Is Daily Pairs Trading of ETF-Stocks profitable?
Chrilly Donninger
Chief Scientist, Sibyl-Project
Sibyl-Working-Paper, July 2016

http://www.godotfinance.com/

Economics is extremely useful as a form of employment for economists.
(John Kenneth Galbraith)

Abstract:
Pairs trading is a venerable trading strategy. There is agreement that it worked fine in the far past. But it is less clear if it still profitable today. In this working paper the universe of eligible pairs is defined by the holdings of a given ETF. It is shown that the stocks must be from ETFs which select high-quality, low-volatility stocks. The usual closeness measure presented in the literature performs poor. The paper presents a simple and clearly superior alternative based on zero-crossings. The strategy performs with the correct universe and the improved pairs selection rule before trading costs quite fine. It depends on the assumed trading costs if this is also in real-trading life the case.

Introduction:
The seminal paper on pairs trading is Gatev et. al. [1]. They authors did not invent the strategy. It was in common use since the 1980s. The pairs are formed from a universe of stocks. There is a one year formation period. Each stock is normalized to 1 at the beginning of this period. One selects for each stock the closest neighbor. The distance measure is the summed up squared daily difference of the normalized prices. From this closest-pairs one selects the best 10 or 20 and uses them in the trading period. If the spread moves out of a 2 standard-deviation band one enters a contrarian trade. The position is closed once the mean of the spread is reached again. A pair can be traded several times during the trading period. In [1] the trading period is set to 6 months. A pairs stays for this amount of time within the trading-pool. But the formation-period calculation is done every month. There are 6 overlapping selection pools. If a position has not closed at the end of the trading period, it is closed. There is otherwise no stop-loss rule. The spread is set to 0 at the beginning of the trading period. The strategy never trades immediately after the formation period.
Gatev et al. report nice profits in the long run. But the profits are declining in the last period of their study. Do&Faff [2] get in 2010 similar results. The distance method is a shirt-sleeved method for determining the best pairs. A more statistical sound approach is cointegration. Cointegration was proposed in a popular trading book by Vidyamurthy [3] and various other papers. Rad, Low and Faff investigate in a recent publication [4] the performance of the distance, the cointegration and a sophisticated copula approach. The performance of the distance and cointegration methods are similar, the copula is clearly worse. Similar studies were done by Perlin for the Brazilian market [5] and by Karvinen [6]. The later authors do not set the spread to zero at the beginning of the trading period. There is no gap between the formation and the trading period. Additionally the stock prices are normalized by subtracting the mean and dividing by the standard-deviation of
the formation period. The normalized stocks have zero mean and a standard deviation of 1. But they are usually not N(0,1) distributed. The normalization is during the trading period the same. Karvinen takes also analyst recommendations into account. A pair is excluded, if the recommendations are conflicting. Or with other words the recommendations are a part of the distance measure. This should solve a principle problem of pairs trading. There is usually no intrinsic reason that the gap closes again. The situation is in this case similar to a divorce. It gets very expensive if the relation of a pair breaks apart. Including the recommendations should reduce the probability of divorces.

**The data:**
I downloaded the recent holdings of industry sector ETFs like IYE, IYF, ITB, IBB, IDU or XLU and general style ETFs like OEF, QQQ, DIA, USMV, SPLV from the ETF providers. No attempt was made to reconstruct the historical holdings of these ETFs (retrieving and processing the current holdings is cumbersome enough). The historic simulation is from 2011-01-01 to the current date 2016-07-26. Some of the stocks were during these period not a member of the ETF. The historic simulation has no notion of this. It is not in the strict sense a simulation of e.g. the IYE-stocks. Some of the stocks were even not traded from the beginning. The universe is in most cases at the beginning of the simulation period smaller than the current one. Some stocks which were initially in the ETF are not selected any more. One has a survival bias. This bias is difficult to quantify. The changing weights of the stocks have no effect. The strategy ignores this information. Besides the stocks the ETF itself can also be a part of a pair. All data are downloaded from yahoo. The calculation uses the adjusted closing prices. This avoids the nasty handling of stock splits but it introduces some small Profit&Loss errors due to dividend adjustments. This effect is certainly minor in comparison to the assumptions about the trading costs.

**The Strategy:**
The initial results with the distance method were rather disappointing. Pairs trading is based on mean-reversion. The distance measures if the stocks stick together. But sticking together and mean-reversion are two different concepts. Vidyamurthy proposes in [3] zero-crossings as an alternative. One counts the number of times the spread moved above or below the mean-spread. But this measure is also not satisfactory. It is known from the theory of Brownian-motions that zero-crossings are much more likely in the first few steps of the motion. If one starts at zero a small up- followed by a larger down-move is a zero crossing. The path moves in the following away from zero and a crossing gets very unlikely. The situation is somewhat different for a mean-reverting process but the general behavior is still the same. A zero (or mean) crossing does also not create a profit. The interesting case is a crossing which started initially outside the two-sigma band. This is the main distance function. A larger number of crossings is of course better than a lower one. For two pairs with the same number of crossing the distance is used as a secondary measure. But a pair with 5 crossings is always closer than a pair with only 4. The strategy defines also a minimum number of crossings (usually 4). A pair with less crossings is never traded.
The strategy does not use overlapping formation periods. The set of tradeable pairs is determined each month (every 21 trading days). The formation window is like in most studies a year (252 trading days). But an open position is not automatically closed at the end of the trading period. An open position is – if mean reversion does not happen before – closed after 30 trading days. There are usually pairs from the previous formation period open. It makes no sense to close a position which was entered at day 20 of the trading period just because a new formation calculation is performed. The strategy does not reset the spread to zero at the end of the formation phase. It uses the mean and the standard deviation from the formation period also in the trading phase. A position is only opened, if the spread is between 2 and 4 standard deviations. It is unlikely that the spread is by chance larger than 4 deviations. A very large spread is a sign that the pair is in divorce. As an additional stop-loss an already open position is closed if the spread gets larger than 8 standard deviations. This stop-loss is only triggered a few times but it avoids some really disastrous losses.

Gatev et al. normalize the prices at the beginning of the formation period. I tried two normalizations. The first uses the median of the first 5 days. The other the mean over the whole formation period. The mean performs in most cases slightly better. But normalizing with the standard-deviation like in [5] and [6] is clearly worse.

**Performance:**

As already noted simulated trading is done from 2011-01-01 till 2016-07-26. One starts with an initial cash of 500.000$. This is a convention I use in all my trading papers. In the standard setting a short and long position of 10% of the current cash is entered. Initially one buys and sells stocks for 50.000$ each. Graphic-1 shows the performance of the ETF USMV (iShares Minimum Volatility US). The red line is the performance if the pairs are sorted according crossings (and secondary with distance). A pair is only selected if it has at least 4 crossings in the formation period. There are up to 20 pairs considered. The USMV contains 170 stocks plus the ETF itself which is traded as a plain stock. There are hence 170 possible pairs. But due to the 4 crossings filter there can be less than 20 pairs in the selection. The strategy has an overall profit of 144.4%, a monthly Sharpe ratio of 1.16 and a max. relative drawdown of 8.2%.

![Graphic-1: Performance of USMV, Crossings-Sorted (red), Distance (yellow)](image-url)
The performance is almost as fine as the funds of Bernie Madoff. But this is without trading costs (which will be considered later). The yellow line shows the erratic performance of a purely distance based selection criterion. There are always 20 pairs considered. The performance is sometimes spectacular, sometimes disastrous. At the end there is a loss of -20%. There are other ETFs where the situation is not as bad. But the crossings based selection process is always much better. One can improve the performance by setting an additional threshold of 4 crossings and sorting the pairs which pass this filter by distance. This almost boils down to the first method. There are only few pairs with more than 4 crossings. The standard case of 4 crossings is sorted in both cases in the same way. It is usually argued that Pairs-Trading is beta-neutral. This is true for the left side of the graph. There was a severe crash in August 2011. One does not notice this crash in the performance of the red-line. The yellow line shows at this time a marked drop. But the effect of the short term Brexit-crash is clearly visible at the right end of the red graph. The Brexit effect disappears if one selects only the top 10 pairs. But in this case the overall profit – and also the Sharpe-ratio – is lower. The lower number of pairs increases the overall volatility. But there are no pairs which are hit during the Brexit. It is simply good or bad luck how the strategy performs during a crash. The performance of the working pairs strategies is usually smoother than the long ETF.

Graphic-2 shows in red the performance of the SPLV holdings, of the XRLV (yellow) and QQQ (green). The parameters of the first two ETFs are the same than for the USMV. The QQQ shows a slightly better performance if one sets the minimum crossings threshold to 3.

The SPLV (S&P 500 Low Volatility) ETF is a selection of 100 low-volatility stocks from the S&P-500. The USMV selects from a larger universe. But there is a great deal of overlap between the SPLV and the USMV. The weights are also calculated differently. The strategy pairs are independent form the ETF-weights. Each pair gets the same weight. I tried also a weighting proportional to the crossings. E.g. allocate 2% per crossing. The difference in performance is minor.

Graphic-2: Performance of SPLV (red), XLRV (yellow) and QQQ (green)
The XRLV (PowerShares S&P 500 ex-Rate Sensitive Low Volatility) is similar to the SPLV. It contains also 100 low volatility stocks from the S&P-500 universe. But the selection process is somewhat different. The QQQ is the Nasdaq-100 ETF. The SPLV has a P&L of 94.3% a Sharpe ratio of 1.06 and max. relative drawdown of 6.2%. The XRLV a P&L of 79.3%, a Sharpe ratio of 0.99 and a drawdown of 8%. The QQQ a P&L of 112.9%, a Sharpe ration of 0.8 and a drawdown of 16.2%. None of the pairs are hit by the Brexit. But the XRLV and the QQQ show some losses during the Aug. 2011 crash. The QQQ performance is especially strong during the weak market of Oct. 2015. High volatility can – but must not - be a beneficial market regime.

Graphic-3 shows in red the performance of the SPMO (PowerShares S&P 500 Momentum), yellow is the SPVU (PowerShares S&P 500 Value) and green the SPHB (PowerShares S&P 500 High-Beta). Each ETF has a holding of 100 stocks. The minimum number of crossings is set to 3. If one sets – like for the USMV – the crossings threshold to 4 there are too few pairs selected. The SPMO has a P&L of 66.7%, a Sharpe ratio of 0.81 and a drawdown of 10.7%. The SPVU a P&L of 48.1%, a Sharp ratio of 0.51 and a drawdown of 13.1%. The SPHB is with a loss of -40.2% simply a disaster.
This is a general pattern. The performance of the strategy depends on the quality of the stocks in the ETF.

Graphic-3: Performane of SPMO (red), SPVU (yellow) and SPHB (green)

There is interestingly also a strong correlation between the ETF performance and the strategy. Graphic-4 shows the performance of the ETFs. The time range is from 2012-01-01 till 2016-07-26. (The USMV was introduced at 2011-10-20). The USMV has with 1.07% the highest Sharpe ratio. Followed by the SPLV and the QQQ. The SPHB performs fine till the end of 2014 but is since that time much worse. SPHB pairs trading does not work over the whole period. The pairs-relation is not stable enough. Karvinen tries to solve this problem in [6] with the analysts recommendation. But a much simpler – and also better working – approach is to rely on the selection process of the ETF. There are no junk-stocks in the USMV.
Another classic approach is to select the stocks according to industry. The stocks within a sector should perform more uniform. This is another way to tackle the divorce problem.

Graphic-5 shows in red the performance of IDU (iShares US Utilities), yellow is the ITB (iShares US Home Construction) and green the IYE (iShares US Energy). The ITB performs best with a P&L of 152.6%, a Sharpe ratio of 0.86 and a drawdown of 15.1%. The IDU has a P&L of 46.1%, a Sharpe ratio of 0.86 and a drawdown of 12.4%. So from the Sharpe ratio the IDU and the ITB are the same. One could boost up the IDU by investing a higher percentage in each pair. The IYE does not work at all. I tried several other industry sectors. But they perform either weak or as bad as the IYE. The IDU and the ITB are the only reasonable sectors I could identify. This is inline with the results in [4]. There is too much junk in most industrial ETFs. The utilities and homebuilder are relative stable sectors.
Joining ETFs:
The holdings in the USMV and SPLV are selected by a similar criterion: Low volatility. The union which I call SPUSMV contains 200 stocks. One could argue that there are more possibilities to form interesting pairs. Graphic-6 compares the performance of the USMV (red), the SPLV (yellow) and the SPUSMV (green). The maximum number of pairs in each formation period is increased for the SPUSMV to 30. The weight per trade reduced to 8%. The performance of the SPUSMV is close to the USMV. There is no significant advantage. If one restricts the maximum number of pairs like for the USMV to 20, the performance of the SPUSMV is worse. I tried several combinations. The union never beats the best component.

Trading Costs:
Graphic-7 shows the performance of the USMV with different trading costs. The trading-cost model assumes a certain amount of cents per trade and share. This is a very simple minded model. But it gives a good feeling for the trading volume involved. The red line is the performance of the USMV without trading costs. The yellow line assumes 1 cent per share and trade, green 5 cents, blue 10 cents and dark blue 13 cents. This is the break-even point.
One can decrease the number of trades and improve the quality of the pairs by setting the minimum number of crossings to 5. The weight of each pair was increased to 14% to get a similar P&L than for the standard settings with 4 crossings. The Sharpe ratio of the case without trading costs (red line) is lower than for the setting above. The higher quality of the pairs does not compensate the lower number of trades. But this setting does not suffer from the short term Brexit crash. The break-even case of above with trading costs of 13 cents per share and trade has now a P&L of 22.8%, a Sharpe-ratio of 0.38 and a drawdown of 14.1%. The break even increases to 16 cents. Or in other words a full trading cycle makes about 64 cents per share (the number of shares are on the short and long side not the same. This is only a gross average figure).

Increasing the number of crossings to 6 does not work anymore. There are too less pairs left. Increasing the band-width does not work either. If one increases the entry point to 2.5 instead of 2.0 standard deviations the number of pairs and especially the number of trades drops dramatically.

---

Graphic-8: Performance of USMV 5 Crossings with different trading costs

One can also change the sorting order. Like before the main sorting is done according crossings. But one can sort the pairs with large deviations ahead of the ones with small...
ones. The red line in Graphic-9 is the standard setting. For the yellow line the large
distances are - within the same crossings group – sorted ahead. The final performance is
the same. But the red line is overall smoother. The Sharpe-ratios are 1.15 to 1.03, the
max. relative drawdown is 8.2% to 12.6%. The green line is the performance of the
inverse sorting method with 13 cents trading costs. It has a P&L of 17.2% and a Sharpe
ratio of 0.22. The break even is at 16 cents. Inverting the sorting order increases
somewhat the win per trade but introduces additional risk.
One could close the position not at the mean but when it has already moved in the other
direction. One possibility is to close if the spread has moved one standard deviation on
the other side. The performance of this strategy is close to the standard one. The break-
even is also around 13 cents. I have tried several other parameter settings. But there
seems to be a barrier around this value.

**Conclusion:**
The strategy is with the right choice of the universe and the improvements due to the
crossings selection rule interesting. The trading costs question can not be empirically
answered with the available daily data. One needs high frequency intraday data with bid-
ask spreads. Intraday trading should improve the performance somewhat because one
optimizes the entry and exit points.
The best way to test the feasibility of the strategy is to do some (paper-) trading with
initially low investment. This has to be done anyway when developing a new strategy.

**References:**
relative value arbitrage rule.
distance, cointegration and copula methods
market.
Thesis.