Statistical Arbitrage Project: Pairs Trading - LATAM vs Commodities

Name: Claudio Gonzalez
NYU-ID: Cag693
Professor: Daniel Totouom

Oct 16, 2018

Introduction

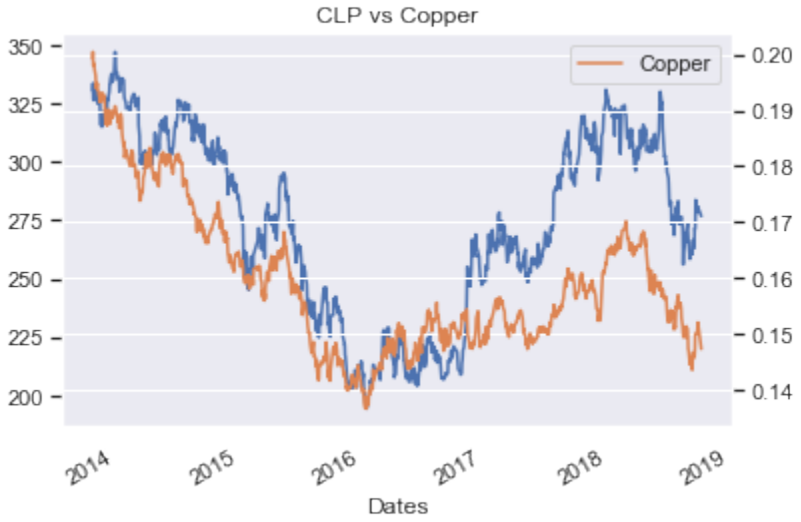
For this Statistical Arbitrage Project, the chosen topic was pairs trading with a mean reversion strategy. Pairs trading is defined as a strategy that monitors the performance of two historically correlated securities. When the correlation weakens a spread is created, meaning the pairs trade would short the outperforming stock and long the underperforming one. The idea is to define this spread or residual for two stocks (a and b) and beta as the hedge ratio as:

$$Residual = Y_a - \beta Y_b$$

The idea is to find two securities or financial assets that behave the same, so for the purposes of this project the chosen area was LATAM currencies and Commodities. Historically economies in Latin America depend immensely in commodities being a big part of their GDP as shown in this example in the following table for Chile:

Sector	GDP %
Mining	15.2
Business services	13
Manufacturing industry	10.9
Personal services	10.6
Retail	7.9
Construction	7.4
Real estate services	5
Public administration	4.3
Financial services	4.2
Transportation	4.1
Agriculture and forestry	2.8
Electricity, gas and water	2.4
Communications	1.9
Restaurants and hotels	1.6
Fishing	0.4

Here is another example of how the currency follows the commodity price with CLP (blue) and Copper (orange) futures. As seen in the graph they are heavily correlated.



Implementation - Filtering and Pair Selection

Then, the futures generic first year data between Oct. 2013 - Oct. 2018 was obtained for the top six LATAM currencies which were Brazil, Mexico, Chile, Colombia, Peru and Argentina. These futures were then changed to the value of LATAM/USD, meaning for example how many dollars would you get for 1 Chilean Peso, showing the real valuation of the currency against the dollar. The same future data was obtained for the top exported commodities for each country which turned out to be copper, aluminum, meat, soy beans, crude oil, iron ore, corn, sugar, coffee, wheat, gold and silver. Then these currencies were paired with each commodity to create a ranking systems that consisted in certain filters to generate possible profitable pairs with their correlations, co-integrations and percent exports.

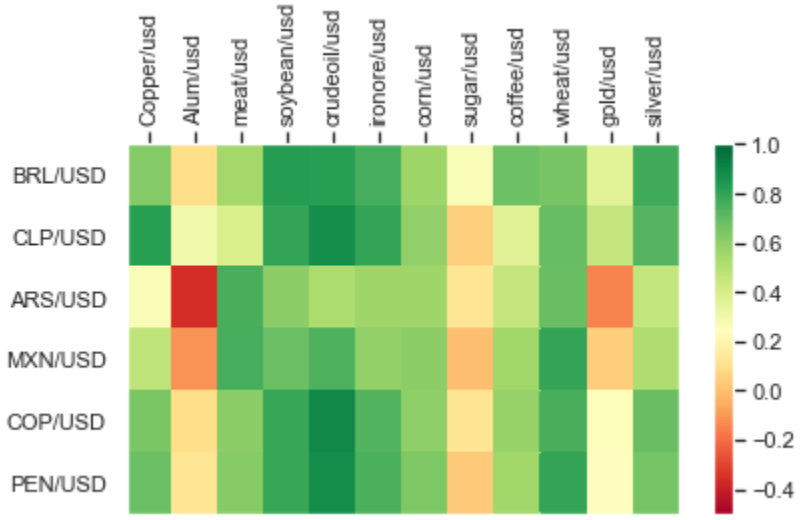
First, the correlation between the currency and the commodity was calculated .This was performed to measure the relationship between the two stocks and if their price follows the same trends. Using the following equation for asset A and B, the correlation is defined as:

$$\rho = \frac{\sum(A_i - \overline{A})(B_i - \overline{B})}{[\sum(A_i - \overline{A})^2 \sum(B_i - \overline{B})^2]^{\frac{1}{2}}}$$

$$\overline{A} = \frac{1}{N}\sum A_i$$

$$\overline{B} = \frac{1}{N}\sum B_i$$

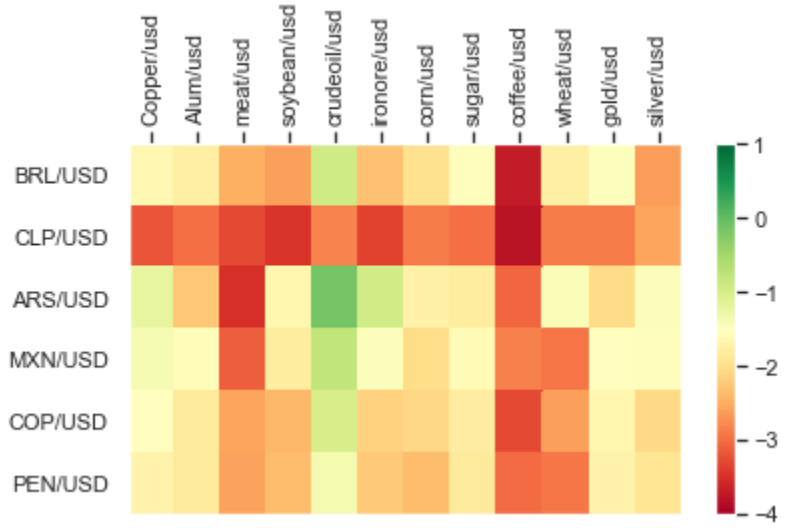
The correlation was calculated for every currency/commodity pair and the following heat map shows the distribution. For example CLP/Copper is highly correlated but ARS/Aluminum isn't.



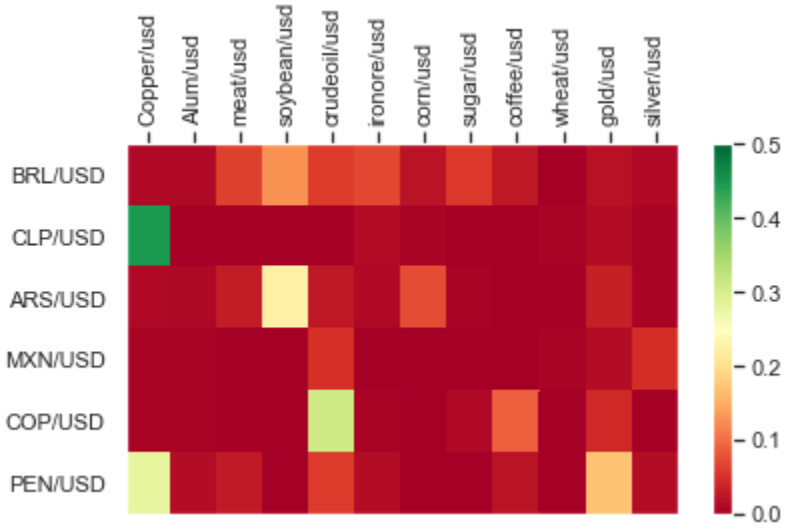
Then the co-integration between currency/commodity pair was calculated. In this case the co-integration identify pairs that tend to have mean-reverting properties relative to its prices. Co-integration states that given two non-stationary time series, there can be a linear combination of the two time series that is in fact stationary. This just shows that the two assets move together in the same way. This is defined as the price of asset A and B, there exists a parameter γ (which is defined as then co-integration coefficient) such that the following equation follows a stationary process:

$$Y_A - \gamma Y_B = \mu + \epsilon_t$$

In this case, a 95% confidence interval was used which is equal to -3.0199. This is the threshold used for the rankings filter, where if the co-integration coefficient was equal or less than that number it would be considered a profitable pair. Once again the co-integration was calculated for every pair of currency/commodity and the following heat map (which in this case the more red the relationship is, the better the co-integration value) shows the results between the pairs.



The final filter was the percent exports that country had for that specific commodities. The hypothesis here is that if the country has a higher percent of exports of that commodity, it would increase the probability of it being an actual profitable pair. The greener the relationship, the better the pair is for the strategy, being CLP/Copper the pair that has the highest percent of exports.



Once these three filters were created, the ranking portion of the project began. The idea was to rank these pairs between these three variables (correlations, co-integrations and exports) meaning that basically if they had values above a certain threshold, the pair would be a possible profitable pair for the strategy. Once the pairs were all ranked, nine pairs were chosen profitable. Out of those nine pairs, seven of them

had exports while two of them didn't it (highlighted in red in the table below). This was done with the intention of proving the hypothesis that percent exports have an actual impact on the possible profits of the pair strategy. The following table shows the pairs chosen to perform the pairs trading strategy:

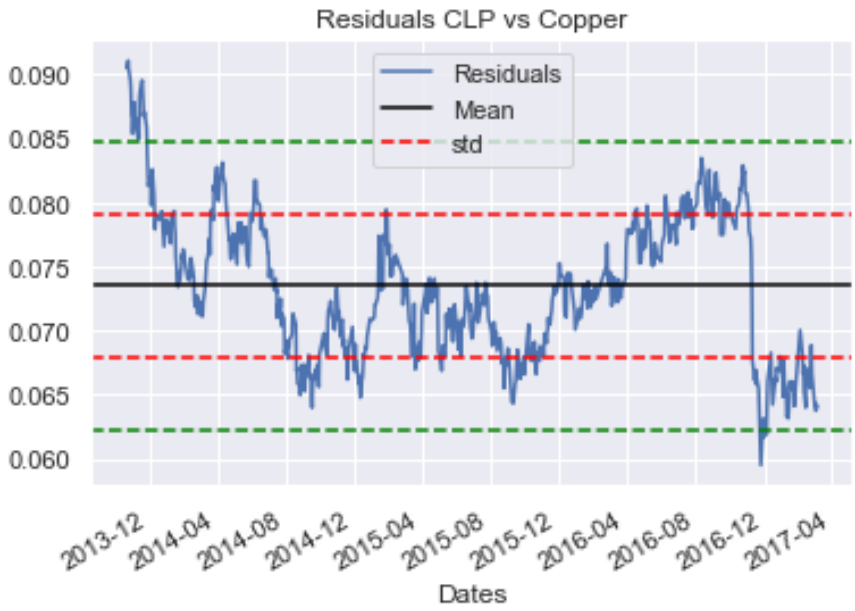
Pair	Correlation	Cointegration	% Exports
CLP/Copper	0.82	-3.18	44.80
COP/Coffee	0.59	-3.29	8.80
ARS/Meat	0.76	-3.50	3.00
BRL/Coffee	0.68	-3.70	2.70
PEN/Coffee	0.56	-3.02	2.10
CLP/IronOre	0.80	-3.35	1.30
CLP/Meat	0.39	-3.29	0.31
CLP/Soybean	0.80	-3.47	0.00
MXN/Meat	0.76	-3.11	0.00

Implementation - Training Strategy and Pairs Trading

After the pairs were selected, the development of the strategy began with a training set of 70% of the data. The first approach was to find the betas in the linear relationship or the "hedge ratios". In order to find a more accurate ratio, the method used was "Total Least Squares" which is defined with a minimization of the objetive function of S, r being the vector of residuals and W the weighting matrix.

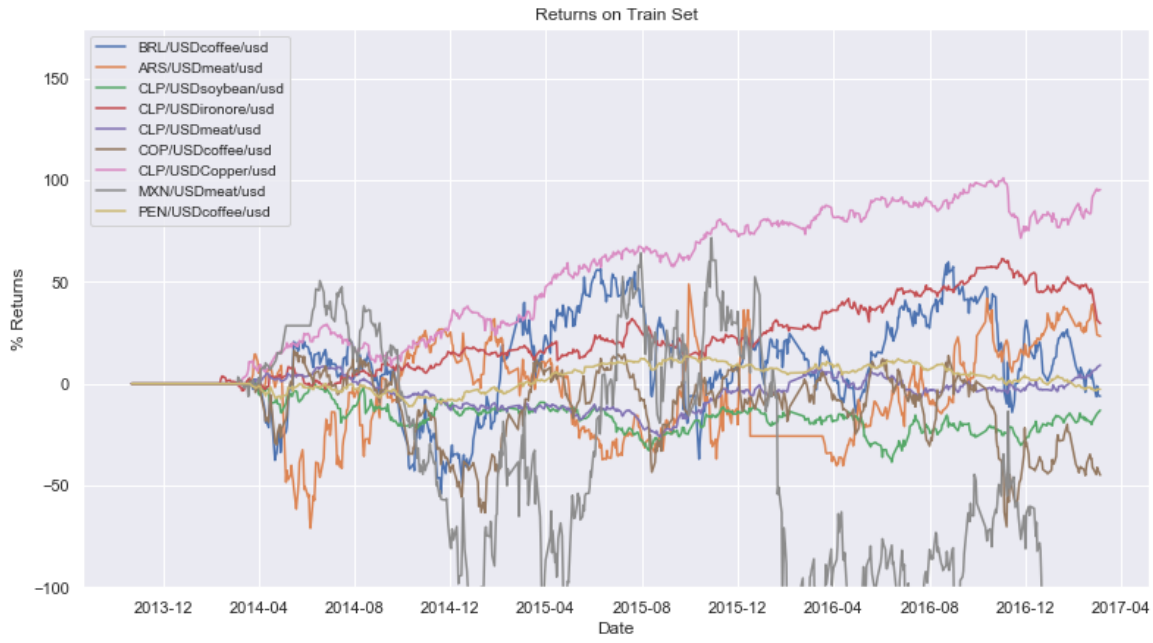
$$S = r^T W r$$
$$r = y - X\beta$$

After the betas were calculated, the residuals were calculated for each pair in order to be used in the pairs trading strategy. The graph below shows the residuals for the relationship between CLP and Copper. As seen in the graph the blue line represent the residuals, the black line the mean, the red line one standard deviation from the mean and the green line two standard deviations from the mean.



The idea here is that if the residual would be between the mean and minus one standard deviation from the mean (the lower bound), this would execute a long signal. If the residual would be between the mean and plus one standard deviation from the mean (the upper bound), this would execute a short signal. In this case as seen from the graph, one standard deviation was chosen instead of two because it would embrace a more consistent and executable strategy. The rolling window used for the rules was 65 days due to it being almost one quarter of the year, which was decided in the strategy that it would be an acceptable standard due to the high volatility of LATAM currencies. Once the parameters were set due to a risk neutral strategy once a currency was long, the commodity was short and vice versa. The output of the strategy would give percent returns, Sharpe Ratio (annualized with square root of 252), Maximum Drawdown, Number of Trades and Hit-Ratio. The following table and graphs shows the performance statistics and returns for each of the nine pairs selected.

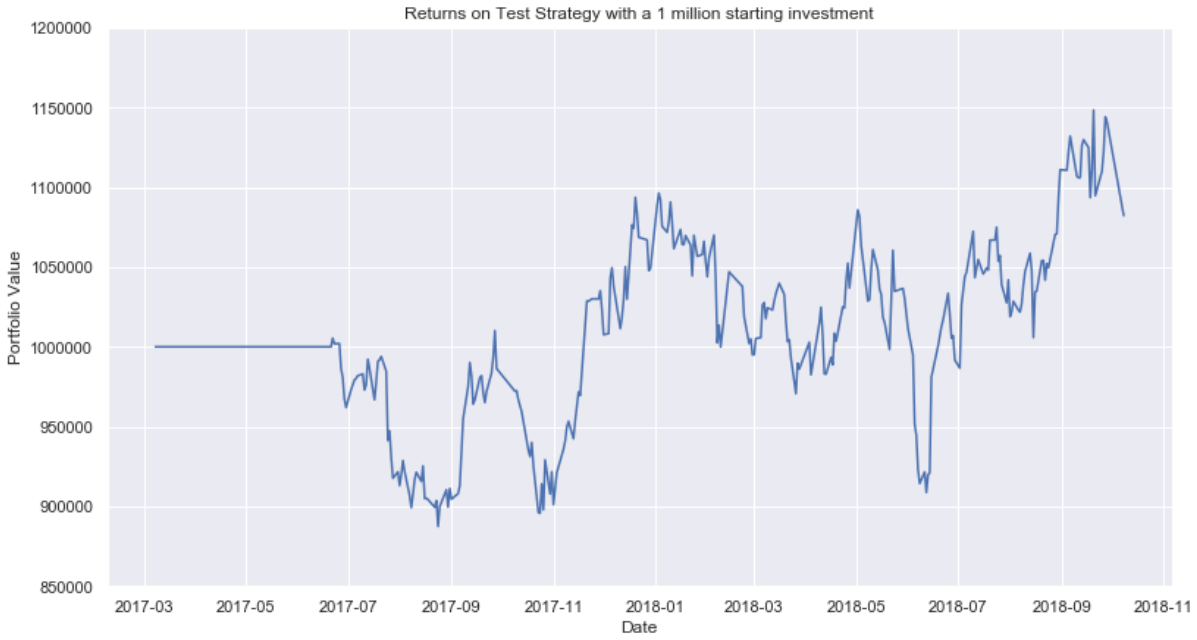
Pairs	Total Returns %	Sharpe Ratio	Maximum Drawdown %	Number of Trades	Hit Ratio %
CLP/Copper	95.4	1.3	15.8	82.0	53.7
CLP/IronOre	29.6	0.5	19.7	63.0	50.3
ARS/Meat	23.4	0.1	74.7	26.0	52.7
CLP/Meat	9.1	0.2	32.0	40.0	50.1
PEN/Coffee	-2.7	-0.1	14.9	43.0	49.8
BRL/Coffee	-6.1	0.0	62.0	39.0	51.9
CLP/Soybeans	-13.0	-0.2	39.1	32.0	51.2
COP/Coffee	-44.9	-0.2	74.4	37.0	49.6
MXN/Meat	-549.4	-0.3	838.2	37.0	49.1



As seen from the table and graphs the pairs varied a lot in terms of performance. The pairs highlighted red were the pairs that had no exports. As you can see both of these pair ranked at the bottom, being MXN/Meat a pair with very low performance statistics in almost every category. This helps the hypothesis that the pairs chosen should have an substantial percent of exports in the country. The best performing pair was with no surprise CLP/Copper having high performance statistics in comparison to the rest. As seen above this pair has great correlation and co-integration but more important almost 50% of Chile's exports are copper which had a clear impact in the strategy. After reviewing the results, the next steps was to choose certain pairs to perform the strategy in the remaining 30 percent of the test data set. In this case, the pairs chosen were the ones with the more positive performance statistics, which were the ones highlighted in yellow, CLP/Copper, CLP/IronOre and CLP/Meat.

Implementation - Test Strategy and Final Results

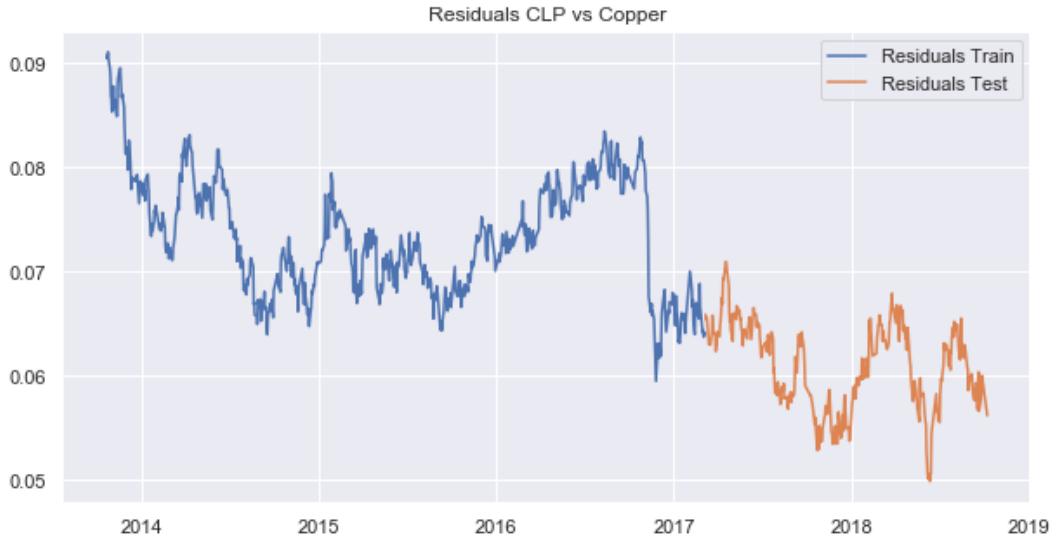
The test strategy with the remaining of the data set (Mar. 2017 - Oct. 2018) was a portfolio constructed of the pairs mentioned above. In this case 1 million dollars was allocated between these three pairs, having $\$800,000$ for CLP/Copper, $\$100,000$ for CLP/IronOre and $\$100,000$ for CLP/Meat. The idea is to use the same beta for the training set in order to calculate residuals and the performance statistics through that period are the following



Pairs	Total Returns %	Sharpe Ratio	Maximum Drawdown %	Number of Trades	Hit Ratio %	Investment Returns
CLP/Copper	11.79	0.29	21.14	20.00	53.93	$94,306.78
CLP/Ironore	7.86	0.30	15.95	17.00	48.83	$7,862.76
CLP/Meat	-20.01	-1.07	21.97	7.00	46.48	$(20,012.21)
Total Portfolio	8.22	0.15	20.71	18.40	52.67	$82,157.33

As seen from the results the strategy allocated a total return of 8.22 percent with a total gain of the portfolio of $\$82,157$. The two first pairs had positive returns but low sharpe ratios with a significant decrease in number of trades, where as the last pair had negative performance statistics. Once again the best performing pair in the portfolio was CLP/Copper, helping the hypothesis that high exports have a high influence on profits of the pair.

Finally, the following graph shows how the residuals change between the training and test set for CLP/-Copper from the example from before. As seen, the test sets values of the residual decrease compared to the training set, which is seen in the actual number of trades performed in the strategy.



Conclusions

The only pair from the chosen pairs that had solid returns and performance statistics was CLP/Copper, which was the pair that had 45% of exports. The pairs that had no exports were the pairs that performed the worse from the nine chosen pairs.The pairs that had exports between 1 and 10% performed below average and had below average performance statistics. The strategy performed with lower returns after 2017, this could be due to the emerging markets turmoil that has been going on through 2018. As mention before there is a clear decrease in the values of the residuals from 2017 till Oct. 2018 which could have had a clear influence in the final results of the portfolio. Overall it is clear that the only profitable pair with solid performance statistics is CLP/Copper and the fact that the percent exports of copper in Chile are very high, it could have a clear influence in the profits of the strategy, which is a very good indication that the hypothesis stated at the beginning was possibly correct.

In order to improve the strategy there are many factors that could be incorporated. First, select different windows for moving mean and std (use solver to find the best window for every pair) or a rolling shifting beta. Also, changing the rankings approach to incorporate a higher weight on exports (use pairs that have exports higher than 20 percent). This means that pairs that have high exports of that commodity should definitely be incorporated as profitable pairs. This could also include a decrease of the confidence in the

co-integration test value to 90 percent. The strategy could also incorporate intra-day prices instead of end of the day and using log prices instead of actual prices but for the case of this study, actual prices were used. Expanding to this, use different generic futures for currencies (not only 1 year futures). Finally, a new approach could be use in the data set where a different training strategy such as slicing per year the inputs and where the rolling window moves depending on the year could be incorporated. In this scenario also the implementation of a Kalman filter to have a shifting window and a moving beta could be an essential implementation to obtain more precise performance statistics and strategy.

This strategy for currencies and commodities seem to be a profitable one, specially in the emerging markets/latin america spectrum. Optimizing this strategy could lead to a very profitable margin, if the pairs chosen are the most effective. The idea is to keep working on this strategy in order to find a great overall final strategy that incorporates changes in an effective matter for the future.

References

A mix of the following articles and websites was used to develop the strategy from the filtering, till the implementation.

1) "Better Hedge Ratios for Spread Trading" by Paul Teetor
2) "Pairs Trading, Convergence Trading, Cointegration" by Daniel Hermont
3) "High Frequency and Dynamic Pairs Trading Based on Statistical Arbitrage Using a Two-Stage Correlation and Cointegration Approach" by George J.Miao