microarray gene-expression*. Proceedings of the National Academy of Sciences of the United States of America, vol. 99 no. 10, 6562-6566.

[5] Christophe Ambroise and Geoffrey McLachlan (2002) *Selection bias in gene extraction on the basis ogf microarray gene-expression*. Proceedings of the National Academy of Sciences of the United States of America, vol. 99 no. 10, 6562-6566.

[6] Chong Ho Yu (2003). *Resampling methods: Concepts, Applications, and Justification.* Practical Assessment, Research & Evaluation, 8(19).

[7] Davison, A. C. & D. V. Hinkley (1997). *Linear Regression.* Bootstrap Methods and their Application. Cambridge University Press. ISBN-0-521-57471-4.