Comparative Performance Analysis of Pairs Trading: A Statistical Arbitrage Trading Method

Statistics 107
Prof. Michael Parzen
Harvard University

Spring 2013

Jimmy Field
Graham Topol
Eric Westphal

Abstract

Pairs trading is a popular, market neutral strategy to profit from deviations in price relationships between two securities that tend to move together. This paper sets out to assess how sensitive the performance of such a strategy is to certain decisions that an investor must make when implementing it. First, we analyze the relative performance of two strategies to select pairs, correlation and co-integration, finding that co-integration is more effective given our assumptions. Then, we analyze three methods to define the stock price combination for the selected pairs, and find that the ratio of log prices and the normalized price ratio methods yield the highest profits. Finally, we test different standard deviation cutoff rules to trigger trades, finding that the lower the threshold, the more trades are executed, and the riskier the strategy becomes. We find that a cutoff level between 1.5 and 2.0 standard deviations tends to be most profitable, suggesting this to be an optimal point in the tradeoff between the number of trades and the percentage of winning trades.

I. Introduction

Pairs trading is a statistical arbitrage method designed to exploit transitory fluctuations from the long-run equilibrium between two stocks. This paper sets out to assess how sensitive the performance of such a strategy is to the many decisions that an investor must make when implementing a pairs trading routine. Chan (2010) tells us that co-integration was developed in order to model dynamic co-dependencies in multivariate time series. Chan (2010) tests whether co-integration is a superior method of picking pairs to correlation, and concludes that is the measure that yields the better performance. This study will conduct a similar analysis on the NASDAQ 100 stocks, assessing the relative performance of correlation and co-integration as a measure of the relationship between two stocks. Using standardized time series data we also find some evidence for highly co-integrated pairs performing better than pairs that are only highly correlated, as the technique identifies a certain "hedge ratio" that makes the spread of two non-stationary stock price series constant, following a common stochastic trend.

We test the two sets of stock picks in a trading simulation routine that is discussed in detail in the Methods section. The trader could set a trading rule in which she selects a multiple of the historical standard deviation of the normalized spread, and when the spread widens above or narrows below these multiples a position will be opened (Chan 2010). Caldeira and Moura (2012) describe pairs trading as an attractive market neutral (cashless) strategy, as one in which given a widened spread, the investor goes long and short the appropriate assets in the same magnitude, costing a net of $0 in the opening period, and giving a chance for a profit at a future time. It is important to note that the term "arbitrage" in statistical arbitrage is used loosely, as there is no net cash outlay in the present but there is the chance for a loss in the future, making the strategy not riskless. Nevertheless, in general, profits of both highly correlated and co-integrated pairs are indeed positive with net zero cash outlays in the first period.

Additionally, we pick a pair that is highly co-integrated and correlated to analyze the relative performance of two other decision criteria an investor implementing a pairs trading routine must evaluate: 1) how to define the relationship between the two stocks to be traded and 2) which standard deviation cut-offs shall be used to trigger trades. Our analysis will analyze which standard deviation cutoff tends to yield the highest profit given the tradeoff between number of trades and winners as a percent of trades.

II. Methodology

Trading the relative movements of two securities that have historically moved in a close relationship is a popular strategy. This strategy employs mathematical models to generate profits in financial markets, as opposed to fundamentally driven approaches. Nevertheless, despite its statistical approach, some key assumptions must be made for pairs trading algorithms, which can have significant impacts on the performance of the trading routine. In this paper, we examine three parts of the decision process an investor faces in implementing a pairs trading strategy:

which pairs to trade, how to define the combination of their prices, and how to select a standard deviation band cutoff and trading method to signal trades.

A. Correlation vs. Co-integration

The first key decision this paper explores is perhaps the most crucial: selecting which pairs to trade. There are two dominant approaches for selecting pairs: the use of within-pair correlation and within-pair co-integration. In either method, the idea is to find pairs of stocks that, given a combination of its prices, will either move together (correlation) or display a mean reversion (co-integration).

This mean reverting process for a security X can be expressed as

$$dX_t = \alpha(\bar{X} - X_t)dt + \sigma(X_t)\, dW_t \text{ -}$$

where $\sigma(X_t)$ is a general volatility function.[1] As long as the modeled relationship holds, an investor at time t should short asset X if $X_t > \bar{X}$, signaling a downward trend in the price of X ($dX_t < 0$), and buy asset X if $X_t < \bar{X}$, signaling an upward trend in the price of X ($dX_t > 0$).[2] The same idea holds for a combination of stocks that historically exhibits such mean reverting patterns.

In picking stocks for a pairs trading strategy, investors can use correlation as a proxy for identifying stock pairs that may fit the mean reverting pattern. When stocks are historically highly correlated, one may expect that deviations from the pattern are temporary and the combined price series will revert to its true mean, giving the opportunity for a profitable trade. However, this method is not seen as very robust, since it is possible that two stocks following a

[1] Ngai Hang Chan (2010), Chapter 15.
[2] Ibid.

random walk will be correlated due to a shared component, but whose price difference may also be a random walk, making the combined series non mean reverting.[3]

An alternative method involves applying the notion of co-integration to the time series, which may allow for the modeling of "dynamic co-dependencies."[4] With co-integration we find a linear combination of two stocks that is stationary (mean reverting) in the long run, even if the individual price series are non-stationary.[5] In practice, one can look for a pair of stocks such that initiating opposite positions in each of them in a certain ratio (called the hedge ratio), will return a market value of the pair that will be stationary. To do this, one can regress the first stock on the second stock and extract the Beta coefficient as the hedge ratio (the number of shares one must short for every one share of the first stock). Secondly, one can conduct a Dickey-Fuller stationarity test to determine whether the spread of the two stocks proves to be stationary in the long run.[6]

In the following section, we display the results of a routine implemented in R to undertake a comparative study of correlation and co-integration as methods for choosing pairs.[7] The routine takes a list of stocks and ranks their combinations based on highest correlation coefficient or lowest p-value in a Dickey-Fuller test of series stationarity.[8] In "Time Series: Applications to Finance with R and S-Plus", Chan conducts this exercise using the 42 Hang Seng Index Component stocks. For our analysis, we use the NASDAQ 100 Index components. Like Chan, we rank the top 10 pairs using correlation and the top 10 using co-integration as the criteria. The same routine can be applied to other stock indexes, such as the S&P 500, which due

---

[3] Michael Parzen, Harvard University, Statistics 107, Lecture 16, Spring 2013.
[4] Ngai Hang Chan (2010), Chapter 15.
[5] Ibid.
[6] Parzen, Michael. Ibid.
[7] Ryan, Jeffrey A., {quantmod} package.
[8] H_o = non-stationary.

to their size improves the chances of finding highly correlated and co-integrated pairs, but this requires significant machine computational power, as the terms of the correlation and co-integration matrices grow exponentially.

Having selected the top 10 stocks by correlation and co-integration, we test how each selection method performs in a simulation that will be discussed in the Results & Analysis section. To make our analysis realistic, we do not use the co-integration and correlation coefficients from the same period in which we simulate trading, since an investor would have to select pairs based on historical price movements that may or may not hold in the future. For this reason, we conduct our "measurement period" to rank pairs from January 1, 2011 to January 1, 2012. The performance of the pairs is assessed by simulating the pairs trading strategy in the following year, from January 1, 2012 to January 1, 2013.

B. Trade Signals

Having selected pairs, the investor must decide how to define and identify the trading signals to trigger investment positions in his routine. We implemented a function in R that takes a pair of stocks, a trading method, a trading time window, an amount to be neutrally invested in each trade, a standard deviation band threshold to initiate trades, and other parameters that are detailed in Appendix E that determine program and output functionality. This routine then loads daily-adjusted stock price data for each stock. This data is used to compare pairs selected with the correlation and co-integration methods, as well as to undertake analysis of standard deviation cutoffs and price combination definitions in order to determine which perform best in our sample.

First, we compare the performance of three popular strategies of defining the price relationship to be used in the routine. These include: the (1) "normalized price difference" approach, used by Marcelo Perlin in “Evaluation of Pairs Trading Strategy at the Brazilian Financial Market”, (2) the “price ratio”, used by Pairslog.com, and (3) the “log price ratio”, used by Chan in “Time Series: Applications to Finance with R and S-Plus.” The three strategies are defined formally in Appendix A.

Secondly, given a pair combination strategy, we compare how setting different standard deviation cutoff triggers in the normalized series affects performance. Particularly, we conduct these sensitivity analyses using ±1, ±1.5, ±2, ±2.5 and ±3 standard deviations from the mean as triggers. For example, if a series crosses the upper band of a given standard deviation cutoff, the algorithm initiates a short position in stock 1 and a long position in the same amount in stock 2.

Each position in our comparative analyses is closed when the series crosses zero, or when it reverts entirely to the mean. Another choice an investor must make is when to close the position, which can significantly impact performance, but we limit our analyses to this “close trade” signal definition. Additionally, the algorithm records an open trade for trades that may have still been open at the end of the simulation period, but does not automatically close the trade and ignores that trade in all other performance metrics. Analyzing the average length of trade and number of trades for each simulated pairs allows one to gauge whether only a couple of problematic positions were entered in the case of a pair that is very uncorrelated or non-co-integrated, but this does not prove to be a problem in our analysis. Finally, our analysis ignores all trading and transaction costs.

Performance is measured in the following categories: average length of an open trade position, number of trades, percentage of trades that were profitable, maximum profit on a trade, minimum profit on a trade, average profit per trade, total profit. By default, the simulation routine assumes that the investor goes long the appropriate stock $10,000 and shorts the other by the same amount, making the strategy market neutral.

Using the previously stated performance measures and described routine, we run the top 10 pairs from each of the correlated and co-integrated groups. We also select the top pair in each selection category and conduct a deeper comparative analysis of the different trade signal strategies the pairs' price movements.

III. Results & Analysis

A. Correlation vs. Co-integration

Using the NASDAQ 100 from the dates January 1, 2011, to January 1, 2012, we ranked all pairs by correlation and co-integration and then selected the 10 pairs with the highest correlation and highest co-integration. Because there were more than 10 pairs with a co-integration coefficient of 0.01, we randomly selected 10 of the pairs with a p-value of 0.01 for the Dickey-Fuller stationarity test. The results are displayed in Table 1 with the correlation ranked in descending order on the left and the co-integration on the right.

Table 1: Correlation and Co-integration Pairs

Top 10 Pairs by Correlation		
Pair	Correlation	Co-integration
ADSK/ADBE	0.9635	0.1608
WCRX/LIFE	0.9635	0.0286
QGEN/LIFE	0.9573	0.2611
ILMN/BMC	0.9544	0.3034
PCAR/ LIFE	0.9536	0.0578
LRCX/AMAT	0.9493	0.0323
MU/AMAT	0.9455	0.2083
WCRX/QGEN	0.9452	0.2968
NIHD/ILMN	0.9450	0.2401
PCAR/EXPD	0.9421	0.2429

Top 10 Pairs by Co-integration		
Pair	Co-integration	Correlation
DLTR/ATVI	0.01	0.7293
EBAY/ATVI	0.01	0.1434
MAT/QCOM	0.01	0.3159
CELG/CHRW	0.01	-0.1185
MAT/GILD	0.01	0.3529
DELL/CHKP	0.01	0.7358
NVDA/CSCO	0.01	0.6373
INFY/FISV	0.01	0.8040
CA/ISRG	0.01	-0.5465
SYMC/CHKP	0.01	0.1720

Note: Co-integration value is p-value of Dickey-Fuller stationarity test

As the results show, the pairs with the highest correlation do not necessarily have the highest co-integration.

We then test the performance of the 10 pairs using both the correlation method and the co-integration method. The results are shown in Table 2. Note that the year being used to test the performance of the methods is 2012-2013, whereas the pairs were selected using the year 2011-2012, and thus the co-integration and correlation is different from those listed in Table 1. This is previously discussed in the Methods section, but is also highlighted in Appendix B.

Table 2: Correlation and Co-Integration Methods on 10 Pairs for 2012 Trading Period using ratio of log prices

Performance of Top 10 Correlation Pairs, Measured in 2011-12 and Traded in 2012-13

Pair	Correlation	Co-integration	Winners	Number of Trades	Win %	Max Profit	Min Profit	Profit	Average Profit	Average Trade Length	Hedge Ratio
ADSK/ADBE	0.3510	0.0400	11	13	85	$ 711.91	$ (656.05)	$ 2,953.30	$ 227.18	8.92	1.063
WCRX/LIFE	(0.5579)	0.7200	4	8	50	1110.76	(3341.89)	(2175.00)	(271.88)	12.63	0.273
QGEN/LIFE	0.4013	0.5630	6	7	86	543.49	(1242.60)	947.57	135.37	10.57	0.361
ILMN/BMC	0.1413	0.6950	9	14	64	1573.21	(1019.75)	3154.03	225.29	8.00	1.146
PCAR/ LIFE	0.7049	0.2360	9	12	75	460.79	(216.64)	2464.48	205.37	6.00	0.888
LRCX/AMAT	0.6116	0.7420	5	9	56	318.77	(1427.33)	(1872.37)	(208.04)	13.89	3.344
MU/AMAT	0.8523	0.2490	9	10	90	859.47	0.00	3928.90	392.89	8.20	0.592
WCRX/QGEN	(0.0378)	0.3600	8	11	73	716.39	(846.88)	1634.07	148.55	7.27	0.756
NIHD/ILMN	0.0179	0.2220	6	11	55	2123.95	(2941.11)	690.87	62.81	12.09	0.253
PCAR/EXPD	0.5882	0.1380	8	10	80	699.25	(1048.75)	972.50	97.25	8.30	1.050
AVERAGE			**7.5**	**10.5**	**71.4**	**$ 911.80**	**$(1,274.10)**	**$ 1,269.84**	**$ 101.48**	**9.59**	**0.973**

Performance of Top 10 Correlation Pairs, Measured in 2011-12 and Traded in 2012-13

Pair	Correlation	Co-integration	Winners	Number of Trades	Win %	Max Profit	Min Profit	Profit	Average Profit	Average Trade Length	Hedge Ratio
DLTR/ATVI	0.4640	0.5770	8	12	67	$ 691.91	$ (711.52)	$ 1,027.71	$ 85.64	8.58	3.988
EBAY/ATVI	(0.6249)	0.0100	9	13	69	756.91	(770.31)	1895.05	147.77	9.39	3.596
MAT/QCOM	0.4484	0.0670	9	12	75	851.85	(175.47)	2771.58	230.96	9.17	0.545
CELG/CHRW	0.4290	0.0510	8	12	67	910.27	(1114.27)	1250.19	104.18	8.33	1.215
MAT/GILD	0.8394	0.3720	8	12	67	1696.43	(1265.63)	1828.98	152.43	11.00	1.142
DELL/CHKP	0.9104	0.0520	7	10	70	1319.03	(641.20)	2832.60	283.26	10.70	0.251
NVDA/CSCO	0.4811	0.5360	6	10	60	538.19	(851.75)	583.47	58.35	10.80	0.733
INFY/FISV	(0.5006)	0.6200	5	10	50	293.48	(505.82)	(473.53)	(47.35)	11.40	0.660
CA/ISRG	0.3335	0.2600	7	10	70	719.79	(1200.56)	485.46	48.55	11.20	0.047
SYMC/CHKP	0.0685	0.3249	11	12	92	1258.25	(1048.63)	4041.07	336.84	8.25	0.325
AVERAGE			**7.8**	**11.3**	**68.7**	**$ 903.61**	**$ (828.52)**	**$ 1,624.26**	**$ 140.06**	**9.88**	**1.250**

Note: Co-integration value is p-value of Dickey-Fuller stationarity test

Both the correlation and co-integration methods' performance are tested using the ratio of log prices in this instance because this is the method used by Chan (2010) that we aim to replicate. Both methods return a similar number of average winners across the 10 pairs; 7.5

winners for correlation and 7.8 winners for co-integration. Both methods return a similar number of average trades per pair, 10.5 for correlation and 11.3 co-integration. Both strategies have similar average percentages of trades that are winners over the 10 pairs, and both methods seem to have very comparable maximum profits. The major difference in the performance of the two strategies is in the minimum profit; the correlation method returns an average minimum profit of -$1,274 across the 10 pairs, with the largest loss over -$3,300, based on an original investment of $10,000. In comparison, the co-integration method returned an average minimum profit of -$828 across the 10 pairs with the largest loss only -$1,266. Thus the co-integration method to returns an average profit across the 10 pairs of $1,624, clearly a more profitable strategy than the correlation method, which produced an average profit across the 10 pairs of only $1,270. The average length the position was open across the 10 pairs was also similar for both strategies. The data suggests that the co-integration method is more profitable, as argued by Chan.

We also performed analyses using the normalized ratio of log prices, the normalized price difference, and the normalized price ratio for the top 10 co-integrated pairs using standard deviation cutoffs from 1.0 to 3.0 at 0.5 intervals. These results can be found in Appendix D. It is important to note that the analysis in Appendix D suggests that there is a lot of variability in our conclusion that co-integration is a superior method when different simulation assumptions are made when one considers different pairs. On average, however, co-integration does seem to outperform correlation.

B. Trade Signal Decisions

We then chose one pair, the DELL/CHKP pair, to perform a more robust and in depth analysis of the co-integration trading methods using the standard deviation cutoffs 1.5 and 2.5. Figure 1 provides data on the price movements for both stocks from January to May 2012.

Figure 1: Price of DELL and CHKP from 01-03-2012 to 5-21-2013

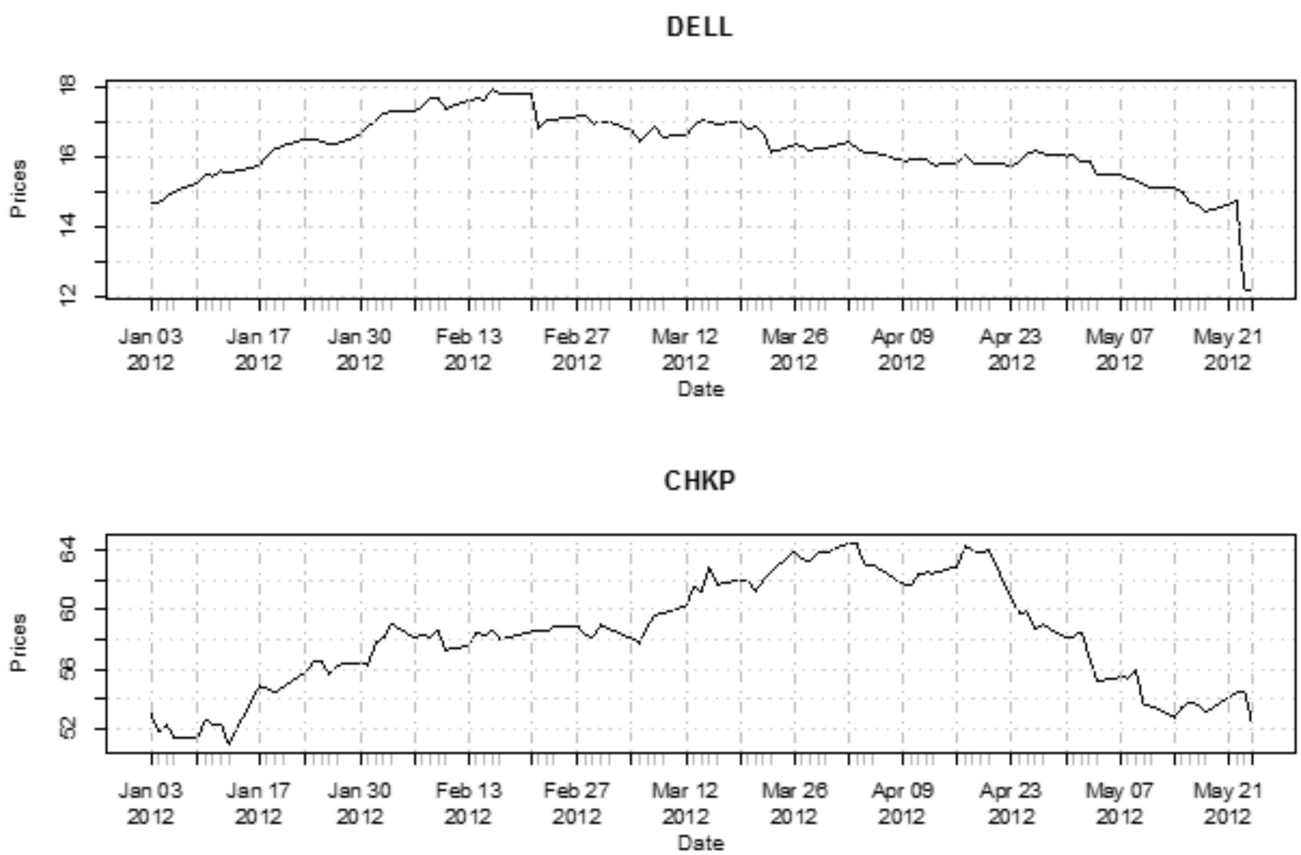
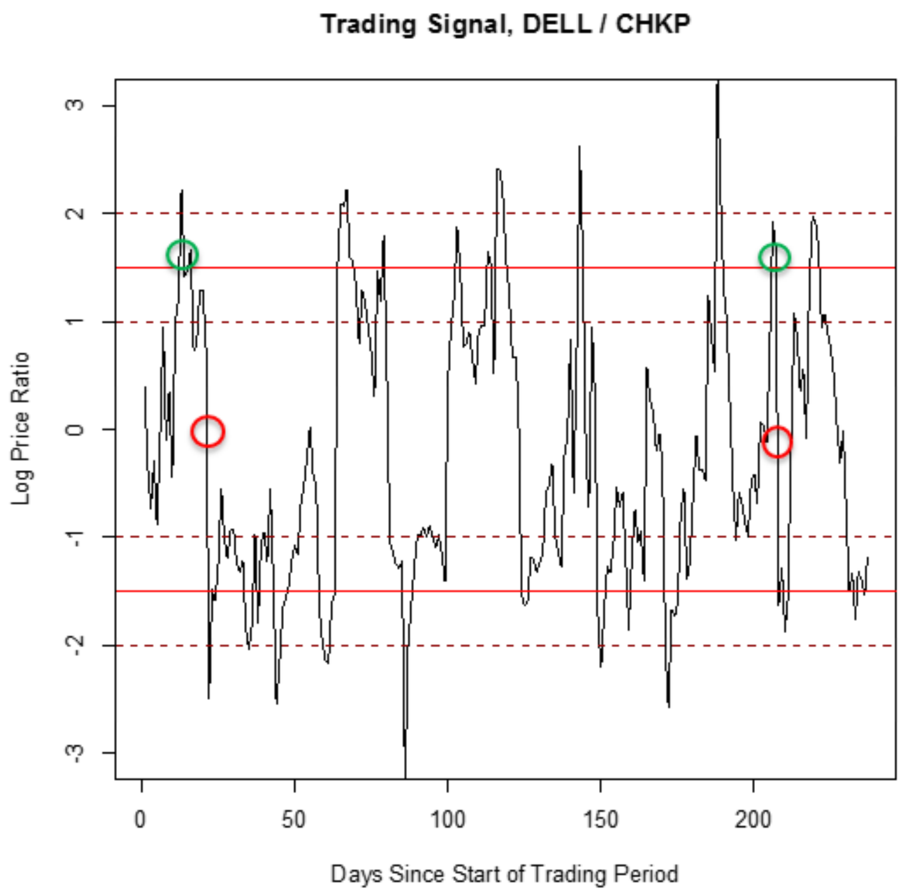


Figure 2 gives us a better idea of how the stocks move together, It specifically looks at the ratio of log prices between the two stocks, with the trading signal cut-off (solid red line) of ± 1.5 standard deviation.

Figure 2: Trading Signals for DELL/CHKP Pair using a 1.5 S.D. Cutoff under the Ratio of Log Prices Method

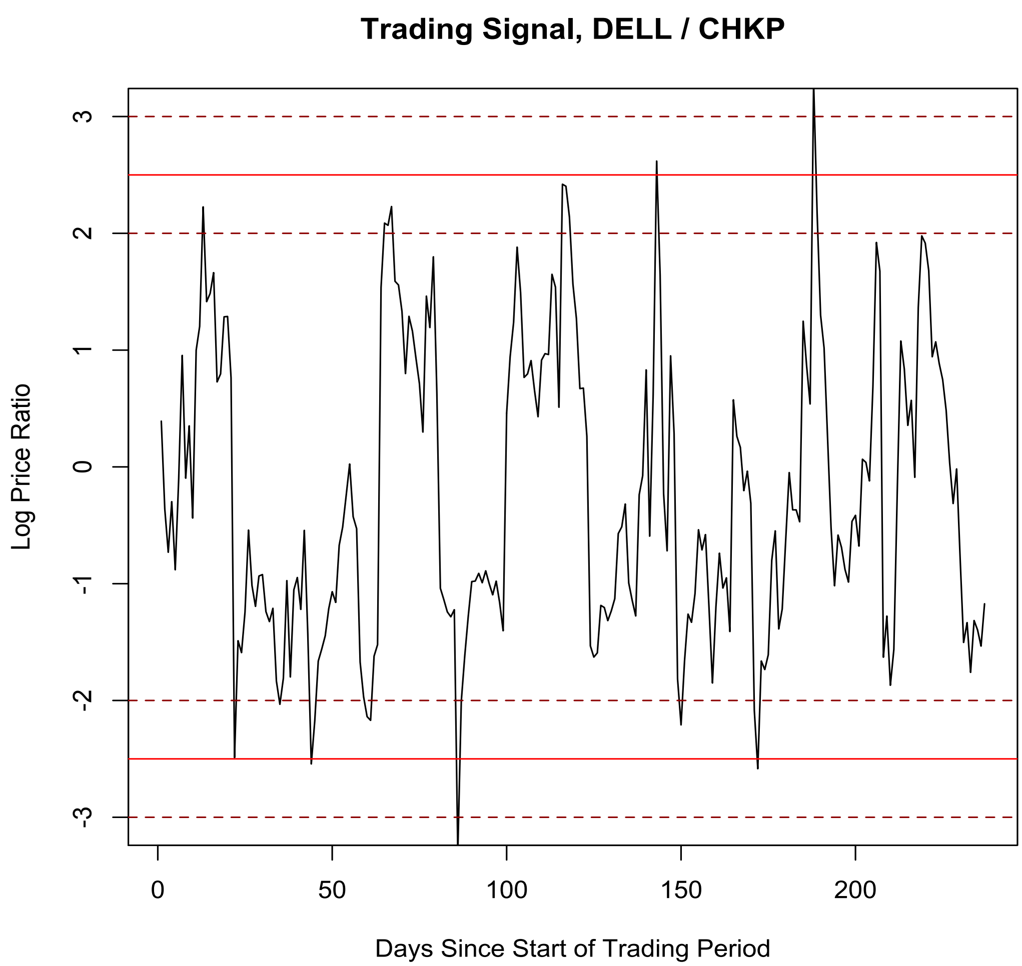


The red bands indicate the ±1.5 standard deviation cutoff. Trades are triggered every time the signal crosses these bands, and the positions are only closed when the spread reverts to zero. For example, at the ±1.5 standard deviation cutoff the first trade is triggered on day 26. The trade is given by the following script; "Short DELL at 17.67 and Long CHKP at 58.09 on day 26" and then position is "Closed on day 35 after 9 days", this trade is highlighted by the first green (position opens) and red (position closes) circles. The most profitable trade, highlighted by the second set of circles is triggered on day 201, where we "Short DELL at 9.72 and Long CHKP at 41.15." The trade then closes on day 206, after five days, producing a new profit high of $1319.03.

At the 1.5 standard deviation level, this method produces 15 separate trades, each one involving a cashless strategy of going long the stock that is below the spread and shorting the stock above the spread. The complete transcript of trades can be found in appendix C. Exactly 9 of the trades were profitable or 60%, the method produced a profit low of -$1148 and a profit high of $1319. The total profit was $3288, the average profit per trade was $219, and the positions were open an average of 10.87 days.

We then reproduce the same test at the ±2.5 standard deviation cutoff. The trading signal is reproduced in Figure 3.

Figure 3: Trading Signal for DELL/CHKP using a 2.5 S.D. Cutoff Under the Ratio of Log prices Model



Again, the trade is only triggered when the signal crosses the band. This produces 5 trades, of which all 5, or 100%, are profitable. The maximum profit is $1319 the minimum profit is $0. The total profit is $2380, and the average gain per trade is $476, while the average length of the position was only 9 days. Table 3 reproduces the trade summaries for the pair DELL/CHKP at the 1.0, 1.5, 2.0, 2.5, and 3.0 standard deviation cutoffs.

Table 3: Results for DELL/CHKP at Different S.D. Cutoffs Under the Ratio of Log Prices Model

Performance of DELL/CHKP in 2012-13 Simulation, Using the Ratio of Log Prices								
St. Dev. Cutoff	**Winners**	**Trades**	**Win %**	**Max**	**Min**	**Profit**	**Avg. Length**	**Avg Profit**
1.0 Cutoff	7	16	44	$1,319.03	($1,448.04)	$788.31	11.25	$49.27
1.5 Cutoff	9	15	60	1319.03	(1148.02)	3288.58	10.87	219.24
2.0 Cutoff	7	10	70	1319.03	(641.20)	2832.60	10.70	283.26
2.5 Cutoff	5	5	100	1319.03	0.00	2380.48	9.00	476.10
3.0 Cutoff	2	2	100	1319.03	0.00	1647.87	9.50	823.93

Naturally, the number of trades decreases as the standard deviation cutoff increases, as the necessary move of the trade signal becomes larger, and thus less likely. There is also a balance between the number of trades and the profitability of trades. For example, only 60% of the trades were profitable under the 1.5 standard deviation rule, but the total profit under this rule was $3288, whereas all trades under the 2.5 standard deviation rule were profitable, but total profit was only $2380. However, with an average profit per trade of more than two times the profit than the average profit for the 1.5 standard deviation cutoff, the 2.5 standard deviation cutoff is more profitable per trade. This highlights the different reasons for using different cutoffs; at lower standard deviation cutoffs, there will be a lower success ratio for trades, but there may be higher overall profit. Therefore, lower standard deviation cutoffs may turn away a risk averse investor who might prefer the higher cutoffs which have fewer trades with losses, but lower overall profits. This conclusion does not hold for all pairs tested in Appendix D.

Figure 4: Total Profits for DELL/CHKP at each S.D. Cutoff

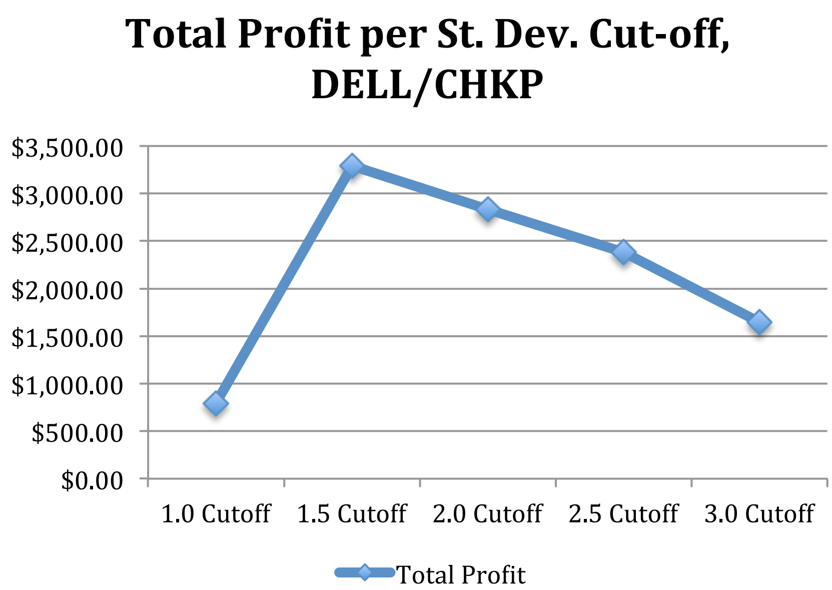


Figure 4, demonstrates that increasing the percentage of wins by raising the standard deviation cutoff produces higher profits up to a point, after which the reduced number of trades impact the total profit. In this instance, we find that the 1.5 cutoff is the most profitable rule, but our broader analysis suggest that the highest profitability is generated by cutoff rules in the range of 1.5 to 2.0 standard deviations.

We also ran this same test for the top 10 co-integration pairs from Table 1 using the ratio of log prices method, the normalized price difference method, and the normalized price ratio method. The condensed results are highlighted in Table 4, with complete results exhibited in Appendix D.

Table 4: Comparing Total Profit Under Different Cutoffs and Different Methods Using Co-Integration Pairs

Comparing Total Profit Under Trigger Cut-offs and Price Relationship Methods					
	St. Dev. Cut-offs				
	1.0	1.5	2.0	2.5	3.0
Log Price Ratio	$1,460.48	$1,427.52	$1,624.36	$840.12	$544.74
Norm. Difference	1133.42	1168.66	793.02	485.57	603.48
Price Ratio	1414.32	1402.20	1851.90	876.45	580.64

Note: Uses the averages of top 10 pairs by co-integration.

Table 4 shows that under the co-integration pairs selection model, determined previously a more effective method for selecting pairs, the 2.0 standard deviation cutoffs are the most profitable under the ratio of log prices and ratio of prices model. Additionally, Table 4 shows that the log price ratio and price ratio methods are generally more profitable than the normalized price difference method.

IV. Conclusion

Pairs trading is a statistical arbitrage method that is very popular with hedge funds and quantitative traders. The method involves finding pairs of stocks that exhibit a mean reverting trend to a static spread. Traders can define such a spread in different ways, such as normalizing a price difference between the two stocks, or taking a ratio of its prices or log prices. Additionally, they must establish a trading rule to open a position when the trading signal breaks through a set standard deviation cutoff. First, this paper analyzed two ways to select pairs, correlation and co-integration. We then tested our pairs for performance using three different methods. These included the ratio of log prices method, the normalized price difference, and the normalized price ratio for both the correlation and co-integration pairs using multiple standard deviation cutoffs.

Our analysis suggests that, in agreement with Chang's analysis, the top pairs ranked by co-integration on average performed better than the top pairs ranked by correlation. We also found that the ratio of log prices and normalized price ratio methods performed better than the normalized price difference, and that the standard deviation cutoffs did make a difference. In most cases there was a middle standard deviation band (1.5 or 2.0) that produced the highest profit for the ratio of log prices and normalized price ratio methods, proving to be an optimal range in the tradeoff between number of trades and number of successful trades (for the correlation pairs, the 1.0 and 1.5 standard deviation cutoffs were the most profitable). If the standard deviation bandis set too low, too many trades are initiated and the risk of losing money is greatly increased, and if the threshold is too high, although each trade is profitable (often highly so) the small number of trades do not generate as much total profit. As with other investment strategies, the trader must assess his own level of risk and let that guide his decisions in implementing his pairs trading strategy, primarily the standard deviation cut-off trade trigger rules.

References

Blitz, David, Joop Huij, Simon Lansdorp, and Marno Verbeek. "Short-Term Residual Reversal." 2012.

Caldeira, João F., and Guilherme V. Moura. "Selection of a Portfolio of Pairs Based on Cointegration: The Brazilian Case." 2012

Chan, Ngai Hang, "Time Series: Applications to Finance with R and S-Plus, Second Edition." Wiley, 2010.

Parzen, Michael. "Statistics 107: Lecture 16." Harvard University, Cambridge. 28 Mar. 2013. Lecture.

Perlin, Marcelo Scherer, "Evaluation of Pairs Trading Strategy at the Brazilian Financial Market", ICMA Centre/Reading University, UK.

R Code adapted from lectures and sections from Michael Parzen, Robin Gong, Ka Yeon Kook, Seong Hwang and Erik Otarola-Castillo

Ryan, Jeffrey A. *Quantitative Financial Modelling Framework*. 01 Jan. 2013. Quantmod 4.0.

Pairslog.com

Appendix A

Strategy Definitions

(1) Normalized Price Difference

$$N.P_a = (P_a - (\bar{P}_a, n = 14))/(\sigma(P_a, n = 14))$$

$$N.P_b = (P_b - (\bar{P}_b, n = 14))/(\sigma(P_b, n = 14))$$

$$N.P.Diff = P_a - P_b$$

where, $N.P._i$ is the normalized price for each stock, P_i is the stock's price, $\bar{P}_i$ is the stock's mean price, σ is the standard deviation, and $n = 14$ puts the mean and the standard deviation on a rolling 14-day basis. This notation is consistent in the next two methods.

(2) Normalized Price Ratio

$$Ratio = P_a/P_b$$

$$N.Ratio = (Ratio - (\overline{Ratio}, n = 14))/(\sigma(Ratio, n = 14))$$

(3) Normalized Price Log Ratio

$$L.Ratio = log(P_a/P_b)$$

$$N.L.Ratio = (L.Ratio - (\overline{L.Ratio}, n = 14))/(\sigma(L.Ratio, n = 14))$$

Appendix B

Table 5: Correlation and Dickey-Fuller Stationarity Test Values Compared from 2011-2012 to 2012-2013

Correlation and Co-Integration Dickey-Fuller Test Coefficients in Ranking Period (2011-12) and Trade Simulation Period (2012-13)				
Top 10 by Correlation	**2011**		**2012**	
	Correlation	**Co-integration**	**Correlation**	**Co-integration**
ADSK/ADBE	0.9635	0.1608	0.3510	0.0401
WCRX/LIFE	0.9635	0.0286	(0.5579)	0.7202
QGEN/LIFE	0.9573	0.2611	0.4013	0.5634
ILMN/BMC	0.9544	0.3034	0.1413	0.6954
PCAR/ LIFE	0.9536	0.0578	0.7049	0.2360
LRCX/AMAT	0.9493	0.0323	0.6116	0.7414
MU/AMAT	0.9455	0.2083	0.8523	0.2487
WCRX/QGEN	0.9452	0.2968	(0.0378)	0.3598
NIHD/ILMN	0.9450	0.2401	0.0179	0.2217
PCAR/EXPD	0.9421	0.2429	0.5882	0.1381

Top 10 by Co-integration	**2011**		**2012**	
	Correlation	**Co-integration**	**Correlation**	**Co-integration**
DLTR/ATVI	0.7293	0.0100	0.4640	0.5767
EBAY/ATVI	0.1434	0.0100	(0.6249)	0.0100
MAT/QCOM	0.3159	0.0100	0.4484	0.0712
CELG/CHRW	(0.1185)	0.0100	0.4290	0.0509
MAT/GILD	0.3529	0.0100	0.8394	0.3723
DELL/CHKP	0.7358	0.0100	0.9104	0.0522
NVDA/CSCO	0.6373	0.0100	0.4911	0.5359
INFY/FISV	0.8040	0.0100	(0.5006)	0.6200
CA/ISRG	(0.5465)	0.0100	0.3335	0.2603
SYMC/CHKP	0.1720	0.0100	0.0685	0.5421

This Table highlights the variability of Correlation and Co-integration values from year to year.

This could be a cause of variability in our results as compared to a different time frame.

Appendix C

Trading Scripts for DELL/CHKP using Ratio of Log Prices Method

Trading Script 1.5 S.D. Cutoff

Short DELL at 17.67 and Long CHKP at 58.09 on day 26
Closed on day 35 after 9 days
New Profit High 580.649642754259
Short CHKP at 58.87 and Long DELL at 17.06 on day 37
Closed on day 68 after 31 days
New Profit Low -1148.01704074649
Short CHKP at 62.48 and Long DELL at 15.77 on day 71
Closed on day 77 after 6 days
Short DELL at 15.84 and Long CHKP at 59.8 on day 78
Closed on day 94 after 16 days
Short CHKP at 54.54 and Long DELL at 12.23 on day 99
Closed on day 113 after 14 days
Short DELL at 12.16 and Long CHKP at 49.16 on day 116
Closed on day 137 after 21 days
Short CHKP at 49.93 and Long DELL at 11.97 on day 138
Closed on day 153 after 15 days
Short DELL at 11.93 and Long CHKP at 47.94 on day 156
Closed on day 158 after 2 days
Short CHKP at 50.25 and Long DELL at 11.43 on day 162
Closed on day 178 after 16 days
Short CHKP at 47.32 and Long DELL at 9.94 on day 184
Closed on day 198 after 14 days
Short DELL at 9.72 and Long CHKP at 41.15 on day 201
Closed on day 206 after 5 days
New Profit High 1319.02504637755
Short DELL at 9.45 and Long CHKP at 44.23 on day 219
Closed on day 221 after 2 days
Short CHKP at 45.22 and Long DELL at 8.87 on day 223
Closed on day 226 after 3 days
Short DELL at 10.17 and Long CHKP at 45.16 on day 232
Closed on day 241 after 9 days
Short CHKP at 48.03 and Long DELL at 10.36 on day 244

Trading Script 2.5 S.D Cutoff

Short CHKP at 62.62 and Long DELL at 16.13 on day 57
Closed on day 68 after 11 days
New Profit High 16.9138366549505
Short CHKP at 54.54 and Long DELL at 12.23 on day 99

Closed on day 113 after 14 days
New Profit High 328.841142495279
Short DELL at 11.93 and Long CHKP at 47.94 on day 156
Closed on day 158 after 2 days
New Profit High 500.471915683931
Short CHKP at 48.12 and Long DELL at 9.77 on day 185
Closed on day 198 after 13 days
Short DELL at 9.72 and Long CHKP at 41.15 on day 201
Closed on day 206 after 5 days
New Profit High 1319.02504637755

Appendix D

Table 6: Results for Top 10 pairs by Co-integration under the Ratio of Log Price Pairs Trading Method

Ratio of Log Price/ Pairs Trading Routine Simulation Results for Top 10 Pairs by Co-integration, Different Standard Deviation Cut-off Levels

1.0 Cutoff	Winners	Trades	Win %	Max Profit	Min Profit	Total Profit	Avg. Length	Avg. Profit
DLTR/ATVI	12	20	60	$859.78	($877.34)	$1,125.28	8.7	$56.26
EBAY/ATVI	14	21	67	492.91	(850.23)	239.69	8.6	11.37
MAT/QCOM	18	24	75	913.43	(881.97)	4314.73	7.6	179.78
CELG/CHRW	12	19	63	910.27	(1378.48)	1396.51	9.2	73.50
MAT/GILD	15	20	75	1827.67	(1499.25)	4656.75	8.9	232.84
DELL/CHKP	7	16	44	1319.03	(1448.04)	788.31	11.3	49.27
NVDA/CSCO	10	19	53	849.80	(851.75)	(142.37)	9.2	(7.49)
INFY/FISV	14	23	61	1396.44	(568.92)	1801.71	7.6	78.34
CA/ISRG	7	14	50	719.79	(1955.89)	(2585.89)	12.6	(184.71)
SYMC/CHKP	19	23	83	1258.25	(2595.24)	3010.03	7.7	130.87
Average	**12.8**	**19.9**	**63.1**	**$1,054.74**	**($1,290.71)**	**$1,460.48**	**9.1**	**$62.00**

1.5 Cutoff	Winners	Trades	Win %	Max Profit	Min Profit	Total Profit	Avg. Length	Avg. Profit
DLTR/ATVI	12	18	67	$859.78	($877.34)	$1,301.63	9.0	$72.31
EBAY/ATVI	10	16	62	413.67	(770.31)	722.18	10.0	45.14
MAT/QCOM	14	18	78	999.41	(820.77)	3499.92	8.0	194.44
CELG/CHRW	8	15	53	910.27	(1378.48)	90.07	10.4	6.00
MAT/GILD	14	17	82	1696.43	(1499.25)	4752.42	9.4	279.55
DELL/CHKP	9	15	60	1319.03	(1148.02)	3288.58	10.9	219.24
NVDA/CSCO	10	16	62	849.80	(851.75)	633.26	9.9	39.58
INFY/FISV	9	16	56	311.24	(505.82)	(95.65)	9.6	(5.98)
CA/ISRG	4	10	40	719.79	(1955.89)	(2384.99)	15.8	(238.50)
SYMC/CHKP	14	17	82	1258.25	(2154.21)	2467.82	8.4	145.17
Average	**10.4**	**15.8**	**64.2**	**$933.77**	**($1,196.18)**	**$1,427.52**	**10.1**	**$75.70**

2.0 Cutoff	Winners	Trades	Win %	Max Profit	Min Profit	Total Profit	Avg. Length	Avg. Profit
DLTR/ATVI	8	12	67	$691.91	($711.52)	$1,027.71	8.6	85.64
EBAY/ATVI	9	13	69	756.91	(770.31)	1895.05	9.4	145.77
MAT/QCOM	9	12	75	851.85	(175.47)	2771.58	9.2	230.96
CELG/CHRW	8	12	67	910.27	(1114.27)	1250.19	8.3	104.18
MAT/GILD	8	12	67	1696.43	(1265.63)	1828.98	11.0	152.42
DELL/CHKP	7	10	70	1319.03	(641.20)	2832.60	10.7	283.26
NVDA/CSCO	6	10	60	538.19	(851.75)	583.47	10.8	58.35
INFY/FISV	5	10	50	293.48	(505.82)	(473.53)	11.4	(47.35)
CA/ISRG	7	10	70	719.79	(1200.56)	485.46	11.2	48.55
SYMC/CHKP	11	12	92	1258.25	(1048.63)	4042.07	8.3	336.84
Average	**7.8**	**11.3**	**68.7**	**$903.61**	**($828.52)**	**$1,624.36**	**9.9**	**$139.86**

2.5 Cutoff	Winners	Trades	Win %	Max Profit	Min Profit	Total Profit	Avg. Length	Avg. Profit
DLTR/ATVI	3	4	75	$777.26	($430.52)	$1,079.23	7.3	$269.56
EBAY/ATVI	2	5	40	413.67	(381.83)	(14.23)	14.2	(2.85)
MAT/QCOM	4	5	80	802.35	(157.91)	1919.02	8.0	383.80
CELG/CHRW	3	3	100	910.27	0.00	1425.70	5.7	475.23
MAT/GILD	3	5	60	731.45	(359.01)	673.80	12.0	134.76
DELL/CHKP	5	5	100	1319.03	0.00	2380.48	9.0	476.10
NVDA/CSCO	1	4	25	315.53	(851.75)	(1112.51)	15.5	(278.13)
INFY/FISV	3	6	50	293.48	(505.82)	(323.23)	11.3	(53.87)
CA/ISRG	2	3	67	630.61	9.00	763.69	5.7	254.56
SYMC/CHKP	3	3	100	980.79	0.00	1609.24	6.7	536.41
Average	**2.9**	**4.3**	**69.7**	**$717.44**	**($267.78)**	**$840.12**	**9.5**	**$219.56**

3.0 Cutoff	Winners	Trades	Win %	Max Profit	Min Profit	Total Profit	Avg. Length	Avg. Profit
DLTR/ATVI	0	0	-	$0.00	$0.00	$0.00	-	-
EBAY/ATVI	1	1	100	413.67	0.00	413.67	-	413.67
MAT/QCOM	3	3	100	802.35	0.00	1925.51	6.7	641.84
CELG/CHRW	1	1	100	910.27	0.00	910.27	3.0	910.27
MAT/GILD	1	2	50	239.45	(359.01)	(119.56)	15.0	(59.78)
DELL/CHKP	2	2	100	1319.03	0.00	1647.87	9.5	823.93
NVDA/CSCO	0	2	0	0.00	(851.75)	(994.68)	18.5	(497.34)
INFY/FISV	2	3	67	293.48	(505.82)	52.91	11.0	17.64
CA/ISRG	1	1	100	630.61	0.00	630.61	5.0	630.61
SYMC/CHKP	1	1	100	980.79	0.00	980.79	5.0	980.79
Average	**1.2**	**1.6**	**79.7**	**$558.97**	**($171.66)**	**$544.74**	**9.2**	**$429.07**

Table 7: Results for Top 10 pairs by Co-integration under the Normalized Price Difference Pairs Trading Method

Normalized Price Difference / Pairs Trading Routine Simulation Results for Top 10 Pairs by Co-integration, Different Standard Deviation Cut-off Levels								
1.0 Cutoff	**Winners**	**Trades**	**Win %**	**Max Profit**	**Min Profit**	**Total Profit**	**Avg. Length**	**Avg. Profit**
DLTR/ATVI	15	24	62	$691.91	($1,423.11)	$822.52	6.7	$34.27
EBAY/ATVI	14	22	64	351.90	(1168.09)	(44.78)	7.7	(2.04)
MAT/QCOM	13	20	65	913.43	(851.55)	2325.87	5.7	116.29
CELG/CHRW	21	27	78	959.61	(1592.50)	3279.89	5.5	121.48
MAT/GILD	12	20	60	1023.07	(1647.61)	1515.54	7.4	75.78
DELL/CHKP	11	19	58	917.24	(846.39)	1801.56	7.3	94.82
NVDA/CSCO	9	18	50	773.30	(851.75)	1153.16	7.3	64.06
INFY/FISV	12	20	60	525.57	(796.85)	(166.91)	8.2	(8.35)
CA/ISRG	10	17	59	630.61	(1526.25)	(2159.72)	7.2	(127.04)
SYMC/CHKP	17	23	74	1258.25	(960.92)	2807.06	5.7	122.05
1.0 Average	**13.4**	**21**	**63**	**$804.49**	**($1,166.50)**	**$1,133.42**	**6.9**	**$49.13**

1.5 Cutoff	**Winners**	**Trades**	**Win %**	**Max Profit**	**Min Profit**	**Total Profit**	**Avg. Length**	**Avg. Profit**
DLTR/ATVI	13	19	68	$691.91	($1,022.15)	$1,394.84	7.3	$73.41
EBAY/ATVI	11	16	69	370.16	(387.78)	1773.69	7.6	110.86
MAT/QCOM	8	12	67	802.35	(802.98)	1538.02	5.8	128.17
CELG/CHRW	11	17	65	959.61	(1592.50)	2071.70	6.4	121.86
MAT/GILD	13	17	76	1023.07	(1499.25)	2609.40	7.3	153.49
DELL/CHKP	10	17	59	917.24	(846.39)	1759.72	7.1	103.51
NVDA/CSCO	9	13	69	773.30	(488.00)	1902.67	7.2	146.36
INFY/FISV	9	15	60	632.92	(796.85)	(392.40)	8.8	(26.16)
CA/ISRG	4	11	36	630.61	(1526.25)	(3081.22)	9.0	(297.19)
SYMC/CHKP	7	12	58	1258.25	(746.32)	2110.22	7.8	175.85
1.5 Average	**9.5**	**14.9**	**62.7**	**$805.94**	**($970.85)**	**$1,168.66**	**7.4**	**$69.02**

2.0 Cutoff	**Winners**	**Trades**	**Win %**	**Max Profit**	**Min Profit**	**Total Profit**	**Avg. Length**	**Avg. Profit**
DLTR/ATVI	8	14	57	$691.91	($853.79)	($216.48)	8.6	(15.46)
EBAY/ATVI	6	11	55	365.48	(381.83)	1773.69	9.1	71.91
MAT/QCOM	4	7	57	578.36	(607.53)	700.13	7.6	100.02
CELG/CHRW	11	15	73	767.91	(1328.91)	2432.43	5.5	162.16
MAT/GILD	10	13	77	991.64	(1499.25)	1537.33	8.3	118.26
DELL/CHKP	9	14	64	917.24	(846.39)	2628.74	7.1	187.77
NVDA/CSCO	6	10	60	849.80	(398.95)	2460.70	7.6	246.07
INFY/FISV	5	10	50	632.92	(505.82)	(658.49)	9.9	(65.85)
CA/ISRG	3	9	33	630.61	(1526.25)	(2674.75)	9.0	(297.19)
SYMC/CHKP	6	9	67	538.21	(888.29)	(53.07)	8.4	184.69
2.0 Average	**6.8**	**11.2**	**59.3**	**$696.41**	**($883.70)**	**$793.02**	**8.1**	**$69.24**

2.5 Cutoff	Winners	Trades	Win %	Max Profit	Min Profit	Total Profit	Avg. Length	Avg. Profit
DLTR/ATVI	6	10	60	$412.46	($573.26)	$261.28	7.3	$26.13
EBAY/ATVI	7	10	70	756.91	(381.83)	1650.92	8.9	165.09
MAT/QCOM	3	6	50	578.36	(607.53)	505.64	8.5	84.27
CELG/CHRW	6	8	75	820.22	(1328.91)	555.24	7.3	69.40
MAT/GILD	7	10	70	798.95	(1499.25)	573.45	9.4	57.35
DELL/CHKP	9	12	75	917.24	(830.83)	2354.43	7.5	196.20
NVDA/CSCO	4	6	67	833.33	(426.89)	1245.20	9.7	207.53
INFY/FISV	1	5	20	43.59	(326.84)	(662.15)	14.2	(132.43)
CA/ISRG	1	7	14	125.47	(1063.76)	(2736.53)	8.4	(390.93)
SYMC/CHKP	5	6	83	444.38	(231.64)	1108.17	7.2	184.69
2.5 Average	**4.9**	**8**	**58.4**	**$573.09**	**($727.07)**	**$485.57**	**8.8**	**$46.73**

3.0 Cutoff	Winners	Trades	Win %	Max Profit	Min Profit	Total Profit	Avg. Length	Avg. Profit
DLTR/ATVI	7	7	100	$777.26	$0.00	$2,046.58	9.1	$292.37
EBAY/ATVI	5	8	62	756.91	(381.83)	1381.78	9.0	172.72
MAT/QCOM	1	3	33	579.36	(175.47)	266.84	8.0	88.95
CELG/CHRW	3	4	75	375.31	(192.01)	749.17	6.3	187.29
MAT/GILD	3	4	75	619.79	(1499.25)	(533.10)	13.8	(133.28)
DELL/CHKP	6	8	75	917.24	(652.77)	1037.90	9.3	129.74
NVDA/CSCO	3	4	75	833.33	(48.82)	1280.78	7.0	320.20
INFY/FISV	2	3	67	244.24	(63.32)	224.51	10.0	74.84
CA/ISRG	1	7	14	125.47	(873.19)	(1452.17)	7.0	(207.45)
SYMC/CHKP	3	3	100	645.99	0.00	1032.53	7.3	344.18
3.0 Average	**3.4**	**5.1**	**67.6**	**$587.49**	**($388.67)**	**$603.48**	**8.7**	**$126.96**

Table 8: Results for Top 10 pairs by Co-integration under the Ratio of Prices Pairs Trading Method

Price Ratio / Pairs Trading Routine Simulation Results for Top 10 Pairs by Co-integration, Different Standard Deviation Cut-off Levels								
1.0 Cutoff	**Winners**	**Trades**	**Win %**	**Max Profit**	**Min Profit**	**Total Profit**	**Avg. Length**	**Avg. Profit**
DLTR/ATVI	12	20	60	$859.78	($877.34)	$1,125.28	8.7	$56.26
EBAY/ATVI	14	21	67	492.91	(850.23)	238.69	8.6	11.37
MAT/QCOM	18	24	75	913.43	(881.97)	4314.73	7.6	179.78
CELG/CHRW	12	19	63	910.27	(1378.48)	1396.51	9.2	73.50
MAT/GILD	15	20	75	1827.67	(1499.25)	4656.75	8.9	232.84
DELL/CHKP	6	15	40	1319.03	(1448.04)	306.32	11.5	20.42
NVDA/CSCO	10	19	53	849.80	(851.75)	(142.37)	9.2	(7.49)
INFY/FISV	14	23	61	1396.44	(568.92)	1834.20	7.6	79.75
CA/ISRG	7	14	50	719.79	(1955.89)	(2585.89)	12.6	(184.71)
SYMC/CHKP	19	23	83	1258.25	(2595.24)	2998.94	7.6	130.39
1.0 Average	**12.7**	**19.8**	**62.7**	**$1,054.74**	**($1,290.71)**	**$1,414.32**	**9.1**	**$59.21**
1.5 Cutoff	**Winners**	**Trades**	**Win %**	**Max Profit**	**Min Profit**	**Total Profit**	**Avg. Length**	**Avg. Profit**
DLTR/ATVI	12	17	65	$859.78	($877.34)	$1,125.28	9.3	$59.28
EBAY/ATVI	10	16	62	413.67	(770.31)	722.18	10.0	45.14
MAT/QCOM	14	18	78	999.41	(820.77)	3499.92	8.0	194.44
CELG/CHRW	8	15	53	910.27	(1378.48)	176.37	10.3	11.76
MAT/GILD	14	17	82	1696.43	(1499.25)	4752.42	9.4	279.55
DELL/CHKP	9	15	60	1319.03	(1148.02)	3288.58	10.9	219.24
NVDA/CSCO	10	16	62	849.80	(851.75)	633.26	9.9	39.58
INFY/FISV	9	16	56	311.24	(505.82)	(95.65)	9.6	(5.98)
CA/ISRG	4	10	40	719.79	(1955.89)	(2384.99)	15.8	(238.50)
SYMC/CHKP	14	17	82	1258.25	(2317.43)	2304.60	8.5	135.56
1.5 Average	**10.4**	**15.7**	**64**	**$933.77**	**($1,212.51)**	**$1,402.20**	**10.2**	**$74.01**
2.0 Cutoff	**Winners**	**Trades**	**Win %**	**Max Profit**	**Min Profit**	**Total Profit**	**Avg. Length**	**Avg. Profit**
DLTR/ATVI	8	12	67	$691.91	($711.52)	$1,027.71	8.6	85.64
EBAY/ATVI	9	13	69	756.91	(770.31)	1895.05	9.4	145.77
MAT/QCOM	9	12	75	851.85	(454.07)	2492.98	9.2	207.75
CELG/CHRW	8	12	67	910.27	(1114.27)	1250.19	8.3	104.18
MAT/GILD	10	13	77	1696.43	(1265.63)	3941.99	10.1	303.23
DELL/CHKP	8	11	73	1319.03	(641.20)	3162.99	10.5	287.54
NVDA/CSCO	7	11	64	538.19	(851.75)	903.40	10.5	82.13
INFY/FISV	5	10	50	293.48	(505.82)	(473.53)	11.4	(47.35)
CA/ISRG	7	10	70	719.79	(1200.56)	485.46	11.2	48.55
SYMC/CHKP	11	12	92	1258.25	(1048.63)	3832.73	8.3	319.39
2.0 Average	**8.2**	**11.6**	**70.4**	**$903.61**	**($856.38)**	**$1,851.90**	**9.7**	**$153.68**

2.5 Cutoff	Winners	Trades	Win %	Max Profit	Min Profit	Total Profit	Avg. Length	Avg. Profit
DLTR/ATVI	3	4	75	$777.26	($430.52)	$1,079.23	7.3	$269.56
EBAY/ATVI	2	5	40	413.67	(381.83)	(14.23)	14.2	(2.85)
MAT/QCOM	4	5	80	802.35	(157.91)	1919.02	8.0	383.80
CELG/CHRW	3	3	100	910.27	0.00	1425.70	5.7	475.23
MAT/GILD	3	6	50	731.45	(359.01)	603.72	11.8	100.62
DELL/CHKP	5	5	100	1319.03	0.00	2380.48	9.0	476.10
NVDA/CSCO	1	3	33	315.53	(851.75)	(679.15)	15.7	(226.38)
INFY/FISV	3	6	50	293.48	(505.82)	(323.23)	11.3	(53.87)
CA/ISRG	2	3	67	630.61	9.00	763.69	5.7	254.56
SYMC/CHKP	3	3	100	980.79	0.00	1609.24	6.7	536.41
2.5 Average	**2.9**	**4.3**	**69.5**	**$717.44**	**($267.78)**	**$876.45**	**9.5**	**$221.32**

3.0 Cutoff	Winners	Trades	Win %	Max Profit	Min Profit	Total Profit	Avg. Length	Avg. Profit
DLTR/ATVI	0	0	nan	$0.00	$0.00	$0.00	nan	nan
EBAY/ATVI	1	1	100	413.67	0.00	413.67	nan	413.67
MAT/QCOM	3	3	100	802.35	0.00	1925.51	6.7	641.84
CELG/CHRW	1	1	100	910.27	0.00	910.27	3.0	910.27
MAT/GILD	1	1	100	239.45	0.00	239.45	11.0	239.45
DELL/CHKP	2	2	100	1319.03	0.00	1647.87	9.5	823.93
NVDA/CSCO	0	2	0	0.00	(851.75)	(994.68)	18.5	(497.34)
INFY/FISV	2	3	67	293.48	(505.82)	52.91	11.0	17.64
CA/ISRG	1	1	100	630.61	0.00	630.61	5.0	630.61
SYMC/CHKP	1	1	100	980.79	0.00	980.79	5.0	980.79
3.0 Average	**1.2**	**1.5**	**85.2**	**$558.97**	**($135.76)**	**$580.64**	**8.7**	**$462.32**

Appendix E
R Code

```
#Eric Westphal, Jimmy Field, Graham Topol
#Statistics 107
#Final Project
#Statistical Arbitrage

## Simple


# Pairs Trading using the Difference of the Means[9]

pairs.trading = function(ticker1, ticker2, from = "2011-01-01", to = Sys.Date(), type
 = "z.diff", quiet = T, threshold = 2, amount = 10000, pgraph = T)
{
  #get prelim. pair information
  require(quantmod)
  require(tseries)
  names = c(ticker1,ticker2)
  y = getSymbols(ticker1, auto.assign = F, from = from, to = to)
  x = getSymbols(ticker2, auto.assign = F, from = from, to = to)
  y.ad = as.numeric(Ad(y))
  x.ad = as.numeric(Ad(x))
  corr = cor(y.ad,x.ad)
  fit = lm(y.ad ~ -1+x.ad)
  beta = coef(fit)[1]
  spread = as.numeric(y.ad-beta*x.ad)
  adf = adf.test(spread, alternative="stationary", k=0)


  #Define pair methods
  #Use 14 day moving averages
  if(type == "z.diff")
  {
    y.norm = (y.ad - runMean(y.ad, n=14))/runSD(y.ad,14)
    x.norm = (x.ad - runMean(x.ad, n=14))/runSD(x.ad,14)
    compstat = y.norm - x.norm
  }
  else if(type == "ratio")
  {
    rat = y.ad/x.ad
    compstat = (rat - runMean(rat,14))/runSD(rat,14)
  }
  else if(type == "logratio")
  {
    rat = log(y.ad/x.ad)
    compstat = (rat - runMean(rat,14))/runSD(rat,14)
  }
  else
  {
  print(paste("Type must be either left blank, z.diff, ratio or logratio"))
  }

  ##Plots
  if(type=="z.diff")
  {
  numdays = length(compstat)
  if(pgraph == T)
  {
```

[9] Code influenced by or adapted from Michael Parzen, Seong Hwang, Erik Otarola-Castillo and Robin Gong (Harvard University, Statistics 107) and Ngai Hang Chan.

```
    par(mfcol=c(3,1))
    plot(Ad(y)[1:100], ylab = "Prices", xlab = "Date", main = ticker1)
    plot(Ad(x)[1:100], ylab = "Prices", xlab = "Date", main = ticker2)
}

  plot(compstat[14:numdays], ylab = "Normalized Difference", xlab = "Days Since Start
of Trading Period", main = paste("Trading Signal,",ticker1,"-",ticker2), type="l",
ylim=c(-3,3))
  cutoff = threshold
  abline(h=c(cutoff,-cutoff), col="red")
    abline(h=c(cutoff+.5,-cutoff-.5), col="dark red", lty=2, lwd = 1)
    abline(h=c(cutoff-.5,-cutoff+.5), col="dark red", lty=2, lwd = 1)


  }
  else if(type == "ratio")
  {
  numdays = length(compstat)
  if(pgraph == T)
  {
    par(mfcol=c(3,1))
    plot(Ad(y)[1:100], ylab = "Prices", xlab = "Date", main = ticker1)
    plot(Ad(x)[1:100], ylab = "Prices", xlab = "Date", main = ticker2)
}

  plot(compstat[14:numdays], ylab = "Price Ratio", xlab = "Days Since Start of Trading
Period", main = paste("Trading Signal,",ticker1,"/",ticker2), type="l", ylim=c(-3,3))
  cutoff = threshold
  abline(h=c(cutoff,-cutoff), col="red")
  abline(h=c(cutoff+.5,-cutoff-.5), col="dark red", lty=2, lwd = 1)
  abline(h=c(cutoff-.5,-cutoff+.5), col="dark red", lty=2, lwd = 1)
  }
  else if(type == "logratio")
  {
  numdays = length(compstat)
  if(pgraph == T)
  {
    par(mfcol=c(3,1))
    plot(Ad(y)[1:100], ylab = "Prices", xlab = "Date", main = ticker1)
    plot(Ad(x)[1:100], ylab = "Prices", xlab = "Date", main = ticker2)
}

  plot(compstat[14:numdays], ylab = "Log Price Ratio", xlab = "Days Since Start of
Trading Period", main = paste("Trading Signal,",ticker1,"/",ticker2), type="l",
ylim=c(-3,3))
  cutoff = threshold
  abline(h=c(cutoff,-cutoff), col="red")
  abline(h=c(cutoff+.5,-cutoff-.5), col="dark red", lty=2, lwd = 1)
  abline(h=c(cutoff-.5,-cutoff+.5), col="dark red", lty=2, lwd = 1)
  }


  #declare variables of performance
  y.traded = x.traded = 0
  current = "neither"
  profit = 0
  maxprofit = minprofit = avgprofit = numtrades = winners = 0
  tradeLength = 0


  #implement trade signals
  for(i in 14:numdays)
  {
    if(compstat[(i)] > threshold & current == "neither")
      {
```

```
      y.traded = (-amount/y.ad[i])
      x.traded = (amount/x.ad[i])
      current = "x"
      numtrades = numtrades + 1
      enterDate = i
      enterPrice1 = y.ad[i]
      enterPrice2 = x.ad[i]
      if(quiet != T)
      {
      print(paste("Short",ticker1,"at",y.ad[i],"and Long",ticker2,"at",x.ad[i],"on
day",i))
      }
    }

  if(compstat[(i)] < (-threshold) & current == "neither")
    {
      y.traded = (amount/y.ad[i])
      x.traded = (-amount/x.ad[i])
      current = "y"
      numtrades = numtrades + 1
      enterDate = i
      enterPrice1 = y.ad[i]
      enterPrice2 = x.ad[i]
      if(quiet != T)
      {
      print(paste("Short",ticker2,"at",x.ad[i],"and Long",ticker1,"at",y.ad[i],"on
day",i))
      }
    }

  #closes position, ensuring nothing is updated if the last trade isn't closed other
than numtrades
  if((compstat[i] < 0 & current == "x") | (compstat[i] > 0 & current == "y"))
    {
      profit.temp = y.traded*y.ad[i] + x.traded*x.ad[i]
      profit = profit + profit.temp
      winners = winners + (profit.temp>0)
      temp.tradeLength = i - enterDate
      tradeLength = tradeLength + temp.tradeLength
      if(quiet != T)
      {
       print(paste("Closed on day",i,"after",temp.tradeLength,"days"))
       if(profit.temp > maxprofit)
       {
         print(paste("New Profit High",profit.temp))
       }
       if(profit.temp < minprofit)
       {
         print(paste("New Profit Low",profit.temp))
       }
      }
      maxprofit = max(maxprofit,profit.temp)
      minprofit = min(minprofit,profit.temp)
      y.traded=0
      x.traded=0
      current="neither"
    }
 }

 avgprofit = profit/numtrades
 avglength = tradeLength/numtrades
 avglength

 #return results
 results = c(round(winners,0), round(numtrades,0), round(100*winners/numtrades,0),
```

```
 round(maxprofit,2), round(minprofit,2), round(profit,2), round(avglength,3),
 round(corr,4), round(beta,4), round(avgprofit,2), round(adf$p.value, 3))
  names(results) = c("winners", "numtrades", "winpct", "max", "min", "profit",
 "avgtradeLength", "corr", "beta", "average profit", "adf.test")
  return(results)

}

#SAMPLE CALLS
#pairs.trading("AMP","OZM", from = "2011-01-01")
#pairs.trading("JNJ","HD",from = "2012-01-01", to = "2013-01-01", pgraph = F, type =
 "logratio")
#just use logratio
#pgraph T means three graphs, pgraph F means 1
# quiet=F, you can see all of the trades
#keep threshold at 2
#amount =$10000
#call each pair twice= 1 for finding period and 1 for trading period

manyCutoffs <- function(ticker1, ticker2, from = "2011-01-01", to = Sys.Date(), type =
 "z.diff")
{
 #create matrix for results
 table = matrix(0,0,nrow=5,ncol=11)
 colnames(table) <- c("winners", "numtrades", "winpct", "max", "min", "profit",
 "Avg.Length", "corr", "beta", "Avg.Profit", "adf.test.p")
 rownames(table) <- c("1 SD", "1.5 SD", "2 SD", "2.5 SD", "3 SD")

 #get results for different cutoffs
 for (i in 0:4)
 {
  results =pairs.trading(ticker1, ticker2, from = from, to = to, type = type, quiet =
 T, threshold = (1 + i*.5))
  table[(1+i),] = results
 }
 return(table)
}

#SAMPLE CALL
#manyCutoffs("AMP","OZM", from = "2011-01-01", type="logratio")


## ANALYSIS PART 2 - Selecting Pairs
## $5,000 each way in each pair, top 10 pairs using Corr and Coint from NASDAQ,
 totalling a potential $100,000 in portfolio


#Correlation Method
selectpairsCORR <- function(from = "2011-01-01", to = "2012-01-01")
{
 npairs<-10 #number of top pairs to trade
 cutoff<-threshold

 #Read file and prepare matrix with price series
 stocks<-read.csv(file.choose(), header=T)
 tickers<-as.character(stocks[,2])
 n<-length(tickers)
 temp.stk<-getSymbols(tickers[1], auto.assign = F, from = from, to = to)
 temp.prices<-as.numeric(Ad(temp.stk))
 len<-length(temp.prices)
 Data = matrix(0,0,nrow=len,ncol=n)
 colnames(Data) <- c(tickers)
 for(i in 1:n)
 {
  stock<-getSymbols(tickers[i], auto.assign = F, from = from, to = to)
```

```
  prices<-as.numeric(Ad(stock))
  Data[,i]<-prices
 }

 #Pick stocks
 nstock<-dim(Data)[2]
 cor<-cor(Data[1:len,])
 cor[upper.tri(cor)]<-NA
 diag(cor)<-NA
 cor<-as.matrix(cor)
 toppairs<-matrix(0,0,nrow=npairs,ncol=2)
 for(i in 1:npairs)
 {
  a<-sort(cor, decreasing=T)[1:npairs]
  b<-which(cor==a[i],arr.ind=T)
  toppairs[i,1]<-rownames(cor)[b[1]]
  toppairs[i,2]<-col<-colnames(cor)[b[2]]
 }
 return(toppairs)
}

#SAMPLE CALL
#topcors<-selectpairsCORR()


#Cointegration Method

 from = "2011-01-01"
 to = "2012-01-01"
 npairs<-10 #number of top pairs to trade
 cutoff<-2

 #Read file and prepare matrix with price series
 stocks<-read.csv(file.choose(), header=T)
 tickers<-as.character(stocks[,2])
 n<-length(tickers)
 temp.stk<-getSymbols(tickers[1], auto.assign = F, from = from, to = to)
 temp.prices<-as.numeric(Ad(temp.stk))
 len<-length(temp.prices)
 Data = matrix(0,0,nrow=len,ncol=n)
 colnames(Data) <- c(tickers)

 #Pick stocks
 coint<-cor(Data)
 coint[upper.tri(coint)]<-NA
 diag(coint)<-NA
 for(i in 1:n)
 {
  for(j in 1:n)
  {
    if(i > j)
    {
      p1=rownames(coint)[i]
      p2=colnames(coint)[j]

      y = getSymbols(p1, auto.assign = F, from = from, to = to)
         x = getSymbols(p2, auto.assign = F, from = from, to = to)
         y.ad = as.numeric(Ad(y))
         x.ad = as.numeric(Ad(x))
         corr = cor(y.ad,x.ad)
         fit = lm(y.ad ~ -1+x.ad)
         beta = coef(fit)[1]
         spread = as.numeric(y.ad-beta*x.ad)
         adf = adf.test(spread, alternative="stationary", k=0)
      coint[i,j]=adf$p.value
```

```
    }
  }
 }

 #Makes a list of the top 10 by lowest p-value of co-integration Dickey-  #Fuller Test;
 Flag mechanism is set up to correct for the case of duplicate   #p-values
 coint<-as.matrix(coint)
 toppairs<-matrix(0,0,nrow=npairs,ncol=2)
 flag<-0
 for(i in 1:npairs)
 {
  if(flag != 0)
  {
    for(j in 1:flag)
    {
      a<-sort(coint, decreasing=F)[1:npairs]
      b<-which(coint==a[i],arr.ind=T)
      index1<-b[i,1]
      index2<-b[i,2]
      toppairs[i,1]<-rownames(coint)[index1]
      toppairs[i,2]<-colnames(coint)[index2]
      flag = flag - 1
    }

  }
  else
  {
    a<-sort(coint, decreasing=F)[1:npairs]
    b<-which(coint==a[i],arr.ind=T)
    test<-dim(b)[1]
    if(test>1)
    {
      flag<-test-1
    }

  index1<-b[1,1]
  index2<-b[1,2]
  toppairs[i,1]<-rownames(coint)[index1]
  toppairs[i,2]<-colnames(coint)[index2]
  }
 }

## Access toppairs or coint matrices

##Plotting Coint; adapted from Robin Gong, Stat 107, Section 7 Materials
##
plotcoint<-function(ticker1,ticker2,from = "2011-01-01",to = "2012-01-01")
{
 #Get stock info and prepare for output
 y = getSymbols(ticker1, auto.assign = F, from = from, to = to)
  x = getSymbols(ticker2, auto.assign = F, from = from, to = to)
  Y = as.numeric(Ad(y))
  X = as.numeric(Ad(x))
  n.days = 180

  #cointegration
 p.vals <- 1:(length(Y)-179)
 for(i in n.days:length(Y))
 {
    p1 <- Y[(i-179):i]
    p2 <- X[(i-179):i]
    fit<- lm(p1~-1+p2)
    beta <- fit$coeff[1]
    spread <- as.numeric(p1-beta*p2)
    test <- adf.test(spread, alternative="stationary",k=0)
```

```
    p.vals[i-179] <- 1-test$p.value
 }
plot(EMA(p.vals), type="l", lwd = 3, main = paste("180 Day Cointegration,
 ",ticker1,"and",ticker2), xlab = "Days", ylab = "EMA P-Values", col = "red")
}

#plotcoint("HD","JNJ", from="2011-01-01", to="2013-01-01")
## plot from 2011 to 2013 to show how we determined and then how it traded after- do
 this for the two we are doing in depth


s1="adsk"
s2="adbe"
p1=Ad(getSymbols(s1,auto.assign=FALSE, from="2012-01-01", to="2013-01-01"))
p2=Ad(getSymbols(s2,auto.assign=FALSE, from="2012-01-01", to="2013-01-01"))
fit=lm(p1~-1+p2)
beta=coef(fit)[1]
beta
sprd=p1-beta*p2
sprd=as.numeric(sprd)
adf.test(sprd,alternative="stationary",k=0)
cor(p1,p2)
#repeat this over and over again
s1="wcrx"
s2="life"
s1="qgen"
s2="life"
s1="ilmn"
s2="bmc"
s1="pcar"
s2="life"
s1="lrcx"
s2="amat"
s1="mu"
s2="amat"
s1="wcrx"
s2="qgen"
s1="nihd"
s2="ilmn"
s1="pcar"
s2="expd"
s1="dltr"
s2="atvi"
s1="ebay"
s2="atvi"
s1="mat"
s2="qcom"
s1="celg"
s2="chrw"
s1="mat"
s2="gild"
s1="dell"
s2="chkp"
s1="nvda"
s2="csco"
s1="infy"
s2="fisv"
s1="ca"
s2="isrg"
s1="symc"
s2="chkp"
```