LUCENA
RESEARCH

Scientific Investment Technologies

Trading with Correlated Pairs

Created: December 10, 2013
Updated: July 3, 2014

Weiyi Chen, Quantitative Analyst
Tucker Balch, Ph.D., CTO

Lucena Research, LLC
75 5th Street, Suite 1025
Atlanta, GA 30308

404.907.1702 phone
404.751.0132 fax

Overview

Lucena Research has developed multiple market neutral strategies through the means of pairs trading. In this paper, we will review a strategy that trades pairs of correlated stocks using a reversion method outlined below.

Background and Related Work

In order to build a market neutral portfolio, we find pairs that are strongly correlated with one another will remain correlated for some finite period in the future. An easy to see example of a mean-reverting pair of stocks is given in Fig. 1. Using their correlation, we can perform a linear regression on their returns to determine when they have made a significant departure from their expected relationship (spread) and take mean reverting positions profiting on their return to the expected difference.

Figure 1: A typical pair found using Linear Regression, we see that entrances to positions occur when the pair of returns has diverged (prices have moved apart from one another) and exits occur when the returns have once again converged (underlying chart from Google Finance).

In this paper we will outline the process for one such pairs trading strategy which will have the following pipeline:

- Identify pairs of highly correlated equities
- Select pairs based off of their potential returns
- Simulate performance of these pairs with backtesting

A more thorough introduction to pairs trading and additional investigations into the underlying theory can be found in Ganapathy Vidyamurthy's book on the same topic (see References).

Methodology: Finding and Selecting Pairs

The first step in our pairs trading strategy is to identify pairs of highly correlated equities. This is done by calculating the correlation between all pairs in a given basket over a look-back period prior to the date of interest. For this paper the S&P 100 was used. Once the correlation matrix of all pairs has been calculated, we use a threshold to give us only pairs with a high degree of correlation (around 0.97). Table 1 contains 10 highly correlated pairs from our trial typical of those found on some day.

Index	Equity 1	Equity 2	Correlation
1	LMT	RTN	0.994625
2	LMT	SBUX	0.982215
3	GD	LMT	0.982196
4	MET	WFC	0.979455
5	MMM	TXN	0.97843
6	FOXA	TWX	0.978277
7	BA	SBUX	0.978092
8	GD	RTN	0.978003
9	TXN	UTX	0.977723
10	HON	MMM	0.9776

Table 1: list of top 10 highest correlated pairs on Dec 10th, 2013

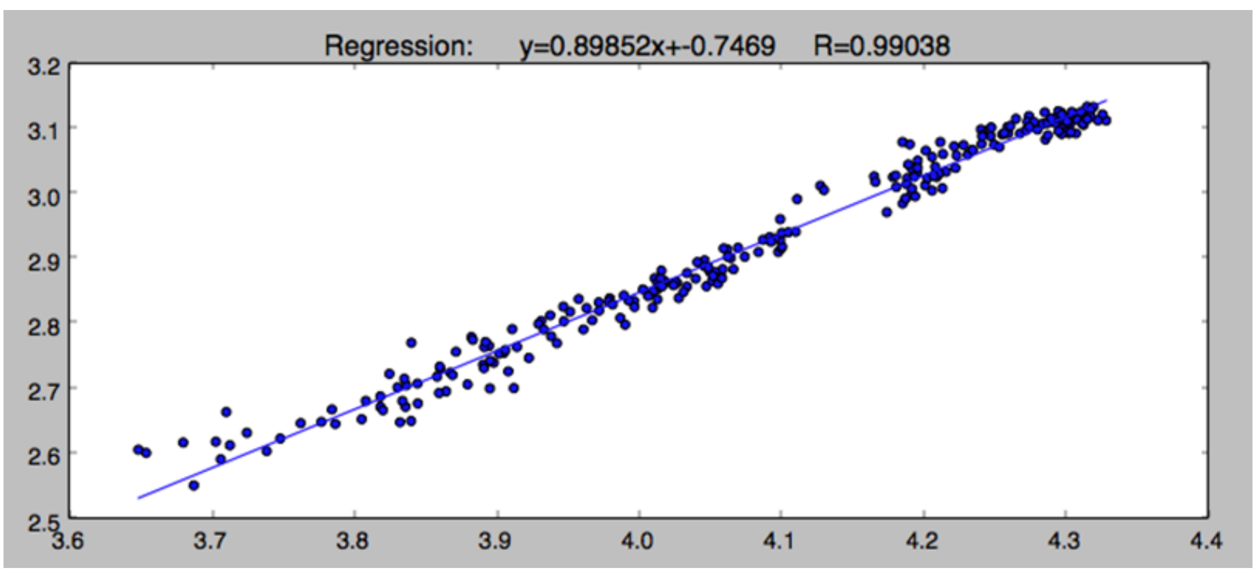


Figure 2: A typical set of returns for a pair (APD, CSCO here) of highly correlated equities with a best fit line plotted.

Once we have a set of highly correlated pairs, we then calculate their cumulative returns over the look back period. We then plot the returns for each pair and find the best fit line for these returns (see Fig. 2). This best fit line is the baseline which we will compare all future return pairs to. For all of the past returns in our preceding period, we determine the standard deviation of their distance from the best fit line which we will use as our signal to determine when to enter a position in this pair. The best fit line will be used to determine when to exit positions.

Equipped with the best fit line and standard deviation of the distance of return pairs from the best fit line (this distance is called the *residual* of the pair), we will enter positions when the residual is more than one standard deviation away from the best fit line, expecting mean reversion to lead this pair's returns back across the best fit line. When the residual crosses the best fit line, we then exit the position having made a profit. See Fig. 3 for an example of the residual for a pair.

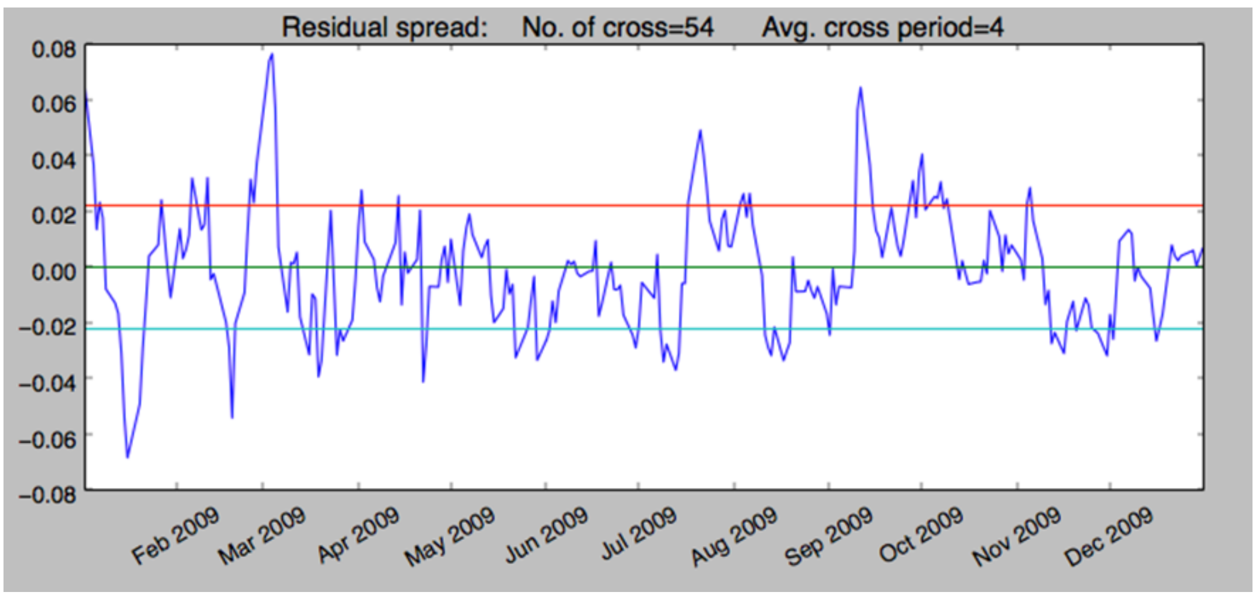


Figure 3: A typical residual spread (the distance each pair of returns is from the best fit line) for with standard deviation lines marked above and below the central line (which is representative of the best fit line).

To determine which pairs will be most profitable, note that profit is made based off of two factors: (1) distance of the residual from the best fit line when we enter a position, and (2) the frequency at which the residual crosses the best fit line. The product of these two factors yields the profitability of a given pair, thus we look for those pairs which are historically more profitable in this sense, and will use these pairs in our backtest trades. See Table 2 for a list of the top 10 most profitable pairs for a given period in our backtest.

Index	Equity 1	Equity 2	Correlation	Profitability
1	DD	GILD	0.97138	69.62

2	MDT	MET	0.971099	63.47
3	FOXA	GILD	0.971349	61.75
4	GD	RTN	0.978003	56.64
5	LMT	SBUX	0.982215	49.76
6	FOXA	TWX	0.978277	46.04
7	RTN	SBUX	0.972998	39.98
8	GD	LMT	0.982196	39.9
9	COP	MA	0.974329	36.99
10	TXN	UTX	0.977723	32.91

Table 2: list of top 10 most profitable pairs on Dec 10th, 2013

Methodology: Back Tests

We implemented a trading strategy in simulation as follows:

1) On each trading day, we generate the most profitable pairs for each two stocks in S&P 100 (as described above). We enter an equally-weighted position when the residual is greater than 1 standard deviation away from the best fit line (e.g., long in A, short in B if above +1 sigma; short in A, long in B if below -1 sigma).
2) When the residual crosses the best fit line or position is held a month, exit the position.

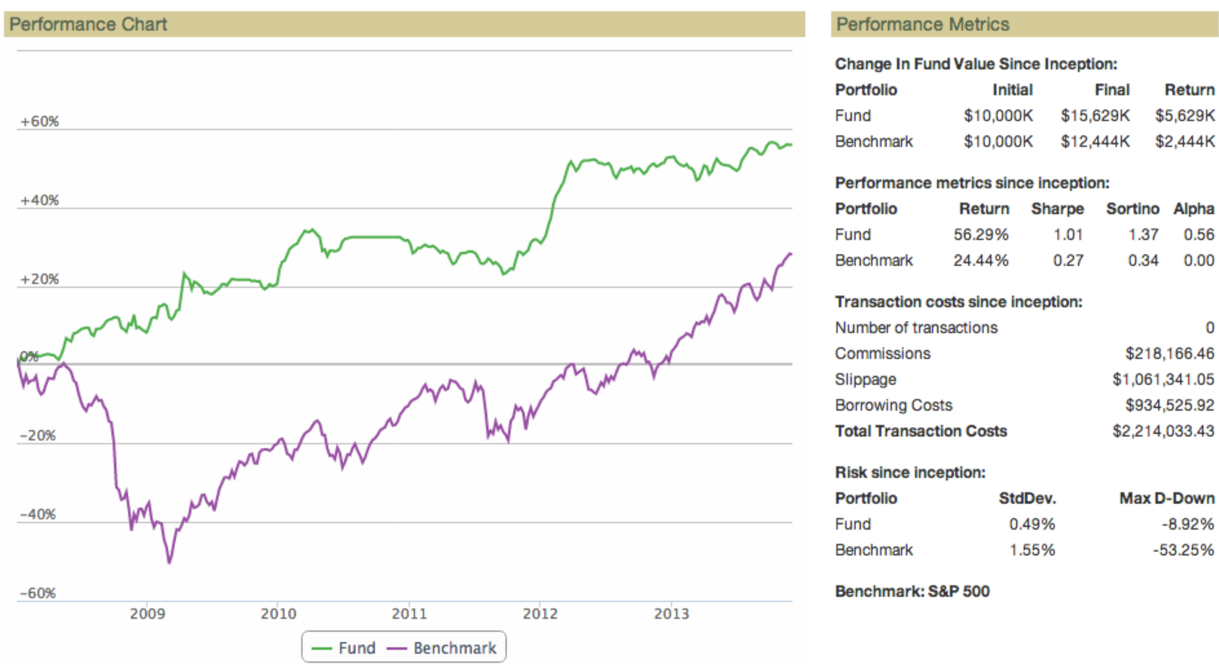


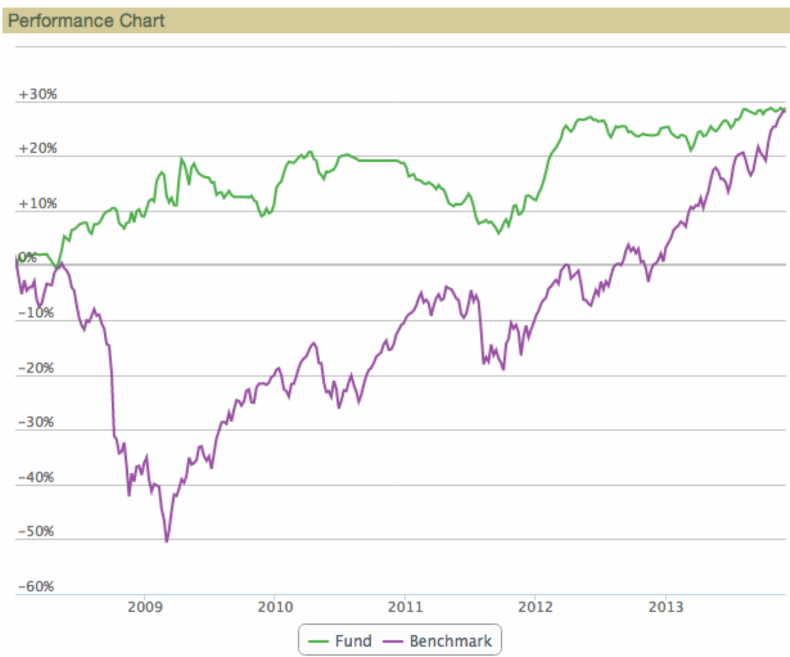
Figure 4: Performance of the ODR pairs trading strategy for 2008 to 2013 with most

profitable pairs.

3) We add transaction costs by modeling commissions, slippage penalties and short interests. Commission are set at $0.0035 per share with a $2.95 minimum per trade, slippage is estimated at 5bps for each transaction (including that the price will move against the trader by 0.05% when entering and exiting a position), and short interest rate is estimated at 3.5% annually.

The first backtest as shown in Fig. 4 shows very positive results, with a consistent average sharpe ratio of more than 1.0 each year.

In our second back test we remove the process to assess which pairs will be most profitable, and solely use the most correlated pairs. As you can see in Fig. 5, performance is still positive but reduced in comparison with the back test considering profitability process.



Performance Metrics

Change In Fund Value Since Inception:

Portfolio	Initial	Final	Return
Fund	$10,000K	$12,897K	$2,897K
Benchmark	$10,000K	$12,444K	$2,444K

Performance metrics since inception:

Portfolio	Return	Sharpe	Sortino	Alpha
Fund	28.98%	0.59	0.78	0.29
Benchmark	24.44%	0.27	0.34	0.00

Transaction costs since inception:

Number of transactions	0
Commissions	$201,736.97
Slippage	$946,605.37
Borrowing Costs	$845,328.80
Total Transaction Costs	$1,993,671.14

Risk since inception:

Portfolio	StdDev.	Max D-Down
Fund	0.49%	-13.00%
Benchmark	1.55%	-53.25%

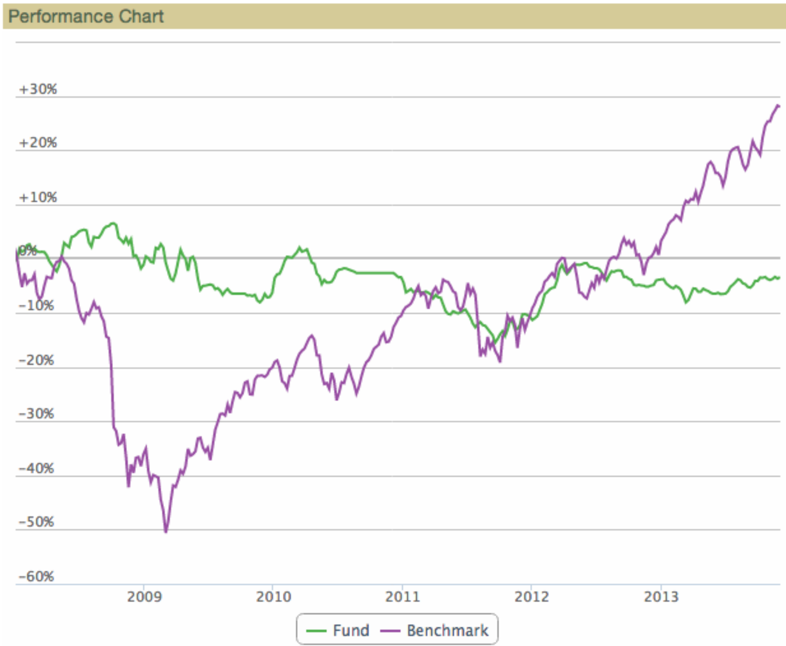
Benchmark: S&P 500

Figure 5: Performance of the ODR pairs trading strategy with most correlated pairs.

Finally, as a control, we ran a back test in which the pairs were selected randomly from the selected correlated pairs each day. The purpose of a randomly selected portfolio is to illustrate the performance we would expect if there were no predictive information. Performance of that test is illustrated in Fig. 6 in the next page.

As you can see, performance for the random correlated pairs is significantly worse than for our test cases.

More detailed back test reports, including specific transactions are available from Lucena Research on request.



Performance Metrics

Change In Fund Value Since Inception:

Portfolio	Initial	Final	Return
Fund	$10,000K	$9,673K	($326,185.27)
Benchmark	$10,000K	$12,444K	$2,444K

Performance metrics since inception:

Portfolio	Return	Sharpe	Sortino	Alpha
Fund	-3.26%	-0.03	-0.04	-0.03
Benchmark	24.44%	0.27	0.34	0.00

Transaction costs since inception:

Number of transactions	0
Commissions	$164,410.10
Slippage	$772,381.61
Borrowing Costs	$724,993.64
Total Transaction Costs	$1,661,785.35

Risk since inception:

Portfolio	StdDev.	Max D-Down
Fund	0.50%	-20.88%
Benchmark	1.55%	-53.25%

Benchmark: S&P 500

Figure 6: Performance of the ODR pairs trading strategy in which correlated pairs were selected randomly.

Discussion

We propose a pairs selecting and corresponded event triggering criteria based on relative pricing and the idea that mis-pricing will correct itself in the future. The event predicts higher returns for equities whose recent performance have had low returns than for equities whose recent performance have over-performanced.

The criteria was utilized in a pairs trading strategy in which we enter equally-weighted long and short positions in event-triggered pairs. The strategy was evaluated in three back tests:

1) With most profitable pairs;
2) With most correlated pairs;
3) With randomly correlated pairs.

As we remove requirements for profitability assessing (Fig. 5) and correlation ranking (Fig. 6), the backtests significantly underperform the former test that includes these factors (Fig. 4). This suggests that there is actionable information in our pairs trading strategy.

References

Pairs Trading: Quantitative Methods and Analysis
Ganapathy Vidyamurthy
Wiley Finance, 2004