The Poverty of Academic Finance Research:
Spread trading strategies in the crude oil futures market.

Chrilly Donninger
Chief Scientist, Sibyl-Project

The implication is that most published empirical discoveries in Finance are likely false. [1]

Abstract:
Harvey, Liu and Zhu argue in [4] that probably most of the Cross-Section of Returns literature is garbage. One can always try an additional factor and will find a significant Cross-Sectional result with enough trial and error. Lopez de Prado argues in a series of articles ([1],[2],][3]) in a similar vein. Theoretically scientific results are falsifiable. Practically previous results and publications are checked only in rare occasions. Growth in a Time of Depth by Reinhart-Rogoff ([5]) was the most influential economic paper in recent years. It was published in a top journal. Although the paper contained even trivial Excel-Bugs it took 3 years till the wrong results and the poor methodology was fully revealed ([6]). The reviewers did not check the simple spreadsheets.
This paper analyzes a less prominent example about spread trading in the crude oil futures market by Thorben Lubnau ([7]). The author reports for his very simple strategy a long term Sharpe-Ratios above 3. It is shown that – like for Reinhart-Rogoff – one needs no sophisticated test statistics to falsify the results. The explanation is much simpler: The author has no clue of trading. He used the wrong data.

1) The trading plan:
West Texas Intermediate (WTI) and Brent-Oil are chemically quite similar. But WTI is born in the USA. Export from the USA is banned by law. There is also a bottleneck in pipeline capacity at Cushing, Oklahoma, the main storage for WTI in the US ([7]). WTI is usually traded at a discount to Brent. At the time of this writing (2015-06-10, 5:55 UCT) it is 61.10 to 65.55 for the N5 future. The current price of the Q5 Futures is 61.52 to 66.11 which corresponds to a contango of 0.67 to 0.84%. There are a similar products, but the price is driven by its own market forces. The Futures are therefore interesting candidates for a mean-reverting spread trading strategy.

2) The Data:
The ticker for WTI is CL, for Brent LCO (COIL on Interactive Brokers). There are futures for each month of the year. CL is traded on NYMEX. It expires 3 business days before the 25th of the preceding month (2015-06-22 for N5). The settlement method is physical delivery. COIL is traded on IPE. Expiry is 15 calendar days before the first day of the contract month (2015-06-15 for N5). The settlement method is Cash. Both futures have a multiplier of 1000. A CLN5 contract is currently worth 61.100$. The market is very liquid and the bid-ask spread is usually 0.01.
Graphic-1 shows the price for CL in the last 10 years from 2005-05-27 to 2015-05-27. The red and yellow lines are back-adjusted values. One always trades the nearest future. For the red line the rollover is fixed at every 15th of a month. For the yellow line one rolls over at maximum volume. The rollover strategy follows the market. Rollover has a significant impact on the long run behavior of the time series. In the green line the rollover is also at max. volume, but the data are **unadjusted**. The green line ignores the effect of contango. The graphic shows that a simple long buy-and-rollover strategy can lose even in an upwards trending spot market. Contango is a significant headwind.
Graphic-2 shows the same time-series for LCO. The fixed date rollover is done at the 5th of each month. The effect of contango (actually LCO is sometimes even in backwardation) is for LCO quite different and less severe. This has important implications for a spread trading strategy.

Graphic-1: CL-Futures from 2005-05-27 to 2015-05-27.
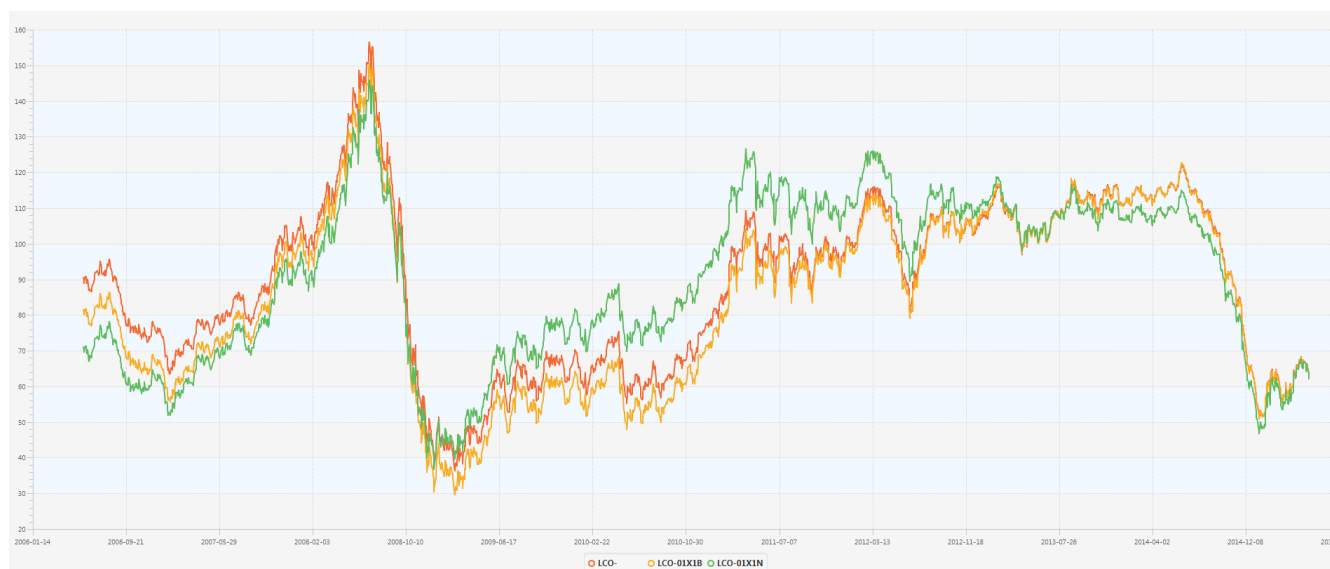
Graphic-2: COIL-Futures from 2005-05-27 to 2015-05-27.

The effect of rolling on the spread is shown in Graphic-3 to 5. Graphic-3 is for the fixed date (15th, 5th) rolling strategy (red-line in Graphic 1 and 2). Graphic-4 for maximum volume (yellow). In Graphic-5 the effect of rolling is ignored, the data are unadjusted (green). If one ignores the effect of contango the time series move till October 2010 in tandem. **The main difference between CL and LCO is the amount of contango.**

Thorsten Lubnau devotes in [7] an own page to the data question. But it is just a summary of well known facts like *"The financial crisis starting in 2008 led to a sharp decline of oil prices"*. The essential question of rolling is not addressed at all. The used time series is shown in figure 1 of the paper (see screen shot after Graphic-5). The figure matches the picture in Graphic-5. They are obviously **unadjusted**.

One can consider the effect of rolling during the course of trading simulation. But everyone with a minimum of quantitative finance experience avoids this nasty and error prone task and uses back-adjusted data. Packages like Unfair Advantage from CSI provide this feature automatically. Lubnau does not mention this point in his further considerations. **He has ignored the effect of rolling at all**.

Graphic-3: CL (red) and LCO (yellow) with fixed date rolling from 2005-05-27 to 2015-05-27.
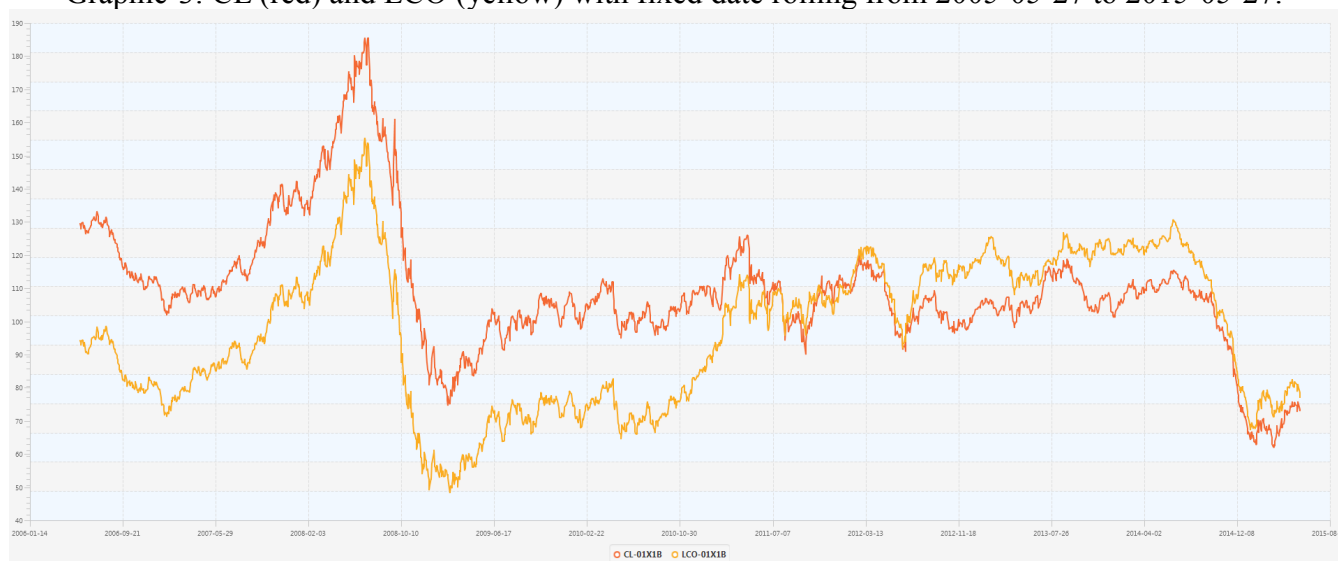
Graphic-4: CL (red) and LCO (yellow) with max. volume rolling from 2005-05-27 to 2015-05-27.

Graphic-5: CL (red) and LCO (yellow) unadjusted from 2005-05-27 to 2015-05-27.

This depressing conclusion is confirmed by the analysis of the trading strategy below. One can argue in favor of a member of the Department of Business Administration at the European University Viadrina in Frankfurt (Oder) that more famous economists from Harvard make stupid Excel-bugs (and nobody notices it).
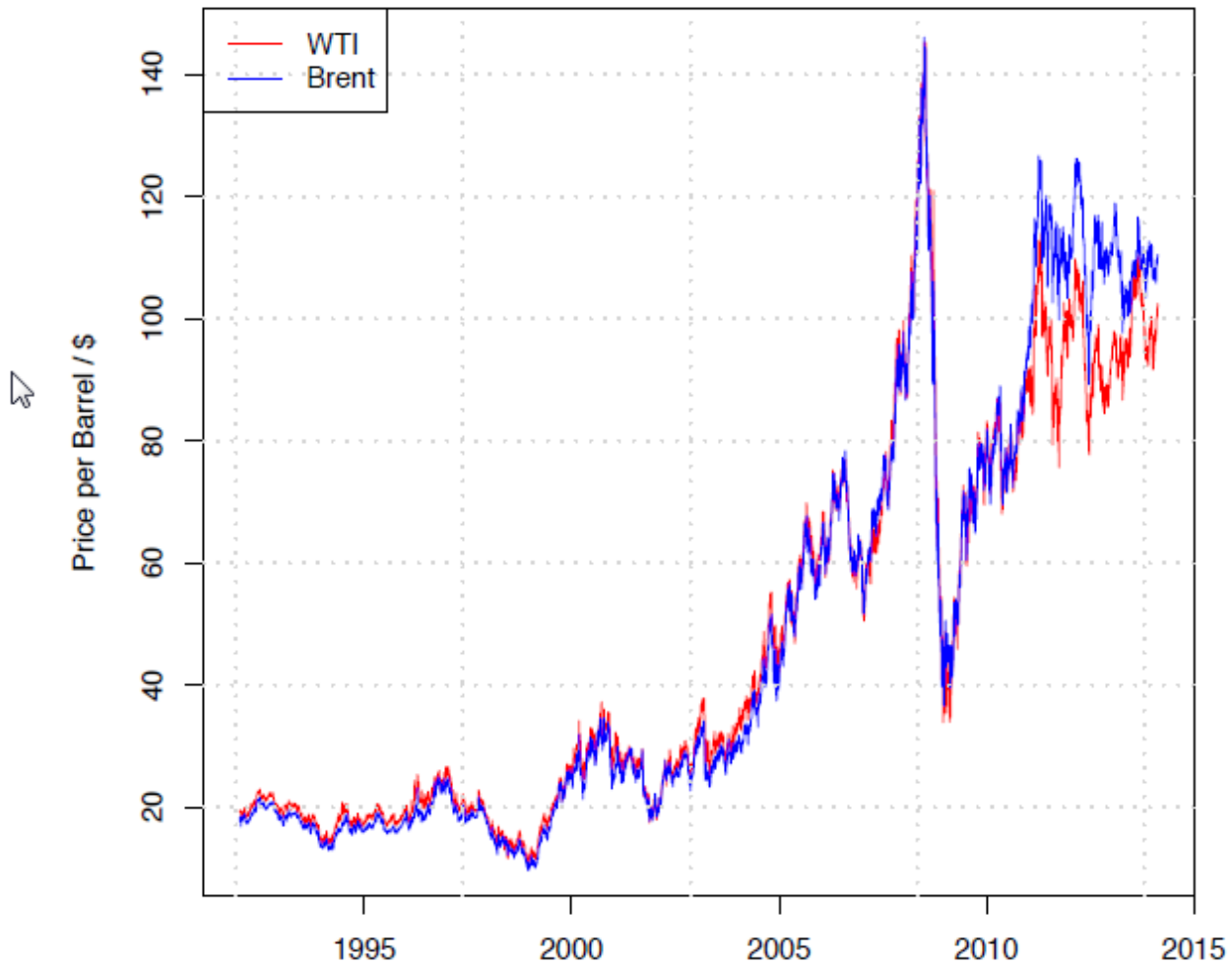
Screenshot of Figure 1 from Lubnau [7].

3) The trading strategy:

As can be seen in Figure 1 Lubnau considers the time range from 1992 to 2015. The time series are initially cointegrated, but the cointegration gets weaker over time.
He defines a CL/LCO long/short portfolio p. From this he defines a signal z as

$$z_t = (p_t - MA_t)/\sigma_t \qquad (1)$$

p_t is the performance of the Portfolio at t. MA_t is a moving average of p over the last k days. σ_t is the volatility of p. Although this value is very critical for the performance of the trading strategy the author gives no details how σ_t is calculated. This is a general rule in the paper. There are a lot of general explanations, references to the literature and all the other redundant stuff of academic papers, but the essential details are missing.
Z_t defines a Bollinger-Band around the CL/LCO portfolio. One goes the portfolio short if $z_t > 2.0$ and long if $z_t < -2.0$. The performance of the portfolio must be reasonable flat. One exploits the swings

around the mean-level. Lubnau starts the historic simulation with 100$ of equity. It is a sad fact of trading live that one can't trade oil futures with 100$ on the broker account. As noted above the nominal value of a future is at current prices 61,000$. The margins are currently approx. 8,000$. The margins are usually not reduced for a spread position, because the short-time VaR of the spread is similar to a single future. Lubnau argues in footnote 5 about private investors. For a private investor, but even for a moderate large hedge fund, the granularity of the futures can't be ignored. There is hence a noticeable error between a theoretical beta and the real one in trading life. A realistic historic simulation has to take this effect into account.

The cointegration has become – even for the unadjusted data – in recent time quite weak. It is hence necessary to adjust the hedging ratio with the following linear regression.

$$R_{B,t} = \beta_{1,t} + \beta_{2,t} * R_{w,t} + \varepsilon_t \quad (2)$$

Where $R_{B,t}$ is the return of the Brent Futures at time t $R_{w,t}$ of the WTI. The author uses the relative involved Kalman filter for this task. He argues that the update has to be done as fast as possible. With the Kalman filter one can calculate $\beta_{2,t}$ recursively. In a footnote he adds *"Given the data processing capabilities even a private investor has at hand, it does not take more than a few seconds to update the hedge ratio"*.

He has obviously not only no clue of trading, but also about the processing power of a modern CPU. His time estimate is wrong by a factor of 10^6. The calculation takes a few microseconds. One can perform for a daily-strategy practically any reasonable calculation without worrying about the time delay. The Kalman filter was developed for real time object tracking. In the realm of finance the argument is only valid for HFT or for a fund which trades thousands of different assets. But this is completely off for a private investor.

Th. Lubnau needs the relative involved Kalman filter to obfuscate how and why he gets impressive results. The Sharpe ratio is – over a period of 5 years – from 1992 to 2015 always above 3. Everybody who has some basic experience in trading knows that this is – even for a backtest – too good to be true. Actually this was the reason why the paper was sent from a buddy-trader. *"Hi Chrilly, a consistent Sharpe ratio of 3+ is impossible. Could you please find out what is wrong"*.

His estimator is – in contrast to the claims in the footnote – also rather time consuming. The performance of the plain Kalman filter breaks down in the last 5 years period. He resorts to a computationally relative involved ML estimator which gives totally different (and erratic) results than the plain Kalman. See Figure 2 from [7] on the left.
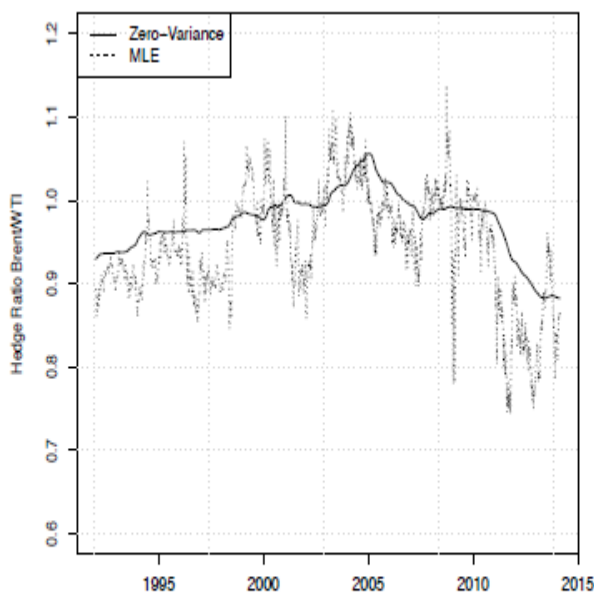
Figure 2: Dynamic Hedge Ratios of WTI and Brent using Kalman Filtering with zero variance (solid line) and MLE variance (dotted line) in the state equation 1992-2013

Graphic-6 shows the rolling calculation of beta from equation (2) with the very robust and efficient

Theil-Sen estimator (see [8]). The estimation window is 126 trading days. The red line is for the fixed rolling date time series, the yellow for the max. volume rolling, the green line for the unadjusted data. The beta of the unadjusted data is much more stable. In Graphic-7 the rolling window length is reduced to 21 days. Beta is like for the ML estimator of Lubnau (dotted line in Fig. 2) quite erratic. But still the behavior of the green line is much "nicer".
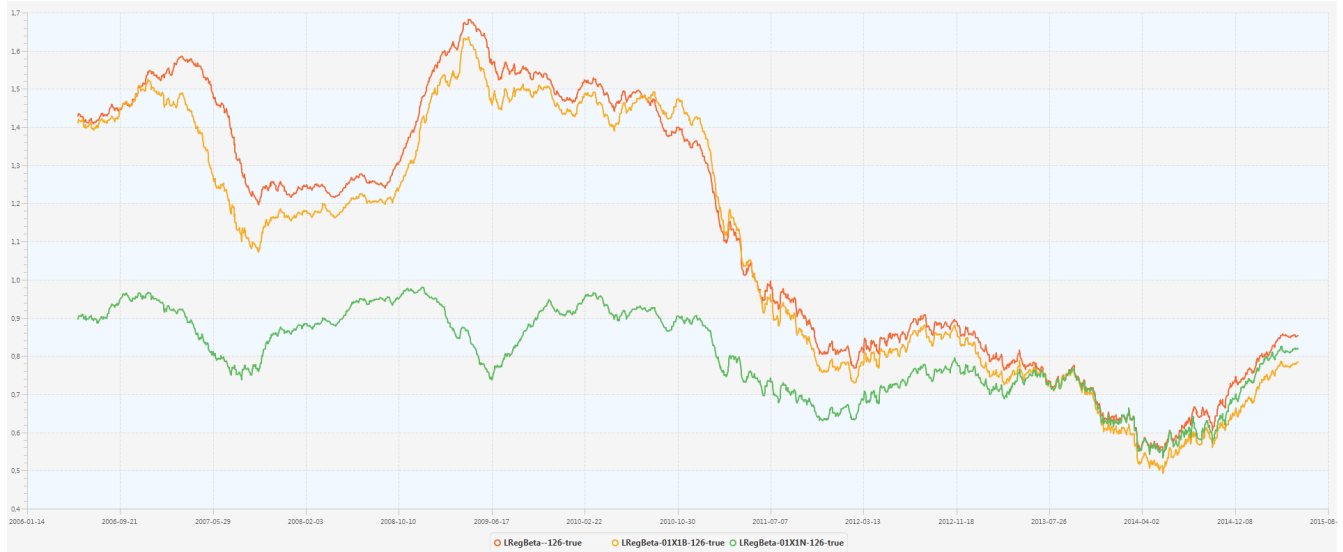
Graphic-6: Rolling beta with Theil-Sen estimator of 126 trading days from 2005-05-27 to 2015-05-27.
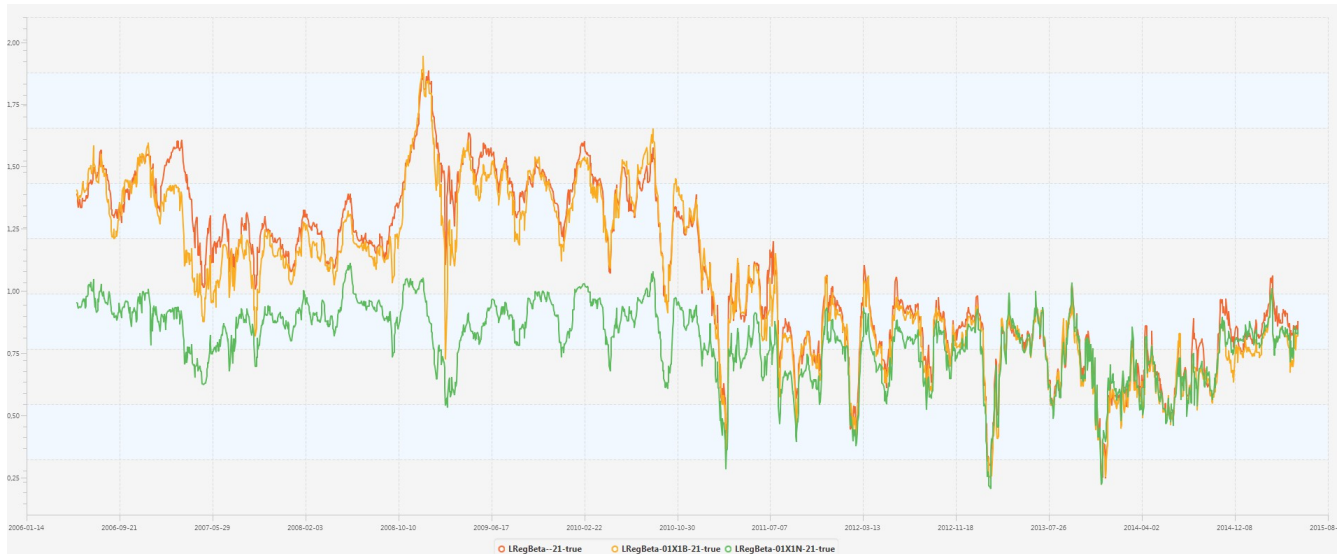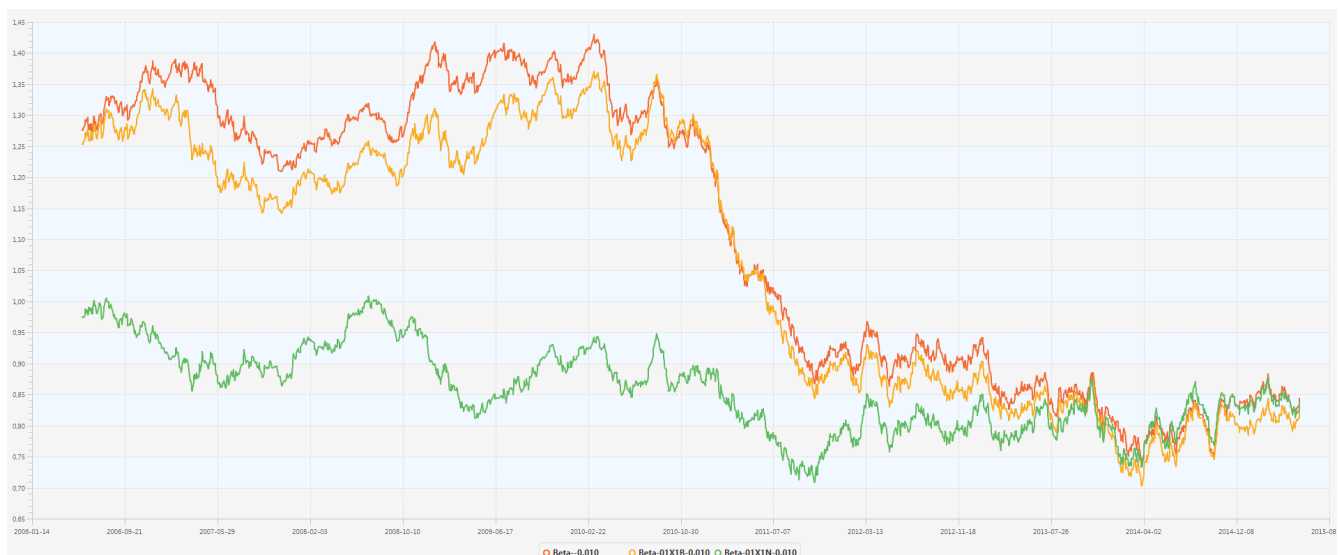

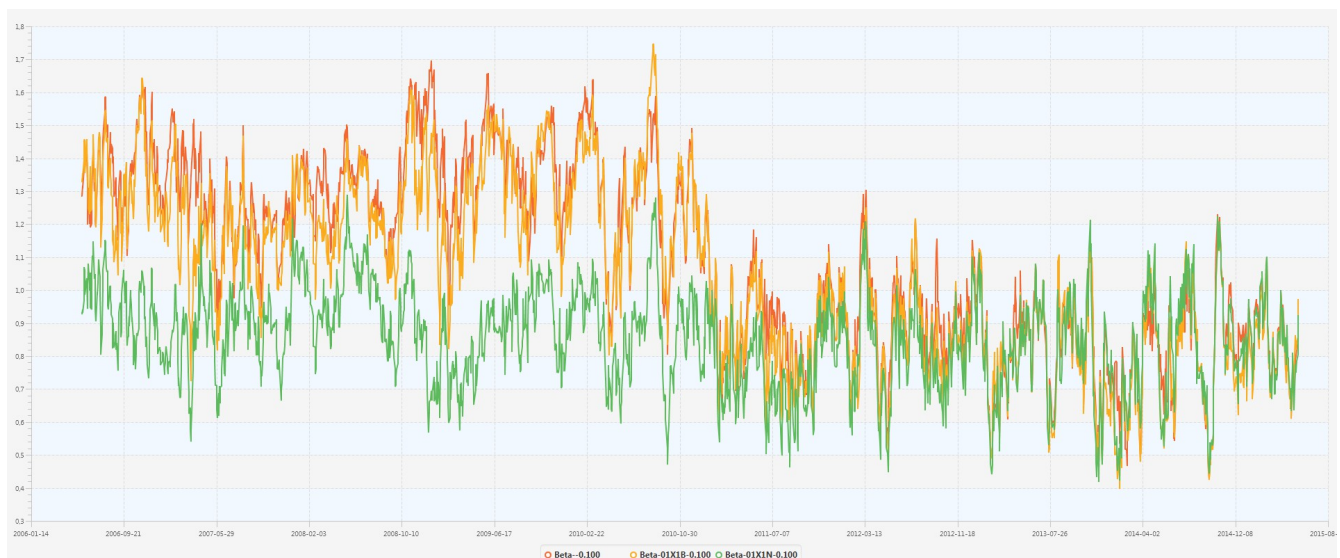Graphic-7: Rolling beta with Theil-Sen estimator of 21 trading days from 2005-05-27 to 2015-05-27.

Graphic-8 shows the calculation of Beta with an exponential filter.

$$\text{Beta}_t = \text{Beta}_{t-1} + K*((R_{B,t}/R_{W,t}) - \text{Beta}_{t-1}) \quad (3)$$

One updates Beta with the error term of the current return ratio. K is the so called gain. It is the impact of the last value. In signal processing (3) is called the adaptive LMS algorithm. It is the working horse in many signal processing applications ([9]). Equation (3) can also be formulated in a state-space representation ([10]). The gain K is fixed in equation (3), whereas in the Kalman filter K is calculated dynamically. K depends for the Kalman filter on the assumptions of the measurement and update noise.

Graphic-8: Beta with Exponential Filter K=0.01 from 2005-05-27 to 2015-05-27.


Graphic-9: Beta with Exponential Filter K=0.1 from 2005-05-27 to 2015-05-27.

The behavior of the Theil-Sen estimator with a window length of 126 (Graphic-6) is quite similar to the results of the adaptive LMS filter with K=0.01 (Graphic-8). For K=0.1 beta is for the LMS estimator even more erratic than the robust Theil-Sen estimator with a window length of 21. The robust method reacts much less to outliers. As we want to follow the non-stationary spread the erratic behavior is not necessarily a bad feature. Lubnau has introduced the MLE version for exactly this reason. The plain Kalman filter smooths Beta too much.

Another approach is to calculate the volatility ratio:

$$\text{Beta}_t = \text{Vola}_{B,t} / \text{Vola}_{W,t} \quad (4)$$

Where Vola is the volatility of the Brent or WTI Future with a given window length. The volatility is calculated with the Yang-Zhang volatility estimator. In Graphic-10 the window length is set to 126 trading days. The fixed date rolling strategy behaves smoother than max. volume rolling. But the overall picture is still the same. I have calculated numerous other measures with similar results.
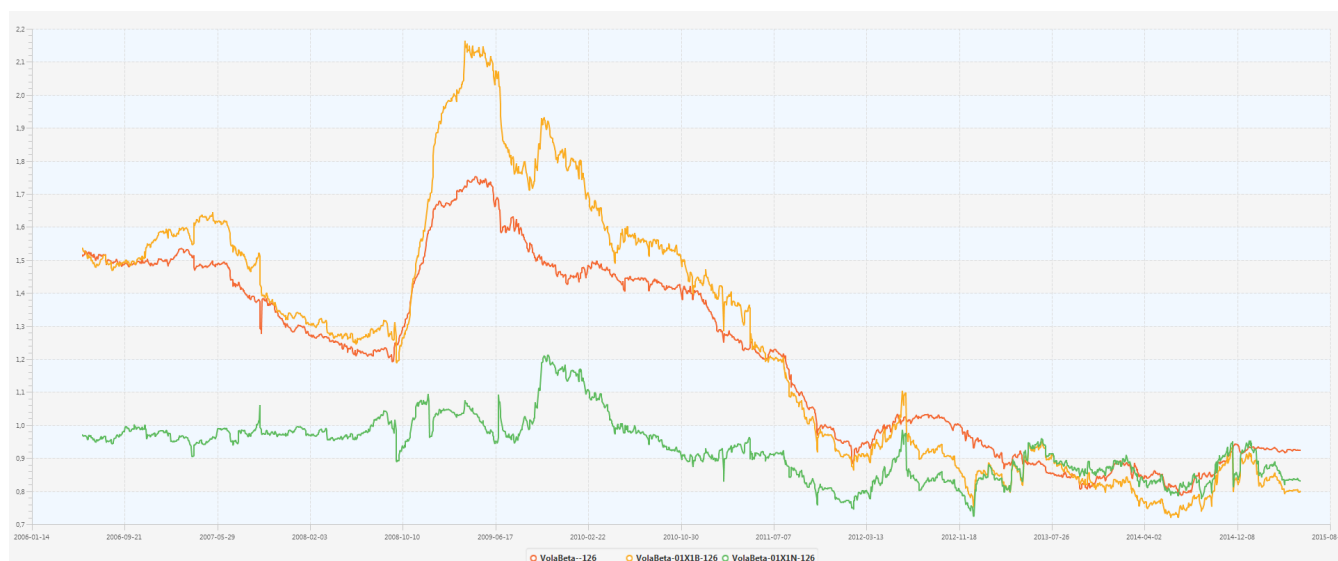
Graphic-10: Rolling beta with Volatility-Ratio of 126 trading days from 2005-05-27 to 2015-05-27.
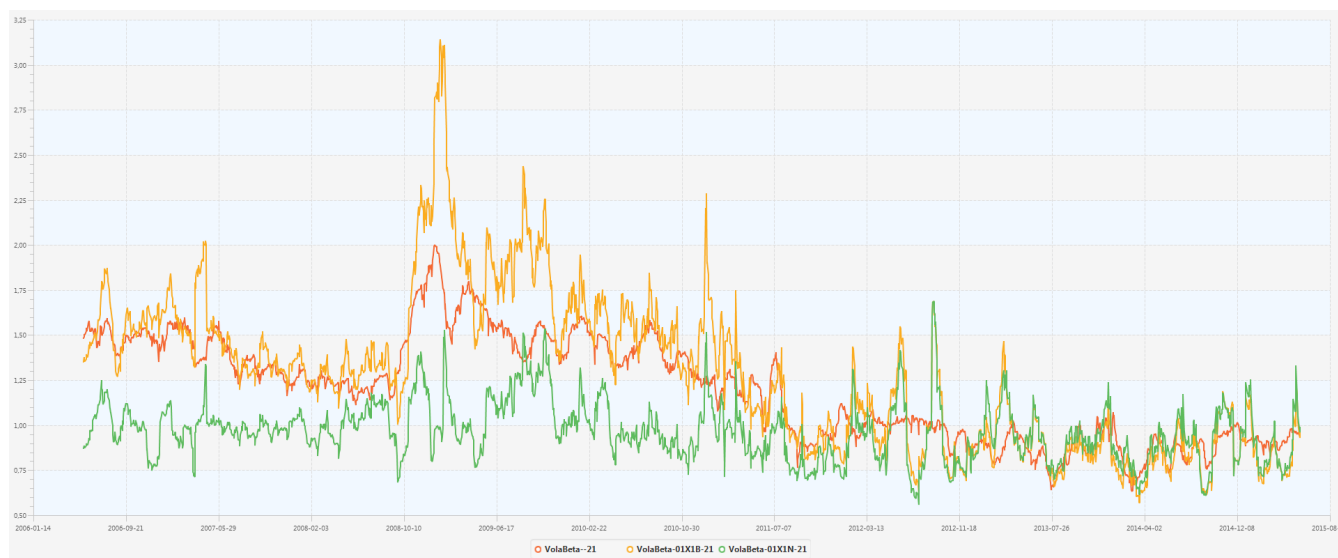
Graphic-11: Rolling beta with Volatility-Ratio of 21 trading days from 2005-05-27 to 2015-05-27.

3.1) The performance of the spread-portfolio:

To simulate a realistic behavior it is assumed that one has a starting value of 1,000,000$. This amount is leveraged by a factor of 2. The current value of a CLN5 contract is 65,550$. One would hence go round(2,000,000/65,550) = 31 CLN5 futures short. Going CL short is just a convention, it has an impact on the spread-portfolio, but only a small effect on the mean-reverting strategy. The calculation of the long LCO futures is analog, but one has to take the beta into account. If the index (cash on the broker account) rises (falls) the number of traded contracts is increased or decreased. It depends of course also on the price level of the futures. The spread portfolio is like in [7] rebalanced every day on the close. There is one major difference. Only an integral number of futures can be traded. So there is sometimes no re-balancing action when beta changes.
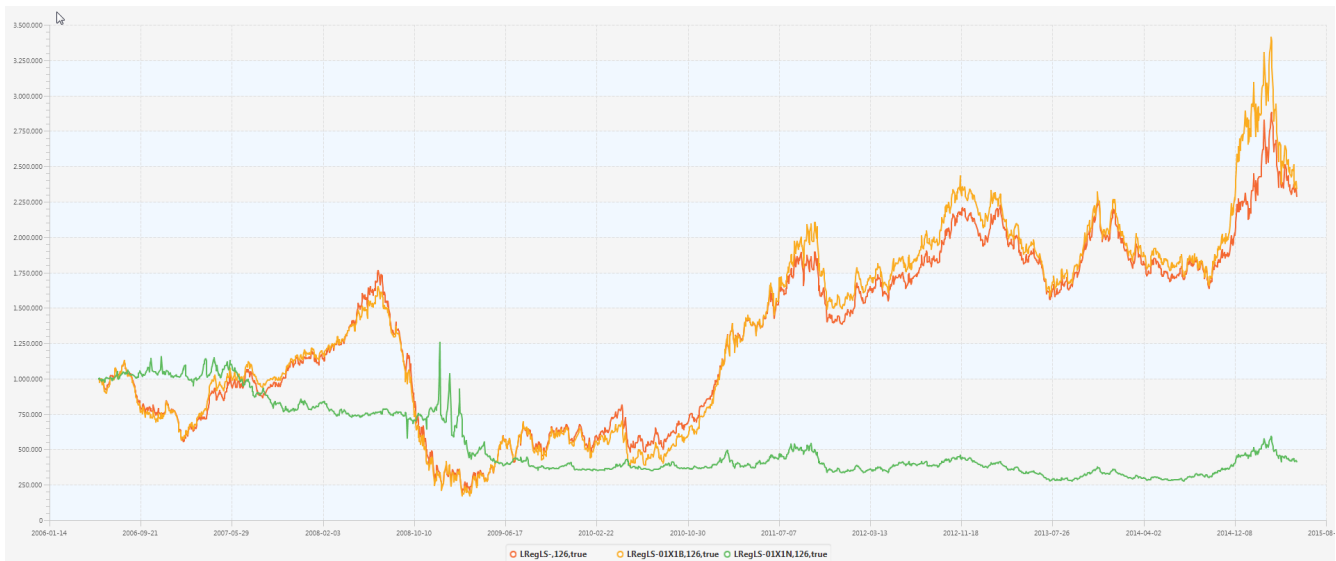
Graphic-12: Spread-Portfolio with Theil-Sen beta of 126 trading days from 2005-05-27 to 2015-05-27.

Graphic-12 shows the performance of the Spread-Portfolio for the Theil-Sen beta with a window length of 126 trading days. The red line is for the fixed date rolling, yellow for max. volume rolling and the green line is for the unadjusted time series. It is clear that the red and yellow lines are not suited for a mean-reverting strategy. There are long up- and down-trends which dominate the short term swings. One could even think about a trend following strategy (see paragraph 4 below). The green line is downwards trending. But overall it is in comparison relative smooth and exploiting the swings around this line could work. The nasty time for the strategy is the 2008 crash. It is a matter of good luck to be on the correct side during this up- and down- spikes.

If one calculates the beta with a 21 trading days rolling window one gets for the Theil-Sen estimator the results of Graphic-13. Although beta is now quite erratic the overall pattern is similar. No matter how beta is calculated: **One can't create a stable mean-reverting portfolio if the spread is not cointegrated.**
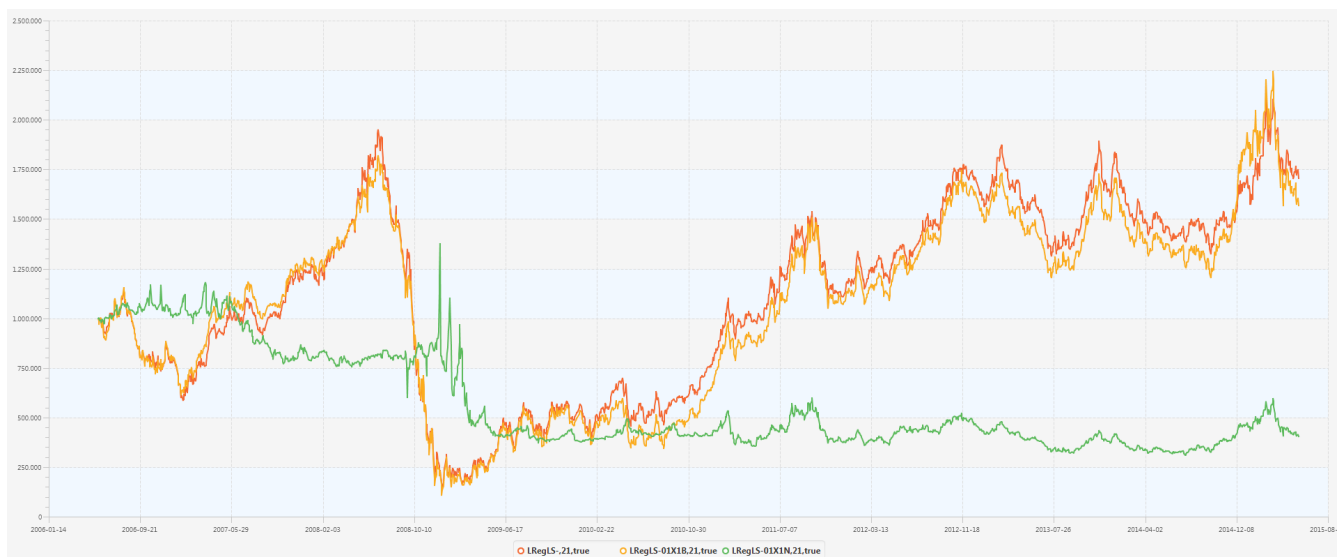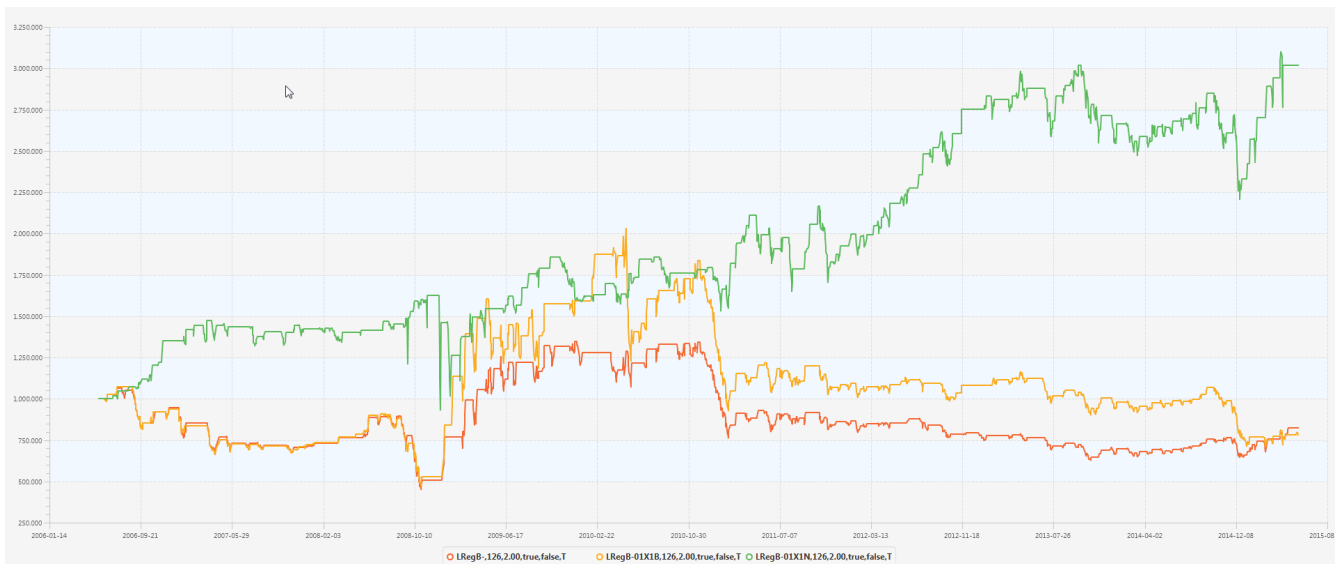
Graphic-13: Spread-Portfolio with Theil-Sen beta of 21 trading days from 2005-05-27 to 2015-05-27.

## 3.2) The performance of the mean-reverting strategy:

The spread-portfolio in paragraph 3.1 is only a theoretical construct to calculate the z-value in equation (1). If the z-value is above 2.0 one goes the portfolio short (reverts the sign of CL and LCO), if it is below -2.0, on goes the portfolio long. Once the z-value crosses 0, the position is closed. It was noted above that the calculation of σ_t is not specified in the Lubnau paper. Here it is calculated as the standard-deviation of the spread-portfolio with the same window-length as the moving average.
It is reported in [7] that the best setting for the Moving-Average MA_t is 20 trading days. I got slightly better results for a moving average of 10 days.
All indexes start with 1,000,000$. The fixed-date rolling time series (red) in Graphic-14 ends with a final value of 821,750$. The max. volume rolling time series (yellow) does in between considerable better, but the final result is with 781,020$ even worse. The different behavior in between is due to good/bad luck. There are large up- and downs. The swings within the Bollinger band are just noise. If the noise hits the band on the right side, the performance jumps up, if it is on the wrong side, the strategy nosedives. The green unadjusted time-series does much better. The final value is 3,019,900$. The overall win is +201,9%. The Sharpe-Ratio is 0.61 and the max. relative drawdown is 42.9%.This is not bad, but it is far from the > 3.0 Sharpe ratio claimed by Th. Lubnau. One can certainly find better (more lucky) beta and σ_t calculations. Lubnau uses the powerful dlm-R package. This package provides an almost unlimited bag of calculations for getting the right beta. If plain Kalman deteriorates, viola, use the MLE estimator, if this does not work, reformulate the measurement equation ...

Graphic-15 shows the performance when beta is calculated with double exponential smoothing. K is set to 0.002. This gives a relative smooth beta, but the double exponential smoother follows trends for beta faster. The green line has a final value of 4,037,200$ (+303.7%) with a Sharpe-Ratio of 0.67. But no matter how sophisticated a lucky beta is calculated. **One can't trade the strategy.** The realistic alternatives are the poor performing red and yellow lines in Graphic-15.
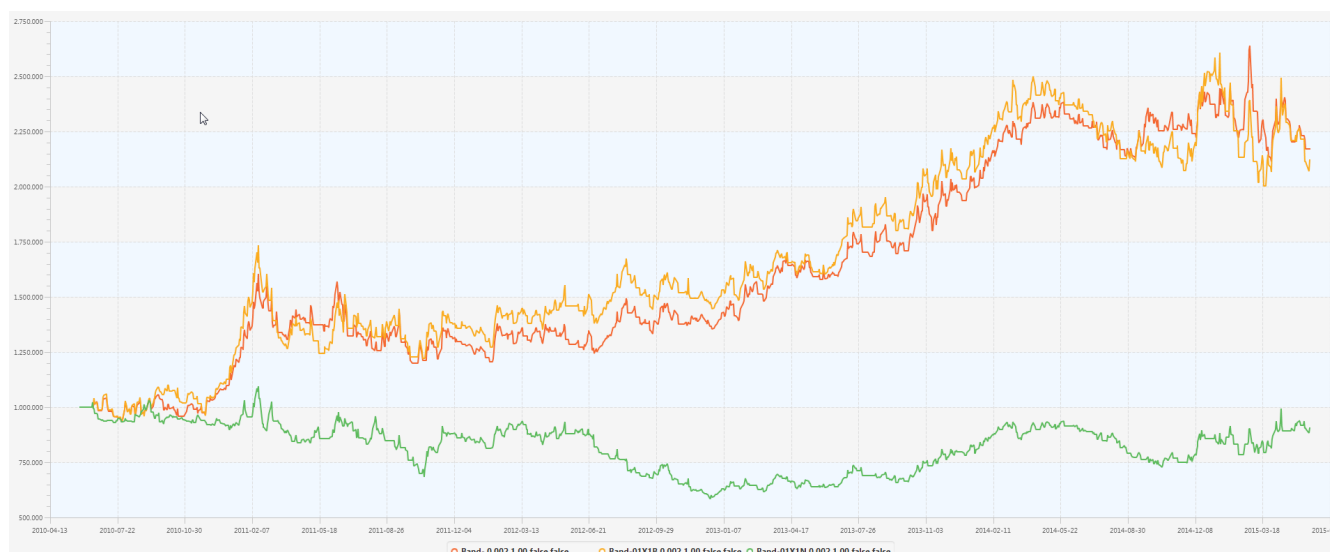
Graphic-14: Mean-Reverting-Portfolio with Theil-Sen beta 126 from 2005-05-27 to 2015-05-27.
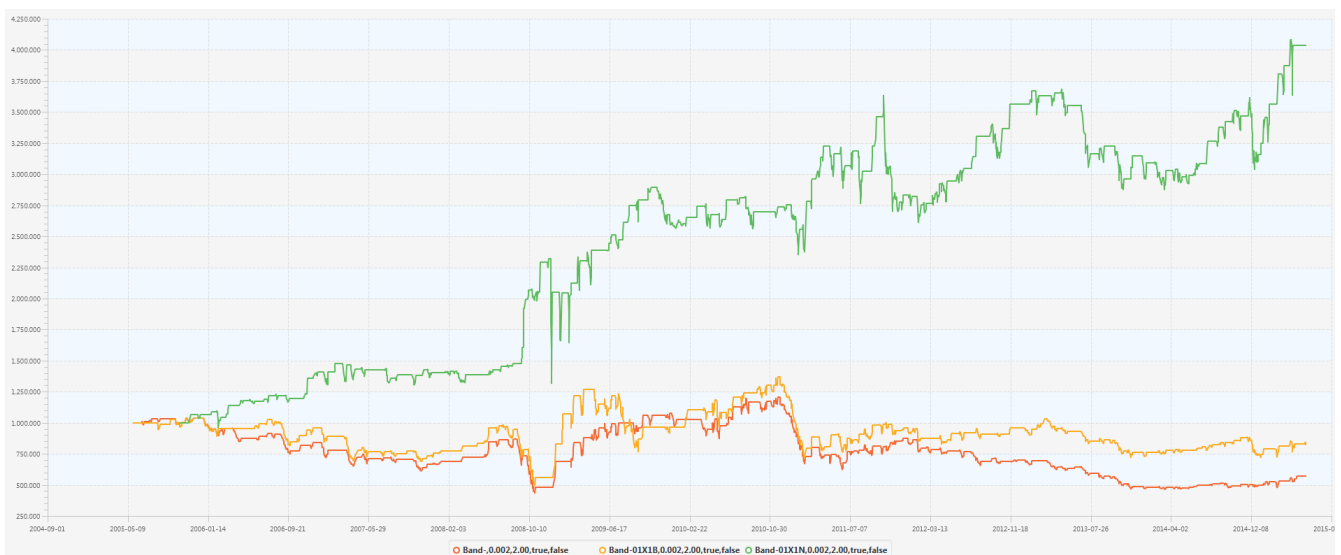

Graphic-15: Mean-Reverting-Portfolio with Double-Exponential beta from 2005-05-27 to 2015-05-27.

4) The trend is your friend:

As can be seen in Graphic-12 and 13 the spread portfolio has significant trends in the last 5 years. Instead of the mean-reverting strategy proposed in [7] one could try on the adjusted data a trend-follower. One goes the portfolio long, if z_t is above 1.0 and shorts the portfolio if z_t falls below -1.0. The position is closed, if the z-value crosses the zero-line. Graphic-16 shows the performance of this strategy from 2010-05-27 till 2015-10-27. One starts as before with 1,000,000$. The final value is for the red line 2,170,020$ (+117.0%) with a Sharpe-Ratio of 0.70. The performance of the yellow line is similar. The different roll-modes have only a minor impact. The strategy does not work for the unadjusted time series. The strategy looks for the last 5 years relative attractive. As can be seen in Graphic-17 it works also reasonable from 2005 till September 2008. But it capsizes in the 2008 crash. There are obviously different regimes. This is maybe one of the reasons the spread and the different behavior exists at all.

Graphic-16: Trend-Portfolio with Double-Exponential beta from 2010-05-27 to 2015-05-27.
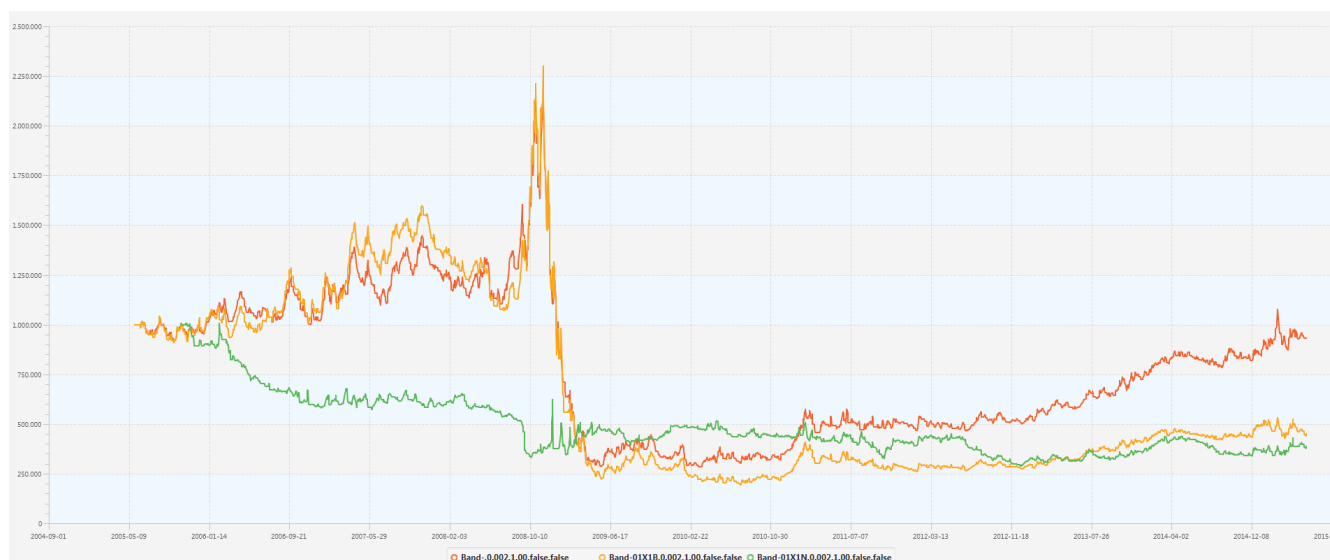
Graphic-17: Trend-Portfolio with Double-Exponential beta from 2005-05-27 to 2015-05-27.

Conclusion:

The impressive results of Th. Lubnau can be explained by the embarrassing fact that the author has ignored the contango of oil-futures. Additionally he has tweaked the calculation to get impressive Sharpe-Ratios. Actually this fact shows already that he has no clue what he is writing about. The results are too good to be true and everyone with a minimum understanding of the trading profession gets suspicious when he reads about a constant Sharpe ratio of 3+ over 23 years. But academic paper does not plush. I have read in my previous profession as a computer-chess professional similar nonsense in the journal of the international computer chess association. Actually there was in the 15 years I was following the journal (one had to subscribe if one wanted to enter a competition organized by the association) only one practical relevant article. But my personal enlightenment was even in this case limited. I was the author ([12]). The original title was "Null move and deep search: Selective-search heuristics for stupid chess programs". The editor considered "stupid" too less scientific and obfuscated it – without consultation – to "obtuse". But stupid was exactly the point of the algorithm. There was no

chess-knowledge involved. This was in complete contrast to the academic claim that computer chess is the drosophila of AI. Actually the claim was that computer chess is the drosophila of the human mind. No professional chess programmer ever took this claim serious. It was all about speed and searching as deep as possible. The communication by academic paper was just noise. But there was an intensive communication by competition within the top programmers. If a new version of a program was published, one played hundreds of games against it and analyzed the behavior. Sometimes one did a complete reverse engineering and disassembled the opponents program. From this experience I coined Chrilly's computer chess law:

Those who publish, know nothing.
Those who know, publish nothing.

The situation is maybe not as bad in quantitative finance. But I have the impression that one can find similar flaws in a lot of academic finance papers. Thorsten Lubenau had the bad luck that I stumbled upon his article. I don't claim that he was cheating. His paper is simply obtuse.

Final Note:

While preparing this article I have watched en passant the real-time behavior of CL and LCO. At least during this very short-time span the time-series show an interesting intraday mean-reverting behavior. As the bid-ask spread is low and liquidity is very high it could be interesting to investigate this behavior further.

Acknowledgment:

Discussions with Johannes Bruski and Jo Fritz were the initial starting point of this paper. Both were also helpful in data preparation. Johannes and Jo do not necessarily share the same view about the Lubnau paper and the academic world in general.

References:
[1] Lopez de Prado M.: Illegitimate Science: Why Most Empirical Discoveries in Finance are Likely Wrong, and What Can Be Done About It. May 1, 2015.
[2] Bailey D., Borwein J., Lopez de Prado M., Zhu Q.: Pseudo-Mathematics and Financial Charlatanism: The Effects of Backtest Overfitting on Out-Of-Sample Performance.
[3] Bailey D. et. al.: Backtest overfitting demonstration tool: an online interface, April 2015
[4] Harvey C., Liu Y., Zhu Heqing: … and the Cross-Section of Expected Returns, April 2015
[5] Reinhart C., Rogoff S.: Growth in a Time of Depth. American Economic Review, 100 (2), 2010
[6] Herndorn Th., Ash M., Pollin R.: Does High Public Dept Consistently Stifle Economic Growth? A Critique of Reinhart and Rogoff. April 2013
[7] Lubnau Th.: Spread trading strategies in the crude oil futures market. April 2014
[8] Wilcox R.: Introduction to Robust Estimation, Chap. 10.2 Theil–Sen Estimator, Academic Press, 2005.
[9] Ifeachor E., Jervis B.: Digital Signal Processing, A Practical Approach, 2nd Ed. Prentice Hall, 2002
[10] Hyndman R., Koehler A., Ord J., Snyder R.: Forecasting with Exponential Smoothing, The State Space Approach, Springer 2008
[11] Yang D., Zhang Q.: Drift-Independent Volatility Estimation Based on High, Low, Open and Close-Prices. Journal of Business, 2000, Vol. 73, No. 3.
[12] Ch. Donninger: Null move and deep search: Selective-search heuristics for obtuse chess programs. ICCA Journal, Vol. 16, No. 3, 1993