Drew University

College of Liberal Arts

Twitter Sentiment Signal Implications on the Russell 1000 Universe

A Thesis in Economics & Mathematics

By

Richa Patel

Submitted in Partial Fulfillment

Of the Requirements

For the Degree of

Bachelor of Arts

With Specialized Honors in Economics & Mathematics

May 2017

To Swamiji,

for continuously inspiring me to believe.

To my committee, Dr. Giandomenico Sarolli, Dr. Jon Kettenring, and Joseph Noto,

for their generous time and mentorship.

* * *

Thank you.

Table of Contents

Foreword

Stock returns have been studied by economists and mathematicians for several decades in an effort to better understand the driving forces behind them. Many researchers have attempted to predict stock returns using various forms of quantitative analysis and past and concurrent information. One very popular model, the Capital Asset Pricing Model (CAPM), says that all securities are driven by two factors: systematic and idiosyncratic risk (Treynor 1961, Sharpe 1964, and Mossin 1966). Given these two factors, the expected return of any one stock should be a function of its relationship with the stock market, otherwise known as beta, and microeconomic factors that impact the firm in particular, otherwise known as alpha. As this theory was studied further, another popular philosophy emerged. Arbitrage Pricing Theory states that while the largest drive in a stock's return could be its relationship with the market, the expected return of a security should also be a function of a variety of macroeconomic factors (Ross 1976). This led researchers to study multiple other factors that could also have strong explanatory power across all securities. Some of these factors include dividends, momentum, market capitalization, positioning changes, and central bank announcements.

It was not until the 1990's that these factor studies started to become more popular among the quantitative finance and wall street community. There suddenly rose a need to fairly compare the impact of multiple different factors, as some did not have a simple linear relationship to returns. Because of this, the Financial Analysts Journal published a paper describing how to be intellectually honest when backtesting a factor using the signal as an input in portfolio construction (Khan 1990). By approaching the problem via portfolio construction with an information coefficient rather than a model with a simple prediction value output, Khan was able to compare the factors in a way that was attractive to practitioners. Over time, this backtesting framework of mimicking the signal being researched in a portfolio became a staple in the way many financial institutions conduct their systematic trading business (PNC, MSCI, Deutsche Bank, Quantopian). Most recently, Deutsche Bank hosted a quantitative trading conference in which they discuss potential mistakes quants can make when researching a factor (QWAFAFEW Presentation 2015). The backbone of this conference, again, uses the same framework set forth in the 1990s as their primary methodology.

This thesis aims to test the theories presented in the works referenced above by using the same methodologies laid out by Khan in 1990 and apply it to the context of Twitter sentiment. In particular, this thesis will evaluate the effectiveness of using Twitter sentiment as a factor to inform portfolio constructing decisions. This approach was used instead of a standard linear regression because it is widely used in finance today to look for a subtle signal in very noisy data. If a standard linear regression was used, then the model may fit the noise rather than the true relationship between the two variables, or it may become excessively complicated and not as easily interpretable. The backtesting methodology is not only widely accepted and used in finance today, but is also well documented and described in literature. Moreover, due to the excessive noise in Twitter data, the nature of this thesis for the most part is exploratory and tests only market neutral trading strategies constructed by stocks from the Russell 1000 universe.

My contribution to the field is four-fold. Fist, no academic literature has been published that looks at Twitter sentiment for 1000 stocks over a six-year time horizon. Most of the studies that look into using Twitter as an indicator only evaluate at most 30 stocks and for at most a fifteen-month time period. I do not believe that sample size and length is enough to make a broader conclusion about the relation, and will discuss why later in this thesis. Second, no one has looked at Twitter sentiment to inform their pairs trading decisions; one of the main focuses of this study is exactly that. Third, many hedge funds and investment banks spend millions of dollars purchasing this type of data and analyzing it. By researching this topic, I can determine whether this is something worthwhile for them to look at and also make recommendations as to how they can allocate their resources for Twitter sentiment data. Finally, in order to conduct data analysis for this, I needed to write my own and modify old python and R code. Part of this code will be released in the appendix, and all of it will be made open source on Github.

In my results, I find that the overall strength of the sentiment signal in predicting the direction of returns of the stocks in the Russell 1000 is weak and decreasing over time. Though some of my results suggest that the strength is likely non-zero. I find that under certain restrictions without transaction costs, using Twitter sentiment to inform stock decisions does have some statistical significance at the 5% level. Incorporating a sector-neutral portfolio helps to further improve results for the signal, but incorporating a beta-neutral portfolio in addition to the sector-neutral portfolio does not improve the results dramatically. I also find that across all portfolios, there were large differences in the individual sector performance, but the best performing year remained unchanged at 2013.

The first two chapters of this thesis focus on economic and market theory. They provide background knowledge on Twitter, sentiment signals, pairs trading and the datasets used throughout the study. Chapters 3 and 4 showcase the methodology and performance of the different portfolios constructed and tested using the Khan framework. The final chapter aims to provide further analysis of the strategies as well as make recommendations as to how financial institutions can effectively use the information discovered in this thesis.

Chapter 1: Economic/Market Knowledge

What is Twitter & Why is it Interesting?

In the last decade, social media's integration into everyday life has helped it to increasingly become the preferred mode of communication, above phone calls, hand-written letters, and even electronic mail. Websites like Facebook, Instagram, Twitter, and Snapchat are used to converse with friends and family, advertise and promote businesses, and even read and receive news updates. One of these social media sources, Twitter, is especially interesting because of its diverse set of users who Tweet about the stock market ("Investor Relations Study: 1 in 4 of Institutional Investors Use Social Media for Research", 2016).

Twitter was founded on March 21st, 2006 by Jack Dorsey. It is a social networking website and mobile application that allows its users to send and read 140 character long messages called "tweets". Each tweet can be favorited, retweeted, and replied to and every user can follow and be followed by other users to receive updates on their new tweets. All tweets can also be categorized by hash-tags and cash-tags. A hash tag looks like "#summer" and is used to differentiate tweets about a particular topic or subject. A cash-tag looks like "$AAPL" and is used to differentiate tweets about particular stocks. Some examples of these being used in practice are shown in **Appendix A**.

The Twitter platform grew rapidly between 2007-2014, with 200 million users sending over 400 million tweets daily by early 2013 ("Twitter Turns Six", 2012). In April of 2013, the Guardian found that news was released faster on

Twitter than on other news outlets. In fact, when the Boston Marathon bombing occurred in 2013, CNN reported the incident a whopping 15 minutes after the first tweet published about it on Twitter ("Twitter Is Becoming the First and Quickest Source of Investment News", 2013). This ability to get news faster on Twitter attracted the attention of several Wall Street traders, and in an effort to gain an edge in receiving information quicker than their competitors, they started to use the platform themselves to help better inform their trades.

Presently, Twitter is used by a variety of different people. The list includes teenagers, college students, adults with full time jobs, celebrities like Jimmy Fallon, politicians like Donald Trump, organizations like UNICEF, NASA, and Macy's, blogs and newspapers, Wall Street traders, and big banks like Goldman Sachs and JP Morgan. Although each user has a different motivation for tweeting, such as promotion, education, influence, and connecting with family, the consumers of Twitter are typically using the platform as a medium of self-expression.

Origin of Sentiment Signals

Sentiment describes a group of people's opinions and emotions about a particular event or situation. Investor sentiment, which is the aggregate opinion or emotion of participants in the stock market, is often used to gauge how investors feel about the market as a whole. Understanding the behavior and mindset of investors is an important factor in forecasting stock price behavior because people tend to follow a herd mentality and mimic what people around them are doing (Dang 2016). This type of behavior can lead to large price movements in the stock market, and consequently, large profits and losses assumed by both retail and institutional investors. Some of the first researchers to try to quantify the impact of investor sentiment on the stock market were Nicholas Barberis, Andrei Shleifer, and Robert Vishny who published their research in the Journal of Financial Economics. In their paper, they use a Markov process to determine how investor sentiment will impact stock prices after positive and negative earnings announcements. At the end of their study, which was cross validated using a random walk model to project price movements, they found that investors generally underreact to positive earnings and overreact to negative earnings (Barberis et. al. 1997). In another study, Baker and Wurgler find that behavioral patterns of retail investors have a significant impact on market returns, and that smaller, unprofitable, and non-dividend paying growth companies are influenced by investor sentiment (Baker and Wurgler 2007).

Since the original study by Barberis et. al., several other techniques have been introduced to help measure the impact of investor sentiment; an extensive list of the various techniques is provided below in Table 1. While each technique is important, this paper will only focus on investor sentiment derived from Twitter. In practice, once these sentiment scores are calculated, they are then used to indicate or signal stock price movement. For example, a bullish sentiment signal on Apple stock would indicate that investors are expecting the stock price of Apple to rise, so buying shares of the stock may be a good move for a portfolio. In addition to the conventional way these signals are used, often times investors use these signals as a contrarian indicator and instead of buying Apple stock when the signal is bullish, they will sell short.

Table 1: Methods of Measuring Investor Sentiment

Approach	Description	Sources
Financial Market Based Measures	Use any of the following: trade volume, VIX, Price/Earnings Ratio, Price Momentum, Realized Volatility, High Yield Bond Returns, Daily Mutual Fund In/Out Flow, Dividend Premium, and Retail Investor Trade Data.	Gervais (2001), Baker & Wurgler (2004), Kumar & Lee (2006), Baker & Wurgler (2007), Barber & Odean (2008), Hou (2009), Vieira (2011), Da (2015)
Survey Based Sentiment Indices	University of Michigan Consumer Sentiment Index, Gallup Index of Investor Optimism, and Conference Board Consumer Confidence Index.	Brown & Cliff (2005), Da (2015)
Text Mining via News and Social Media	Use natural language processing on newspapers, blogs, and social media websites to determine investor mood.	Tetlock (2007), Barber & Odean (2008), Dougal (2012), Ahern & Sosyura (2015), Zhang (2011), Bollen (2011), Nasseri (2014)
Internet Search Behavior of Households	Keep track of how many people are searching similar terms on search engines through apps like Google Trends.	Simon (1955), Bordino (2012), Preis (2013), Da (2015), Curme (2014), Dimpfl & Jank (2015)
Non-economic Factors	Factors such as health, lunar phases, and season/weather impact an investor's risk aversion and trading behavior.	Kamstra et al. (2003), Yuan (2006), Edmans (2007), Kaplanski & Levy (2010), Akhtari (2011)

What is Twitter Sentiment?

Twitter sentiment can be thought of as how the users of Twitter feel about a particular situation. In a financial context, investor sentiment from Twitter can be extracted by using the text mining approach mentioned in Table 1 as well as by using cash-tags to differentiate tweets about stocks. The method that was used for the purposes of this paper will be discussed in chapter 2. Analyzing Twitter sentiment can give valuable insight on how prices of certain stocks will react when, for example, Donald Trump tweets about Boeing ("Boeing Speaks in Trump Terms on Iran Deal: It's About Jobs", 2016). Moreover, many quant funds and investment banks invest millions of dollars in purchasing sentiment data. The findings of this paper could potentially help inform these financial institutions on whether or not this data is worth investing in.

In order to honestly research a topic, it is important to first try to understand and learn from the work that has already been done related to it. Though, since studying social media and its effect on financial markets is still a relatively new topic in quantitative finance, only a handful of peer-reviewed works have been published related to it. One study performed by a team at the IMT Institute for Advanced Studies finds that there is little correlation between the Dow Jones index and Twitter sentiment over a 15-month period, but there exists a significant relationship between sentiment and returns during periods of large tweet volume (Ranaco et al., 2015). Another study found that not only can Twitter sentiment predict stock price movement of Dow Jones stocks for one day, but also it can predict price movement for three days after (Bollen et al., 2010). On the other hand, Nuno Oliveira of University of Minho in Portugal found that using Twitter sentiment in a linear regression to predict the returns of six large-cap stocks in the S&P 500 is not statistically significant, possibly due to the complex nature of the interaction between stock prices in the S&P 500 (Oliveria, 2013). One recent study looked at the relationship between the number of tweets a particular company tweeted daily and its stock's share price—the results of which were not statistically significant (Rennie 2016). While the studies are conflicting, I hypothesize that the samples sizes and techniques used in each study were not the best for determining the true relationship. Due to the noisy nature of both Twitter sentiment data and stock price data, it is extremely difficult to determine whether the price of a stock increased because someone tweeted about the stock, or because the entire technology sector happened to be performing well that day. Therefore, filtering the tweets and being careful about the universe of stocks being researched is imperative. More importantly, choosing the day or time period when evaluating the sentiment signal from Twitter is key because the signal may be stronger on days when earnings are reported, but weaker when there is little news about a particular stock or company.

Thus, in an effort to take into consideration the potential noise in the data, only market neutral strategies will be researched. In particular, this thesis will test out a simple long/short strategy and a statistical pairs trading strategy using daily Twitter sentiment data. In order to assist the reader, the following section describes the most common example of a market neutral strategy and why it helps in minimizing the impact of day-to-day noise in the market.

Pairs Trading Origin & Practice

Simply put, all investors can be long or short a security to express a directional view of where they think a stock's price is headed. A long position in a particular stock indicates that the stockholder expects the stock's price to increase, while a short position indicates that the stockholder expects the stock's price to decrease. If a person only had a long and short position in the same stock for the same dollar notional amount, then the stockholder is considered neutral in that stock because an increase in the stock's price will be canceled out by the short position, and any decrease in the stock's price should be canceled out by the long position.

Pairs trading, which consists of simultaneously taking a long position in one security and short position in another security in a predetermined ratio, is a type of market neutral strategy (Vidyamurthy, 2004, 8). The ratio is typically determined by the securities' betas, a stock's correlation coefficient of its price to a market index like the S&P 500, over some specified period of time. Pairs trading was developed in the 1980s by a Morgan Stanley team of mathematicians, physicists, and computer scientists. The motivation behind this was to take advantage of any mispricings in financial markets. Some famous members of the team included Gerald Bamberger, who taught Law at University of Buffalo after working in finance, Nunzio Tartaglia, and David Shaw, who later went on to start his own hedge fund called D.E. Shaw.

While there are two types of pairs trading strategies that are commonly used in finance today, statistical arbitrage and risk arbitrage. For the purposes of this thesis, only statistical arbitrage will be studied.

At the core of any long or short trade, the theme is to sell a security that is overvalued, and to buy a security that is undervalued. That way, when the price of the overvalued security drops or the price of the undervalued security rises, the trader can make a profit. However, the market is noisy and it is near impossible to figure out the *true* value of the security at any given point unless the trader has perfect information about the precise value of the company. A trader may have a general idea of where the company should be trading, but he/she cannot be certain of the true price. Statistical arbitrage eliminates this ambiguity entirely.

If two companies have extremely similar risk characteristics, then they should be priced approximately the same. The particular price of one security is not important, said security can be priced wrong as well. However, the security's price relative to another security is what matters. For example, consider the companies Coca-Cola and PepsiCo, both of which are traded as two separate securities. For the sake of simplicity, ignore, for the moment, the dissimilarities of the two companies. Both companies are in the Consumer Goods industry, and both companies are known for producing soft drinks. On a fundamental level, both companies have similar risk characteristics. Due to their numerous similarities, the price of Coke stock should, in theory, be similar to that of Pepsi stock. The true monetary value of both Coke and Pepsi can be higher or lower than what the two securities are trading at, but as long as their relative prices stay the same, traders who use statistical arbitrage strategies could not care less. Another example in which two companies' prices can be correlated is when one company produces a good that takes what another company produces as an input. For instance, a large component of a tire manufacturing company is oil, since it takes roughly seven gallons of oil to produce one standard rubber tire ("Rubber Faqs"). As a result, it may be more costly to produce tires if the price of oil rises, so the stock prices of an oil manufacturing company and a tire manufacturing company may be highly correlated as well.

A difference in the relative price of the two securities can be caused by one of the securities being over/under priced or a combination of the securities being over/under priced; the greater the difference in the relative prices, the higher the potential for greater profits. A trading strategy to take advantage of this mispricing is simple: sell the higher-priced security and buy the lower-priced security (Vidyamurthy, 2004, 75). The idea is that the difference in relative pricing will correct itself over time and as that correction happens, the trader will profit from the long-short position. Because the strategy is market neutral as discussed above, the exposure to market risk is limited, and the profit from the trade should be uncorrelated with market returns. The way the pairs are chosen is often through a method created by Evan Gatev, who looks at the historical correlation between two securities (Gatev, 1999). One way to do this is to compare the returns of the two securities over time. Arbitrage pricing theory states that the expected return of two securities is the same if these two securities have the same risk profiles (Huberman, 2005). Using this theory, traders who employ a statistical arbitrage strategy hope that the difference in returns (and consequently the relative prices of the two securities), or spread, approaches a long-term equilibrium. At different points in time, however, the spread will fluctuate around the equilibrium spread. For example, the long-term equilibrium spread between APPL-MSFT return could be 0.025 with a standard deviation of 0.005. On a given day that spread could be wider at 0.10, and traders would be able to act on the “unusual” spread by buying APPL and selling MSFT.

Pairs trading may seem to be a safe and profitable strategy that any trader on Wall Street would be foolish to not employ on his/her own books. However, it is important to remember that correlation does not imply causation, and that the spread between two securities may never return to their long-term equilibrium. To further illustrate the drawbacks of pairs trading for the interested reader, a case study on Long Term Capital Management is provided in **Appendix B**.

All in all, pairs trading is a really interesting strategy to look at, whether it be for short-term weekly trades, or for long-term one year trades. Under the proper conditions and guidance, these trades can be extremely successful in flushing out some of the noise in financial markets.

Chapter 2: Backtest Construction

Data Sets

Several datasets were used to research the various trading strategies presented in this thesis. Daily stock price data for the members of the Russell 1000 was taken from Quandl's EOD database as well as Bloomberg. Daily price data for select market indices was taken from Yahoo Finance. Twitter sentiment signals were taken from the Psych Signal database. A detailed description of how the data is collected and adjusted daily from the different data vendors is provided in **Appendix C** and the following section highlights some important characteristics of the Twitter sentiment data.

Every day at 7pm E.T., subscribers are sent a daily sentiment score ranging from -4 to +4 based on the Twitter activity that took place that day. A zero sentiment score has one of two meanings. The first is that the zero can represent an instance in which both the bullish and bearish scores were zero, i.e. there was no news about that stock on that day. The second way this can happen is when the bullish and bearish scores equal each other. In this instance, the sentiment from Twitter is neutral on the particular stock.

One can imagine that in a scenario where there are a couple extremely positive tweets, and several hundred extremely negative tweets, the bull-bearish score would be zero. In that particular case, as well as the reverse, the zero score would be heavily biased by the two extremely positive tweets. It is important to note, however, that an incident like this is unlikely, and possible only when a

Twitter account that is extremely credible and has a very high significance score, posts a very positive tweet. In that case, it may be both useful and interesting to look further as to why that particular institution or individual was so bullish on a stock when the rest of the Twitter universe was so bearish.

In addition to the bias on the final sentiment signal, there are a variety of factors that implicitly bias the intermediate bullish and bearish scores. First, Psych Signal chooses which Twitter handles are the most important, and therefore, lesser known Twitter handles for up and coming news sources may not receive as much importance, despite the fact that they may have a large Twitter following. In contrast to that, some celebrities may also be deemed as more important, but may not have the most accurate news stories. As a result, markets may overreact to some news stories that the celebrity tweets out, but later realize it is wrong or has nothing to do with the company's stock. While this type of market behavior may be good for day-traders who take advantage of extreme intraday stock price movements, the signal used in this study is an aggregate of the entire day's sentiment. Consequently, in these types of situations, the signal may be closer to neutral for the day rather than a strong buy or sell right after the celebrity's tweet. Finally, Twitter's popularity, as well as the presence of certain stocks on Twitter, has changed over time. Stocks that had just created a Twitter account in 2010 certainly have a greater number of followers and tweeters now. Therefore, a strong sentiment signal of 3.5 in 2010 may not mean the same as a sentiment signal of 3.5 in 2015.

All in all, this paper will assume that despite the biases discussed, the sentiment scores are trustworthy enough to capture how the Twitter universe felt about a particular stock on a given day. Moreover, the data was not filtered out for stocks with a minimum number of tweets related to it. This was done because I believe that no news can be good news for a stock that had a series of very strong negative days. Furthermore, if I eliminated these days from the data, I would be introducing some additional selection bias into the data as well and making the model increasingly complex. Table 2 shows basic summary statistics for the signal data for each year.

Table 2: Summary Statistics for Psych Signal's Twitter Sentiment Signal

Statistic	2010	2011	2012	2013	2014	2015
Count	35,986	133,750	169,384	192,332	204,860	221,478
Mean	0.4074	0.3719	0.3502	0.4745	0.3961	0.5451
St. Dev	1.310	1.221	1.1376	1.2588	1.2812	1.2919
Min	-4.00	-4.00	-4.00	-4.00	-4.00	-4.00
Max	3.89	3.89	3.63	3.89	3.63	3.89
25%	0.00	0.00	0.00	0.00	0.00	-0.11
75%	1.56	1.28	1.35	1.58	1.41	1.80

The large jump in the number of observations from 2010-2011 is largely due to an increase in the popularity of Twitter. As more people began to use Twitter, more companies in the Russell 1000 felt compelled to create their own Twitter accounts for both advertising and competitive purposes. It is interesting to note, however, that despite the small number of observations in 2010, the standard deviation and was comparable to the standard deviation of year 2015. The range of the data, as well as the average signal, also stayed relatively stable over time, indicating that popularity of the Twitter platform may not have a large impact on the stocks in the Russell 1000 in aggregate. However, this says nothing about the impact it has on individual stocks, and the effect of this will be studied in the next chapter.

Universal Backtest Methodology

The following methodology was used for the different strategies tested in this paper. First, to avoid overfitting the strategies' models to the data, an in sample and out of sample partition was created. The in sample dataset ranged from January 1st, 2010-December 31st, 2014, and the out of sample dataset ranged from January 1st, 2015-December 31st, 2015. Not only will this help to avoid data mining by keeping the out of sample data clean for only the best performing strategies, but it also will help to see how robust the strategies are with regard to any changes in the economy and the popularity of Twitter. Moreover, in order to fairly assess and compare each strategy's performance as illustrated by Khan 1990, a mock portfolio of $1,000,000 was used. Once a strategy had been identified as the strongest performer using various metrics, the strategy was run again with varying assumptions for transaction costs.

The mock portfolio was constructed using members of the Russell 1000. This index was chosen for a few different reasons, with the first reason being to simplify the universe of stocks without compromising on the number of potential trades the strategy would have on a given day. Second, the Russell 1000 is comprised of the largest companies by market cap in the United States. Because of this, I assumed that these would be the most well-known publicly traded companies, and consequently the most tweeted about. Third, while the S&P 500 companies are also well known, the S&P indices go through a quarterly rebalancing process whereas the Russell 1000 only reconstitutes its index once a year. Choosing the Russell 1000 over the S&P 500 not only minimizes the selection bias incorporated in the data, but also lowers the severity of survival bias as well. Moreover, the additions and deletions data for the Russell 1000 was provided for free from FTSE Russell, and thus was a better alternative to the costly S&P indices.

Does This Factor Have Merit?

Before any sophisticated trading strategies were crafted to see if Twitter sentiment has some impact on stock return, I was curious to see what the cumulative profits and losses (pnl) would look like if the signals from Psych Signal were followed just as is, without filtering the data for any particular days. A mock portfolio was created. The algorithm would buy all of the stocks with a sentiment signal greater than zero and sell short all of the stocks with a sentiment signal less than 0. The portfolio would rebalance daily by getting out of all open positions at the close. This occurred despite the fact that many stocks may have a signal greater or less than zero for multiple consecutive days. The primary reason for this is simply because news does not affect the market for a prolonged period of time; i.e. news that was relevant a week ago is not a major driving force behind market performance today (Smales 2012). Figure 1 below shows the performance of the algorithm on the in sample data.

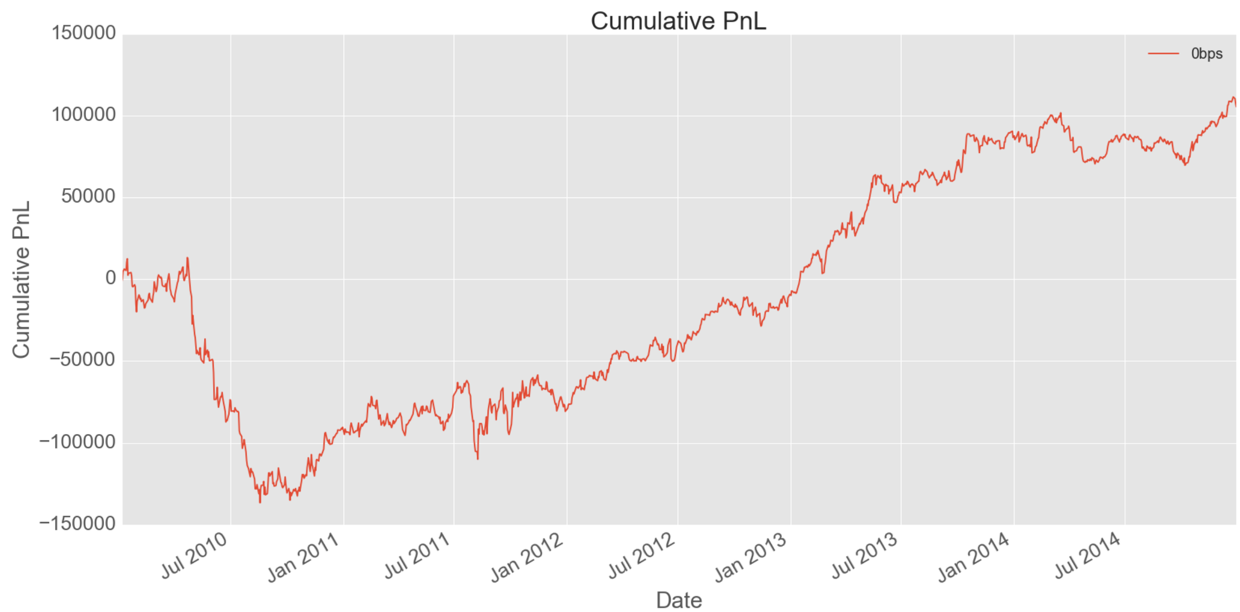


Figure 1: Cumulative PnL With No Modifications to the Data

It is clear that overall, this strategy would make a trading desk some money over the course of 4 years, despite the considerable losses in the beginning of its implementation. The strategy made \$83.76/day on average with a standard deviation of \$3075.74, and had an overall annualized Sharpe ratio of 0.43. The annualized Sharpe ratio for the Russell 1000 for the same time period was 0.69. This metric is a measure of risk-adjusted returns for a portfolio and is calculated using the following equation:

Sharpe Ratio = (Mean portfolio return − Risk-free rate) / (Standard deviation of portfolio return)

It informs an investor of how well a strategy is performing relative to the amount of risk that it takes on. In practice, the higher the Sharpe ratio, the more attractive the strategy is to an investor because it means that the overall return of the portfolio is high while the volatility of the return is low. In other words, the strategy has high reward and low risk. A negative Sharpe ratio is often seen as a sure way to lose money, as the average return of the portfolio, after being adjusted for the risk-free rate, is negative.

Our Sharpe ratio of 0.43 is considered relatively weak, because there is low return and high risk with the strategy, as evident by the small average daily pnl and very high standard deviation. The main reason for this may be that there is a lot of noise inherent within the data and not every signal from Twitter can be treated as the best and most accurate for determining the behavior of stocks in the Russell 1000. For example, some stocks may have a strong buy signal, but the market as a whole could be underperforming that day. As a result, buying that particular stock, while it may be fruitful in the long run, may not have profitable results in the short term. In fact, in that scenario, it may be the case that the stock actually underperforms with the market that day.

Despite these factors impacting the Sharpe ratio, it seems as though there are certain time periods in which the strategy makes positive profits. It may be the case that certain sectors perform better under the Twitter sentiment signal than others, or that the signal has the most predictive power when it is the furthest away from zero. As a result, the following two chapters will attempt to identify the profitable time periods for this factor with exploratory data analysis, though it is important to note that my results are limited to the length of the time series. Market conditions can change and it does not necessarily need to hold that the strategies that perform well in this time period will continue to perform well in the future.

Chapter 3: Simple Long/Short Strategy

The purpose of this strategy is to help filter out some of the noise in the data discussed in chapter 2. Since a signal is calculated every day for all of the stocks that have at least one Tweet about them, it is important to try to differentiate which signals are strong enough to be worth pursuing. For example, if Apple had a signal of 3.5, which indicates a strong buy, and Microsoft had a signal of 0.4, it may be more meaningful to only trade Apple because the signal for Microsoft is relatively weak and trading it may not give you a decent enough return. This chapter introduces two methodologies to explore the data through and help select stocks for our portfolio.

Methodology

For this strategy, the stocks in the Russell 1000 were first ranked each day based on the strength of their signal, as in the Fama and French model (Fama and French 1992). The ranks were such that a rank of 1 would indicate that the stock had the highest buy signal of all the signals that day and a rank of 1000 would indicate that the stock had the highest sell signal. After this, each rank's percentile was calculated by taking the rank and dividing it by the total number of stocks ranked that day. A fragment of the data table is provided below to help visualize this (Table 3).

Table 3: Sample Dataset With Rank Percentiles

Date	Ticker	Sector	Signal	Return	Rank	Rank %
01/01/2010	AAPL	Information Tech	-1.43	-0.10%	918	0.918
01/01/2010	MSFT	Information Tech	-0.56	1.04%	520	0.520
01/02/2010	MSFT	Information Tech	-1.89	-0.027%	802	.802
01/02/2010	XOM	Energy	1.67	0.16%	27	.027

The strategy was simple once the percentiles were found: buy the stocks that were less than some percentile bucket and sell short the stocks that were greater than some percentile bucket. Various percentile buckets were used throughout the research process to test the robustness of the portfolio's returns. Once the buys and sells were determined, the profits and losses of those stocks were calculated using the daily returns and summed. Each trade was put on with equal amounts of the total notional value[1], so if the strategy identified 500 trade opportunities in one day, each stock would be allocated $1mm/500 = $2,000. It is important to note here that

[1] The same strategy was run with a beta neutral portfolio as well. This was done by allocating a proportionate notional amount to the long and the short quantiles such that their weighted betas were equal. The results were not included in this section because they were similar to the equal notional portfolio. Although the beta neutral portfolio does not disproportionately assign higher bets to more volatile stocks, it does add another layer of complexity to the portfolio. Ultimately, I felt that choosing a simpler model was better because every portfolio manager has different rules on how to weigh his/her portfolio. The equal notional portfolio is easier to interpret and modify than the beta neutral portfolio, which is often more valuable when dealing with complex and computationally intensive modelling problems.

since the signal is received at 7pm EST after markets close, the return was calculated with the assumption that the trade would be put on at open the next trade day and taken off at close that same day.

The second methodology tested aimed to ensure each sector within the Russell 1000 was represented in the portfolio. To make this calculation, the stocks were tagged with one of the following 11 sector names: Consumer Discretionary, Information Technology, Energy, Health Care, Financials, Industrials, Telecommunications, Consumer Staples, Materials, Utilities, and Real Estate. After this, the dataset was split into 11 subsets corresponding to these sectors. Each stock in the subset was then ranked daily based on the strength of the sentiment signal. Finally, the percentiles of each stock in each sector were calculated. This helped to guarantee that each sector would be included and that the strategy did not overweight a particular sector, given the strength of the signals in aggregate. For example, at a 10% cutoff point, the strategy would take the top 10% and bottom 10% signals in each sector. Like the first methodology, this was also tested for various cutoff points.

Performance

Table 4 and Table 5 shown below summarize the results of each methodology's cutoff performance in sample, which ranged from January 1st, 2010 through December 31st, 2014. The number of trades represents the total number of individual stocks that the strategy had traded in its lifespan. The maximum and the minimum profit and losses are also reported to showcase the most amount of money the strategy made or lost during the 2010-2014 time period. The percent winners variable represents the percentage of days that the strategy had positive profits. This is to help evaluate the algorithm through another lens as a strategy that makes money 50% of the time may be better than one that makes money only 5% of the time. However, consider the case that a strategy that makes money 50% of the time only makes $10,000/year, while the strategy that makes money 5% of the time makes $1,000,000/year. Then, the strategy that makes money 5% of the time would be the more attractive investment. Because of this, other metrics—namely the average pnl, standard deviation of the pnl, the Sharpe ratio, and the cumulative pnl— were also measured to gauge the strength and resilience of each strategy tested. In order to determine whether or not the returns generated by this portfolio are random, a one sample t-test was conducted to test whether the portfolio's mean profits were statistically significantly higher than zero. This was done because the t-test is one of the most robust inferential tests when there are outliers present in the data. Because the returns of each portfolio have such high standard deviations, the resulting t-statistic that is tested will be smaller than if the outliers were not present. With a lower t-statistic, it is harder for the results to point to statistical significance. Therefore, it was determined that this would be a good approach in testing to see if the returns could have been randomly generated, though there are many other ways this question could have been solved. The resulting p-values of the t-tests are reported below, and a further discussion of the underlying assumptions is provided in **Appendix D**.

Table 4: In Sample Performance Summarized Without Sector Breakdown

No sectors	1%	10%	20%	30%	40%
# of Trades	14,276	148,170	288,121	380,908	587,701
Max PnL	$24,423.97	$17,358.86	$8,746.67	$13,614.92	$9,960.63
Max Loss	$(28,758.70)	$(16,897.31)	$(10,436.43)	$(11,960.97)	$(12,177.58)
% Winners	45.88%	50.00%	50.95%	56.60%	50.32%
Average PnL	$70.78	$(35.35)	$(9.34)	$119.27	$35.07
St. Dev	$5,022.91	$2,286.01	$1,808.89	$2,253.74	$2,127.14
Sharpe Ratio	0.2237	-0.2455	-0.0819	0.8401	0.2617
Cumulative PnL	$83,314.53	$(44,982.92)	$(11,795.09)	$149,377.97	$44,169.20
P-Value	0.621	0.579	0.854	0.062	0.558

Table 5: In Sample Performance Summarized With Sector Breakdown

With sectors	1%	10%	20%	25%	30%
# of Trades	12,251	146,430	275,868	331,754	388,478
Max PnL	$27,161.84	$26,436.97	$12,967.90	$12,198.74	$12,873.67
Max Loss	$(23,161.83)	$(29,250.67)	$(17,703.81)	$(11,180.33)	$(13,136.62)
% Winners	25.99%	50.56%	53.58%	55.01%	56.52%
Average PnL	$(330.02)	$32.61	$111.53	$127.13	$91.56
St. Dev	$6,114.29	$3,350.07	$2,012.90	$1,904.85	$1,915.38
Sharpe Ratio	-0.8568	0.1545	0.8796	1.0595	0.7588
Cumulative PnL	$(236,524.55)	$40,761.71	$139,897.02	$159,275.74	$115,127.53
P-Value	0.152	0.731	0.050	0.019	0.090

Overall, the best performing strategies were at a 30% cutoff for no sectors, and at a 25% cutoff when the data was grouped together by sector. This was evident by the cumulative pnls, the annualized Sharpe ratios, and the p-values. Not only did both strategies have the highest profits in the end and the highest return per unit of risk, both strategies also had the lowest p-value out of the various cutoffs tested. For the 30% cutoff, the p-value, although not statistically significant at a 5% level, was a lot lower than the other p-values. For the 25% sector cutoff, the p-value indicated significance at a 5% level, though this significance should be taken with a grain of salt as discussed in Appendix C. After a 30% cutoff, both strategies stop working. This follows the original hypothesis that including all of the stocks in the strategy would include too much noise in the data, and thus, drive down the portfolio's overall return. On the other hand, the strategy is also not profitable with lower cutoff points (i.e. the most extreme signals). This seems likely because some of the “strongest” signals were from stocks that were the least tweeted about because they have a small Twitter following. As a result, using those signals as an indicator for how they will perform over the next day may not have been the best approach because retail investors may not care to act on a couple positive or negative tweets.

One can argue that some investors will care about the lesser known stocks enough to cause a price movement because when there is a Tweet for them, it has valuable information. Despite this, the data indicates otherwise. At a sector cutoff of 1%, an investor would lose an average of $330/day, with a rather unsteady portfolio throughout the five years given the $6,114 standard deviation. This is because a strategy that only has a handful of trades daily does not have enough data points to make a fair conclusion about the characteristics of the sample population. In other words, the algorithm may have a difficult time discerning between the true relationship and the random noise in the market, and therefore will dampen the average pnl. What is more is that by not trading more names, the potential for a higher standard deviation also increases because there are so few data points, and they may not all follow the same pattern. This is why the Sharpe ratios on the lower cutoffs are smaller compared to those on the higher cutoff strategies.

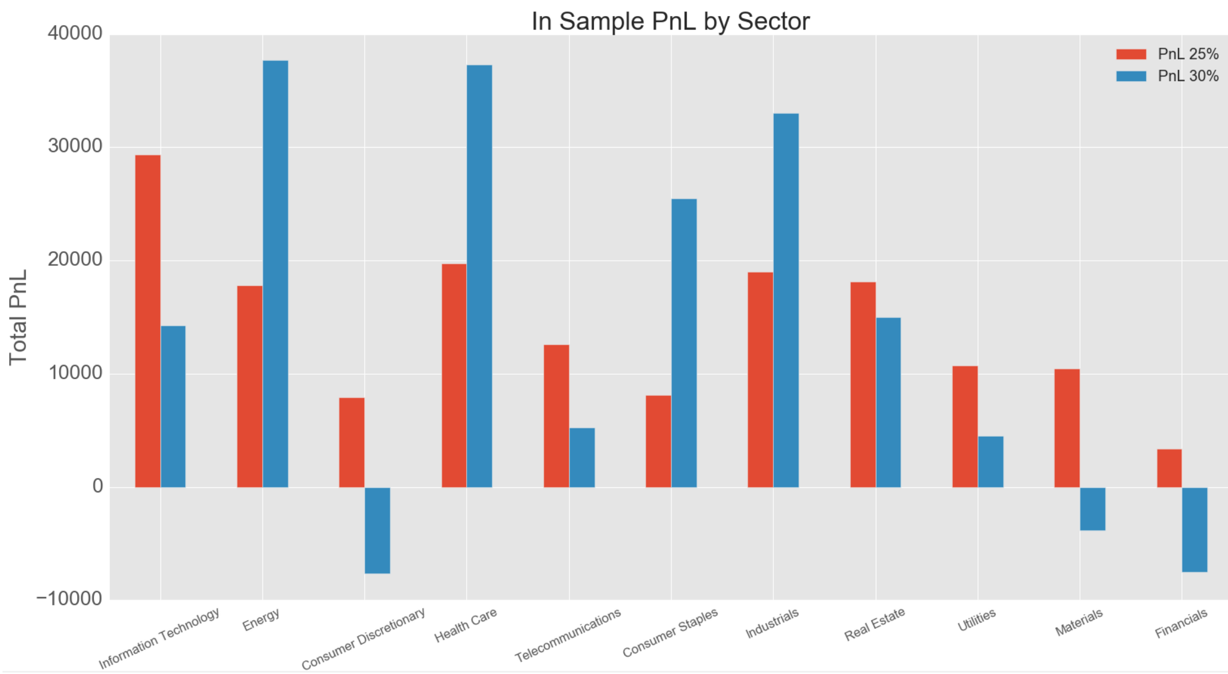
On the flip side, at a higher cutoff point, the strategies are able to capture more profit. This is because the algorithm is able to take advantage of the “true” signals by buying and selling stocks that have a higher presence and following on

Twitter. The law of large numbers, loosely speaking, states that as the sample size increases, the sample mean will approach the population mean. As expected, this contributes to better results in the strategies because with higher cutoff points, more trades are being considered for the overall profit of the portfolio, and the standard deviation of the returns in the five years is lower as well. Both factors lead to higher Sharpe ratios for the strategies.

When the data is grouped by sectors before calculating the percentiles, the Sharpe ratio for the 25% cutoff is the highest at 1.06. Without sector grouping, the Sharpe ratio for the 30% cutoff is the highest at 0.84. While the maximum profits for the 25% cutoff is lower than the 30% cutoff, so are the maximum losses. Conversely, the average profit of $127/day for the 25% cutoff is slightly higher than the $119/day for the 30% cutoff. One possible reason for why the 25% cutoff strategy works better is likely due to the fact that it forced the portfolio to be balanced equally by sectors. By doing this, the algorithm avoided trading some of the extreme signals discussed earlier. In contrast to this, the algorithm may have been forced to trade sectors that had some ever so slightly positive or negative signals. These signals may not have been traded in aggregate, but with fewer signals to compare to when broken up into smaller groups, those signals may have made the cut.

A natural question one might have after looking at the statistics of the strategies in sample is how each sector performed for the two best performing cutoffs. The first graph in Figure 2 below shows a breakdown of each sector's total profits at the end of 2014. Directly under it is a graph showcasing the various

Sharpe ratios for each sector. The red bars represent the 25% cutoff with sector grouping, while the blue bars represent the 30% cutoff with no sector grouping. The sum of all of the sectors' profits represents the total profit of the strategy.



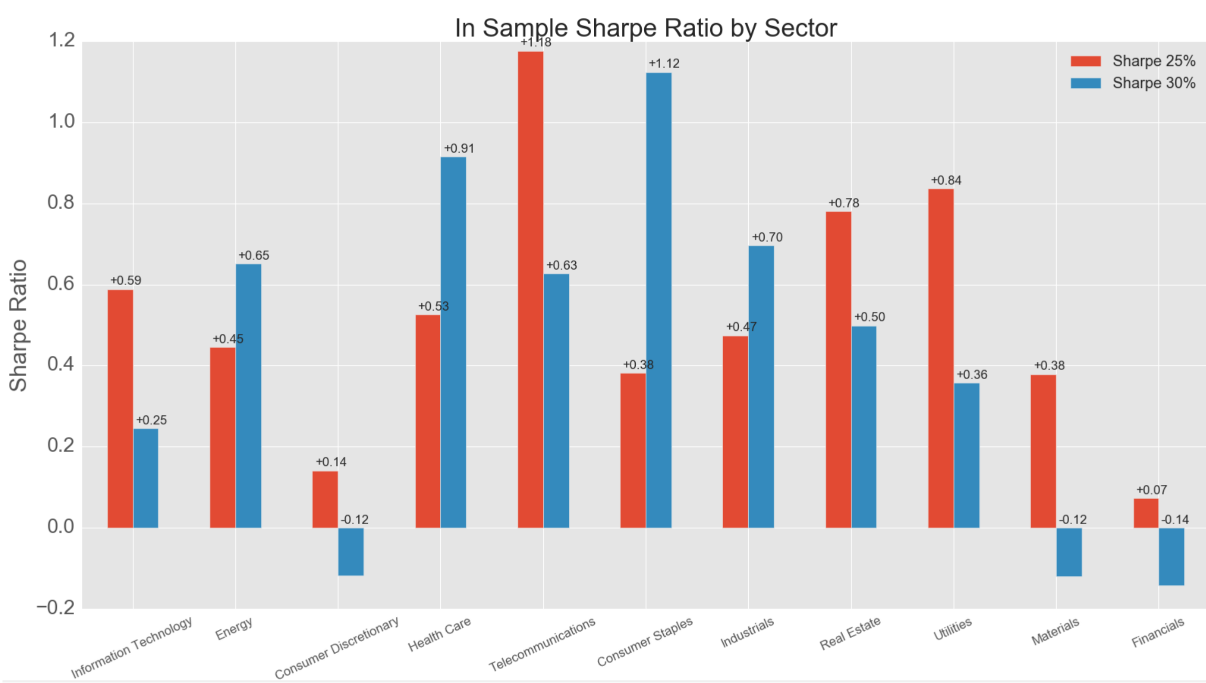


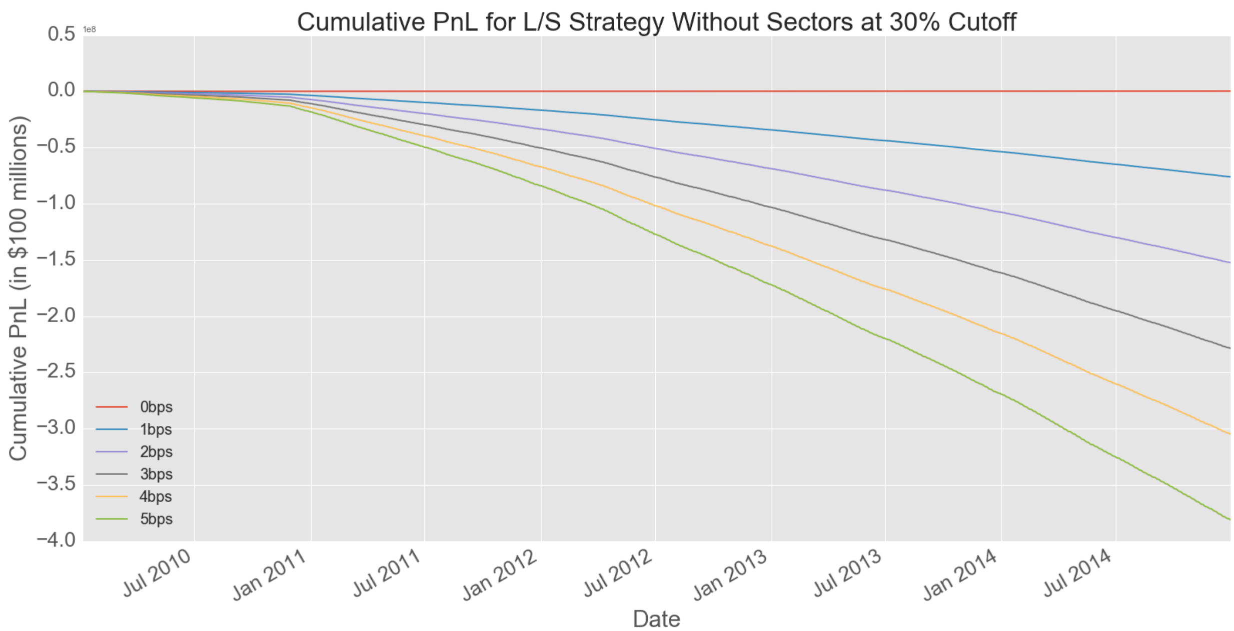
Figure 2: Sector Breakdown Performance

Overall, it seems that some sectors perform better than others. In the 30% cutoff, the most profitable sectors were Energy, Health Care, Industrials, and Consumer Staples. In the 25% cutoff, the most profitable sectors were Information Technology, Health Care, Industrials, and Real Estate. For both cutoffs, Consumer Discretionary and Financials were the least profitable. The distribution of profits for each sector in the 25% cutoff strategy looks more uniform than the distribution of profits for each sector in the 30% strategy. This makes sense because in the 25% cutoff strategy, each sector had an equal weight in the portfolio, so the sectors that were overweight in the 30% cutoff strategy had both their profits and losses minimized. Moreover, it seems that sectors with high Sharpe ratios also have high profits over the five years.

In the 30% cutoff strategy, Consumer Staples had the highest Sharpe ratio of 1.12, but ultimately only contributed to 16% of the total profit from 2010-2014. Since consumer staples includes stocks that produce essential household products such as food and beverages, when there is positive or negative news about these stocks, a large number of people may be inclined to tweet about them. For example, if Proctor & Gamble was found to abuse animals while testing out their beauty products, then several people may not want to purchase their products anymore. Many of those people may also take their anger out on the Twitter platform to encourage others to not support a company that supports animal abuse, which would ultimately strengthen the Twitter sentiment signal. As a result of seeing this consumer dissatisfaction on Twitter, and in anticipation of lower revenues because of the reduced number of buyers of beauty products, the price of Proctor & Gamble stock may decrease. The Consumer Staples sector did not perform as well in the 25% cutoff, showing a Sharpe ratio of 0.38 and contributing to roughly 5% of the total profits. This is likely due to the fact that fewer stocks from consumer staples were traded when each sector was equally weighted in the portfolio. A similar logic can be used to explain the overall profitability and Sharpe ratios of Health Care and Energy stocks. Health Care, whose Sharpe ratio was the second highest for the 30% cutoff strategy, contributed to 23% of the portfolio's total profits. Energy also contributed to 23% of the total profit, but only had a Sharpe ratio of 0.65 in the 30% cutoff strategy. Since most people rely heavily on products produced by both sectors, it makes sense for the sectors to behave similar to the Consumer Staples sector.

On the other hand, the sectors that performed the worst are sectors that many people may still care about, but may not be as directly affected by them. The Consumer Discretionary sector includes stocks that produce products which are not essential to household, but are wanted like a Disney vacation or cable television. Because several people can simply live without these goods and services, the sentiment signal for these stocks may not be as strong as with Consumer Staples simply because many people may not feel passionately enough about them to tweet out their opinion. Moreover, a lot of financial companies took a few years to fully recover from the financial crisis of 2008-2009. In addition to that, many financial institutions, like Goldman Sachs, did not have a Twitter handle until 2012. Thus, any sentiment related to them before then may not have been as strong because without a Twitter account to tweet at, many people may have felt that their voice would not be heard by the actual institution their tweet was meant for.

Naturally, one might now be interested in seeing how these strategies performed with the addition of transaction costs, since in practice, it is impossible to trade without some slippage and transaction costs. Table 4 below showcases the cumulative profits and losses for each of the best performing strategies given various assumptions for transaction costs per one leg of the trade. For example, if the algorithm wanted to trade APPL at 1 basis point, it would cost 1bps to buy APPL and 1bps to sell AAPL.



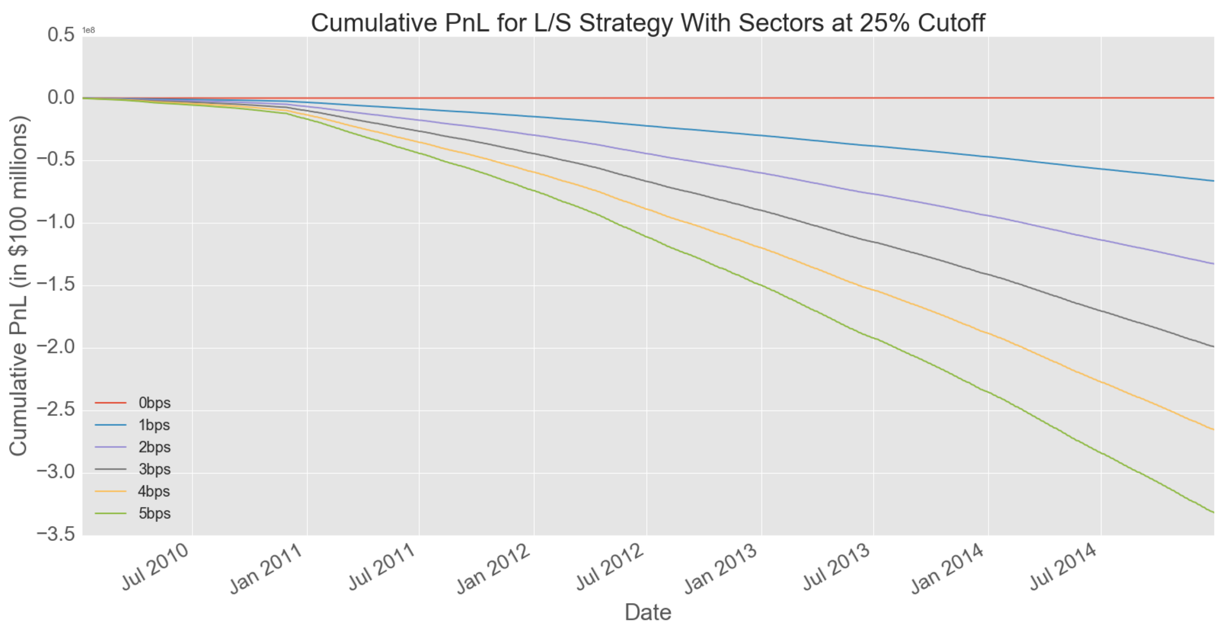


Figure 3: In Sample Cumulative PnL With Transaction Costs

It seems as though both strategies fail to make enough money to offset even 1 basis point of transaction cost. Given that both strategies make an average of $120-130/day and it costs roughly $200 to trade both strategies each day, it would be surprising to see results otherwise. Though this may not seem like an attractive strategy to a trader, in practice, several strategies in quantitative finance are used in combination with other alpha-generating strategies. Thus, I am assuming that when combined with several dozen other strategies, the two that are shown above will have a negligible amount of shared transaction cost. Moreover, because the two strategies were both profitable without transaction costs, I decided to test them out using the out of sample dataset to see if they were also robust to a different time period. Figure 4 below shows the pnl and Sharpe ratios each year for the two strategies compared to the Russell 1000.

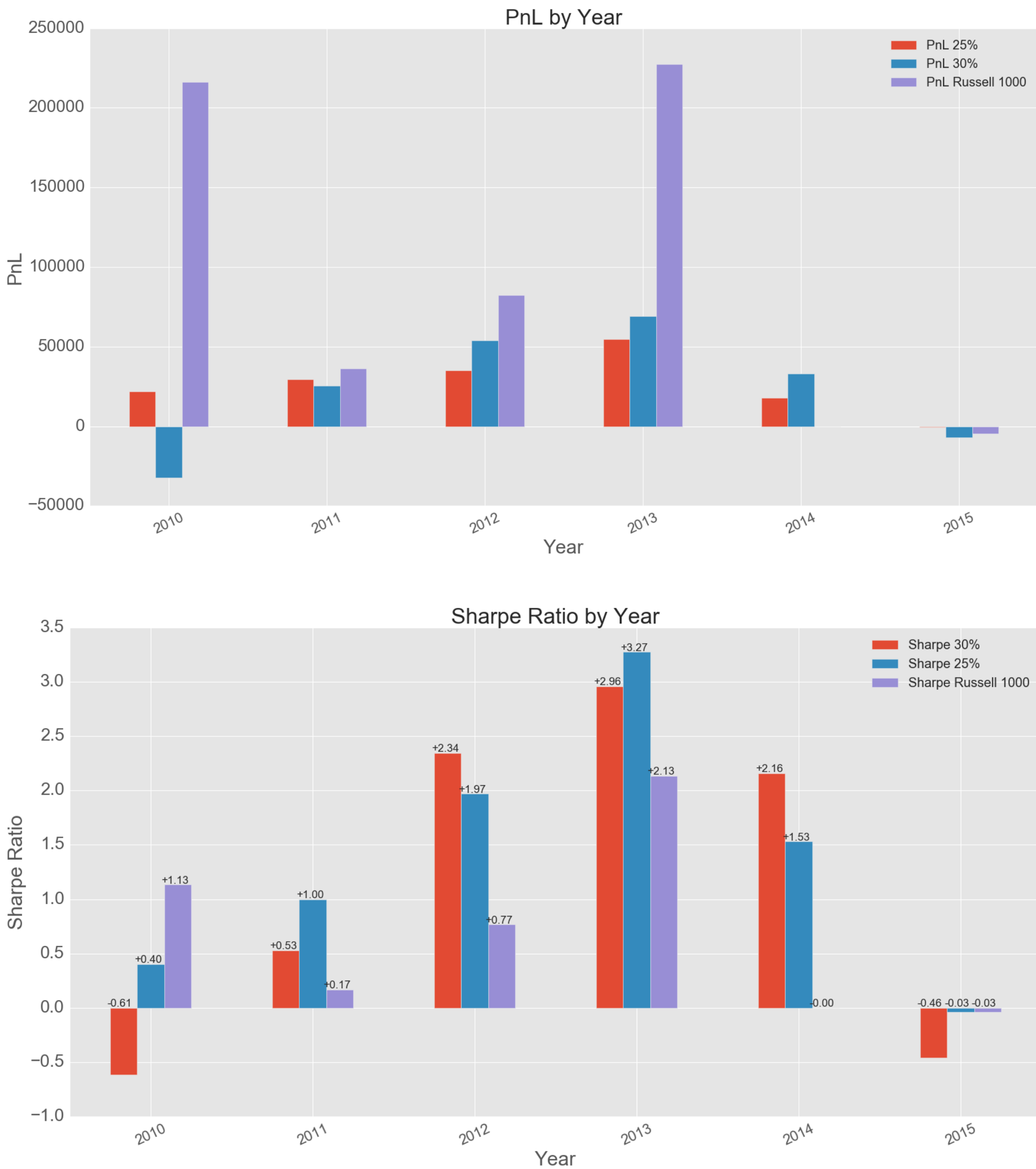
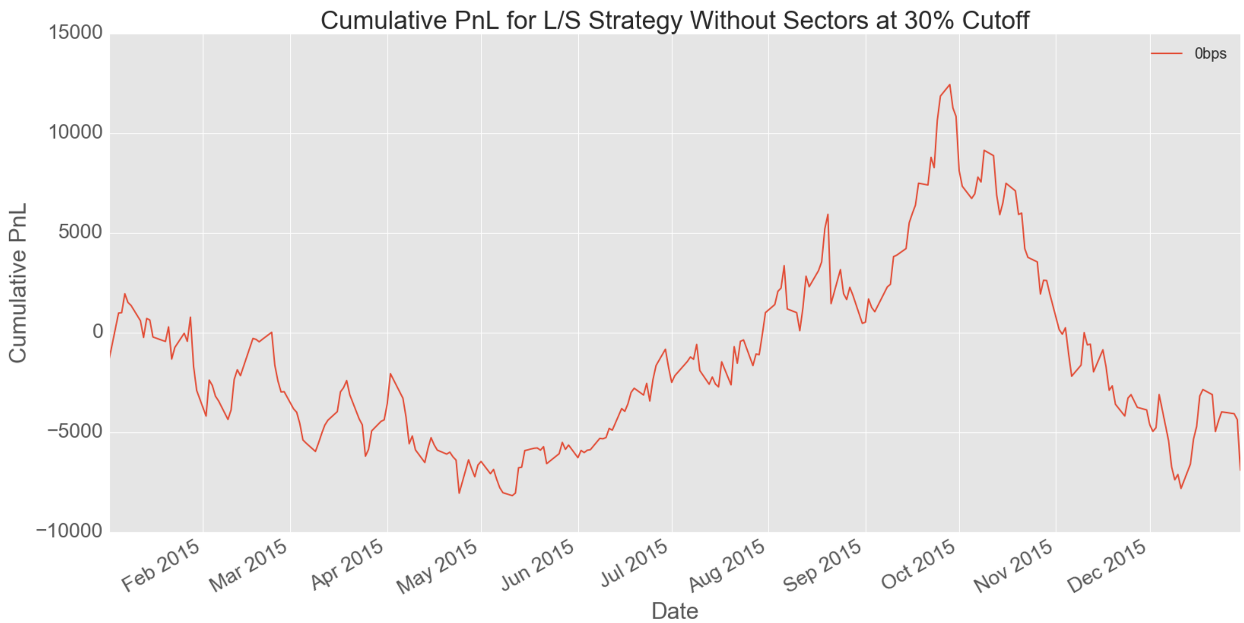


Figure 4: Strategy Performance by Year

Looking at the figure, it is easy to see that higher Sharpe ratios corresponded to higher profits in that year, as they did with the individual sectors. While the strategies were not more profitable than the Russell 1000 each year, they did seem to have higher Sharpe ratios in certain years. This could indicate that the two strategies had more consistent returns during that time than the Russell 1000. In fact, the results suggest that during 2012 and 2013, the Russell 1000, while more profitable in the end, was more risky to invest in than either of the two strategies. Moreover, it seems that while the returns of the Russell 1000 change quite dramatically year over year, the returns of the two strategies are more consistent, especially when the Russell is underperforming one year.

Analyzing the two strategies individually, the two differed in profits and Sharpe ratios by the most amount in 2010. Perhaps this is because the 30% cutoff strategy had unequal weightings for each sector in the portfolio. It could also be the case that the data for the earlier time periods when Twitter was not as popular may not be the most reliable because there were not as many people tweeting about particular companies. In fact, many of the members in the Russell 1000 did not even have a Twitter handle in 2010, so the signals for those stocks may not have been true to how investors felt about them.

2013, on the other hand, looks as though it was an incredibly profitable year for Twitter sentiment, with the Sharpe ratio for the 25% and 30% cutoff strategy at 3.27 and 2.96 respectively. 2013 was the year that Twitter had its IPO. Because of this, it may have been the case that the signals for Twitter sentiment were the strongest in this year because many professionals were attracted to the platform after it went public. Moreover, the fact that the profits for both cutoffs were increasing from 2011-2013 indicates that the hype leading up to the IPO helped the sentiment signal grow in strength. However, after 2013, it may be the case that the strategy began to get crowded as more portfolio managers began to incorporate the Twitter sentiment in their portfolios. 2015 in particular was especially unprofitable. Figure 5 shows the performance of the two strategies throughout 2015.



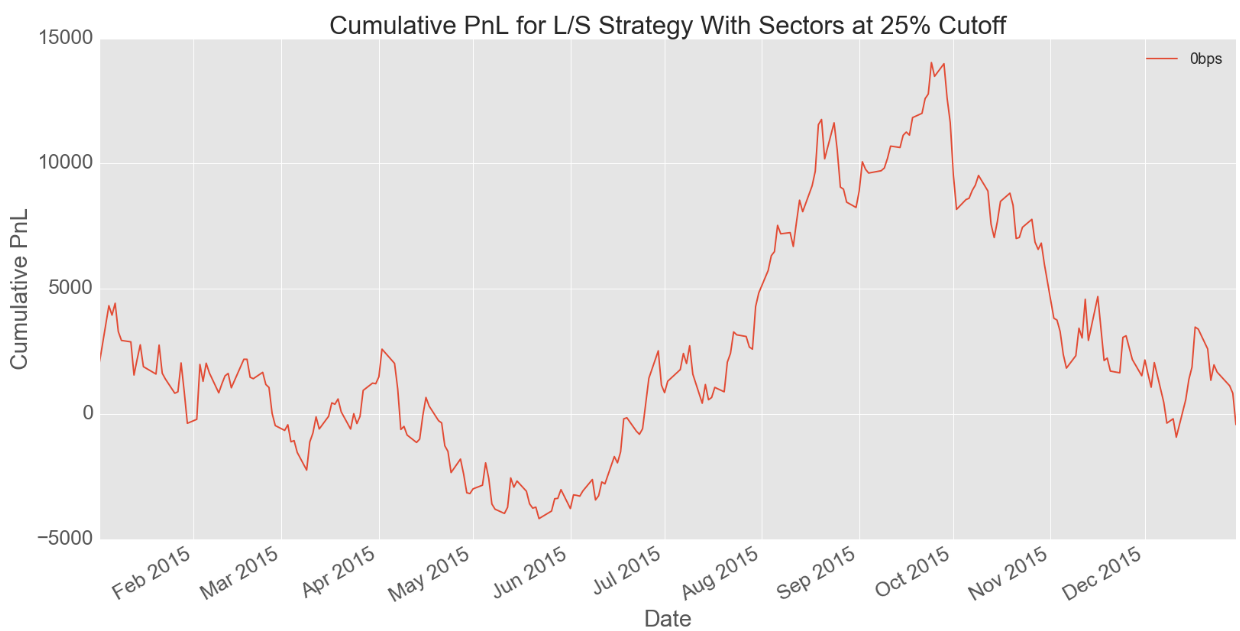


Figure 5: Out of Sample Cumulative PnL (0bps Transaction Cost)

One reason for why the strategy is not profitable in 2015 could be that many stock market participants began using Twitter sentiment to buy or sell certain stocks as the media began to hype up this topic. As more people began to put on the strategy, the profitability of the strategy shrunk.

When Figure 5 is examined closer, both strategies seemed to have performed well during the summer of 2015, when markets were very volatile due to a possible "Grexit" occurring as well as China's decision to devalue its currency. In July, Greece had defaulted on its debt payment and the European Union was hesitant on whether or not they should bail them out ("Referendum Result Takes Greece, and Eurozone, Into the Unknown", 2015). In August, China devalued the Yuan for two days in an effort to help stimulate their economy, as a weaker currency could help increase exports (Wei, Lingling, "China Moves to Devalue Yuan", 2015).

Since markets were extremely volatile during both events, every new piece of information sent markets wild. Going back to the original idea of Twitter being able to spread news faster than any other news source, following Twitter sentiment signals for the stock market would have benefited a trading desk immensely. This is because the sentiment would have been able to more accurately depict what investors will do. As a result, the signals may perform the best during periods of heightened volatility, though it is much easier justifying and crafting a logic behind a past event than it is to know whether it is truly right. Thus, the next strategy will try to strip out even more noise from day-to-day market trends.

Chapter 4: Pairs Trading Strategy

The focus of this strategy is to further filter out noise from the dataset. Consider two stocks, AAPL and AMZN for example, and make-believe that they have a very high correlation of above 0.90. Given their correlation, in theory, these two stocks should move together, i.e. if one of them increases in price, so should the other. However, this becomes interesting when the signal from Twitter contradicts the correlation. Continuing with the AAPL and AMZN example, if the Twitter sentiment signal on AAPL is a strong buy and the Twitter sentiment signal on AMZN is a strong sell, then it seems that the Twitter crowd believes that the two stocks will move in opposite directions of each other. As a result, if the signal truly does have an impact on the two stocks' price, one should be able to capture a small profit by buying AAPL and selling AMZN. Thus, the purpose of this strategy will be to test this very idea of using correlations to filter out the noise in the data.

Methodology

First, the rolling 21-day pairwise correlation of each stock of the Russell 1000 was calculated using each stock's daily close price. Most financial institutions use a 60-day correlation to determine the relationship between two stocks' returns. Bloomberg uses a one-year correlation. I chose to do 21-days because my trades are short term. Since I get in and out of a trade the same day, I care more about the short-term relationship of two stocks than I do about their long-term relationship. 21 business days is roughly equivalent to a month, which I believe is a good length of time to quantify a short-term relationship for a short-term trade based on my own internship experience at Goldman Sachs and the work done by the CME Group (Kamanski 2015). **Appendix E** provides several 21-day scatterplots to help aid the reader in understanding the benefits and drawbacks to using a 21-day correlation.

After the correlations are calculated, pairs are selected based on the following criteria:

1. Each pair must have a correlation of 0.90 or higher.
2. Each stock must not be used more than twice on any given day. This is to prevent overweighting the portfolio of one particular stock that is highly correlated to several other stocks.
3. If there exist multiple pairs that contain overlapping stocks with a correlation of >0.90, the algorithm will choose the pair with the stronger correlation. For example, if AAPL and AMZN have a correlation of 0.97 and AAPL and MSFT have a correlation of 0.91, and AAPL and NFLX have a correlation of 0.95, then the AAPL and AMZN pair and AAPL and NFLX pair will be traded because they have the two highest correlations of the three pairs. This step is to help enforce the second rule.
4. If two pairs have the same correlation, then algorithm will choose the pair that contains stocks in the same sector. This is because there is usually a reasonable fundamental argument for why they are correlated that is fundamental to their lines of business, such as Coca-Cola and PepsiCo.

Like the first strategy researched in this paper, different cutoff points for correlation were used to determine which one performs the best. After this, the algorithm looked at the signal for the selected pairs of stocks. If the signal was the same for the pair, i.e. both buy or both sell, then the algorithm did not trade the two stocks. If the signal was conflicting for the pair, i.e. buy one and sell the other, then the algorithm did as the signal indicated and bought the stock with the buy signal and sold short the stock with the sell signal. Similar to the first strategy with the percentiles, once the buys and sells were determined, the profits and losses of those stocks were calculated using the daily returns and summed. Each trade was put on with equal amounts of the total $1mm notional[2].

Performance

Table 6 shown below summarizes the results of the strategy's performance for each cutoff point in sample, which ranged from January 1st, 2010 through December 31st, 2014.

[2] Like in the long/short strategy, this strategy was run with a beta neutral portfolio as well. While the results of the 0.99 cutoff improved, weighting the portfolio by the beta did not dramatically improve the results of the other cutoffs in Table 6.

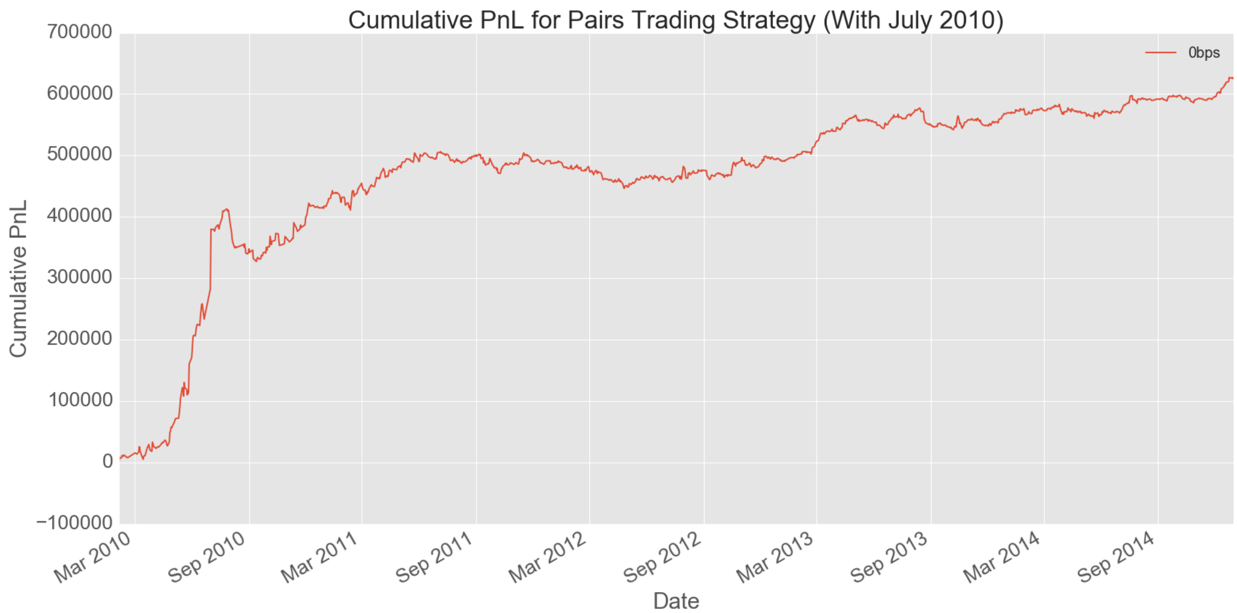
Table 6: In Sample Pairs Trading Performance Summarized

Correlation	0.90	0.96	0.97	0.98	0.99
# of Trades	215,814	78,560	48,408	20,182	3,090
Max PnL	$15,646.20	$97,250.00	$97,250.00	$97,250.00	$97,250.00
Max Loss	$(11,918.63)	$(17,308.45)	$(25,060.66)	$(28,563.27)	$(56,238.53)
% Winners	51.01%	50.25%	53.45%	51.63%	50.00%
Avg. PnL	$47.12	$232.49	$532.81	$240.33	$474.48
St. Dev	$2,079.06	$4,286.05	$5,186.34	$6,426.81	$10,099.38
Sharpe Ratio	0.3598	0.8611	1.6308	0.5936	0.7458
Cumulative PnL	$58,284.01	$281,312.89	$626,047.23	$243,458.76	$224,904.48
P-Value	0.426	0.060	0.000	0.234	0.307

Similar to the simple long short strategy, the most extreme correlation cutoff of 0.99 did not seem to perform the best. By limiting the strategy to only 3,090 trades in the entire in-sample dataset, it may have been the case that the strategy simply did not have enough data points to capture the true relationship between the signal and the stock's return. On the other hand, using a 0.90 correlation cutoff point had one of the lowest Sharpe ratios, lowest cumulative profits, and highest p-values—all indicating that the strategy at 0.90 correlation does not perform well. One possible reason for this may be that the strategy is including too much data, and thus, is introducing excessive noise into the returns. Another reason might be that a correlation of 0.90 over the previous 21 days may not have a corresponding Twitter sentiment signal that is strong enough to affect stock price. It could be the case that the signal for a pair that has a correlation of 0.90 is opposite (buy for one stock and sell for the other) because a correlation of 0.90 is simply not strong enough to capture the relationship between the two stock prices. In that case, it would make perfect sense to have a buy and a sell, rather than two buys or two sells. While there may not be too large a difference between the correlation of 21-day points with 0.95 and 0.96, it is evident that the number of stocks that qualified for that cutoff changed quite a bit with every hundredth of an increase in correlation. It could be that the results shown are random, but I still believe that there is a balance between trading too frequently and not trading enough. That balance is depicted in the 0.97 cutoff.

When the correlation between the pairs was held at 0.97 and higher, the Sharpe ratio, cumulative profits, and p-value beat those same statistics for any other strategy discussed in this paper. The strategy made an average of $532.81/day with an overall standard deviation of $5,186.34. Of the total $626,047 in cumulative profits at the end of 2014, the strategy made $97,250 in July of 2010. There were only two trades towards the end of the month, so each stock got $500,000 in capital to use on the trade. The Twitter sentiment signal was to buy AIG and sell short Tesla, a stock that had its initial public offering just a month before. Tesla was down over 16% that day, and since the strategy was short $500,000 worth of the stock, one trade made a considerable amount of profits. Moreover, the strategy traded Tesla stock multiple times that week to make comparable profits. It is fairly normal to have large movements in price directly after a stock has its initial public offering. Like the first strategy, these large swings in the share price of Tesla further reaffirm the hypothesis that the Twitter sentiment signal seems to be stronger and more useful when markets are volatile.

It might also be interesting to look at how the strategy performs without the Tesla trades in the month of July 2010. Figure 6 below shows the performance of the 0.97 correlation cutoff strategy between 2010-2014 with and without the summer of 2010.



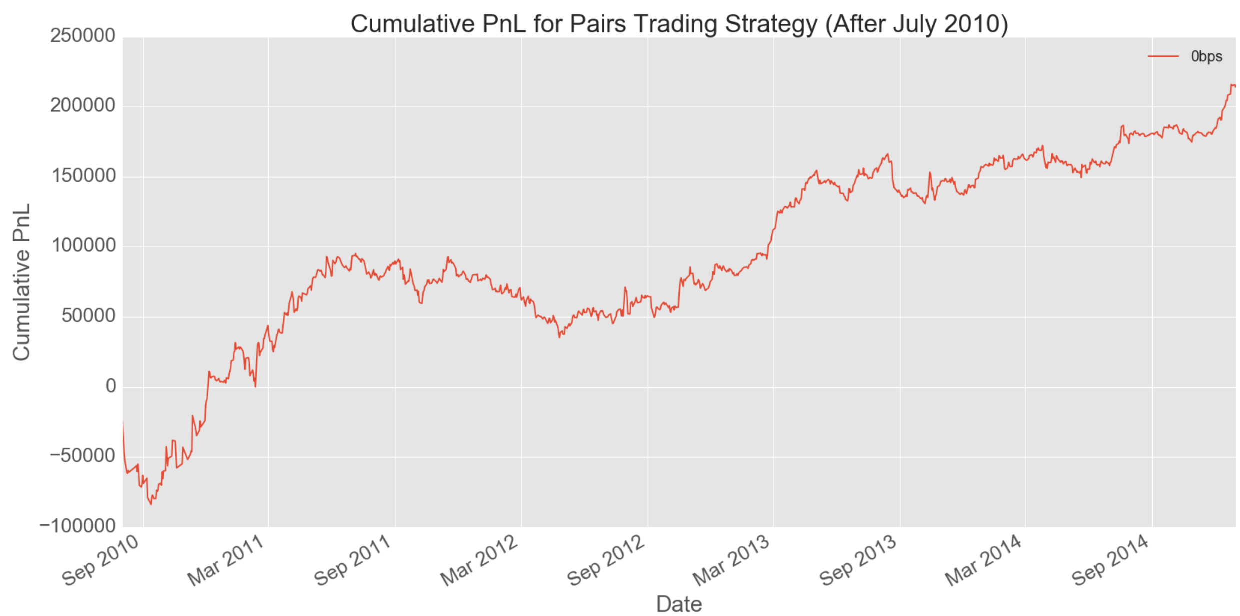


Figure 6: In Sample Cumulative PnL for Pairs Trading Strategy 0.97 Cutoff (0bps Transaction Costs)

When the months before August 2010 are removed, the Sharpe ratio of the strategy drops to 1.10. The strategy's average daily pnl and standard deviation drop to $311.56 and $4,285.10, respectively. The strategy makes positive pnl 52.90% of the trading days between August 1st, 2010-December 31st, 2014. It seems to be the case that the high profit during Tesla stock's early life pulled the average pnl higher by a considerable amount. Also, while the Sharpe ratio decreased for the strategy when July 2010 was taken out, it seems that the strategy still performs the best compared to the other correlation cutoff points. In fact, it beats the statistics of the other strategies in all ways, with one small exception of the average pnl of the 0.99 correlation cutoff.

In practice, taking out outlier days like this is heavily cautioned since removing very large profits or losses can artificially inflate the Sharpe ratio. It can also give one a false impression of the "extremes" of the strategy. For example, if one were to remove a day in which the strategy lost $100,000, then a desk that put on this strategy may not be adequately prepared for a scenario in which this happens. In some cases, that could be the tipping point of a trading desk going out of business. Thus, it is important to be extremely cognizant of the outliers one decides to throw out of the data.

The next test is to determine how robust this strategy was to transaction costs. Figure 7 shows how the 0.97 correlation strategy performs with varying transaction costs. Similar to the first strategy, the strategy stops being profitable after transaction costs are considered, despite the fact that the strategy begins to lose money slower than the simple long short strategy. Again, while it is important to consider transaction costs while testing out the performance of different alpha-generating strategies, most strategies are used in conjunction with others, so it can be assumed that the transaction costs are negligible for each individual strategy.

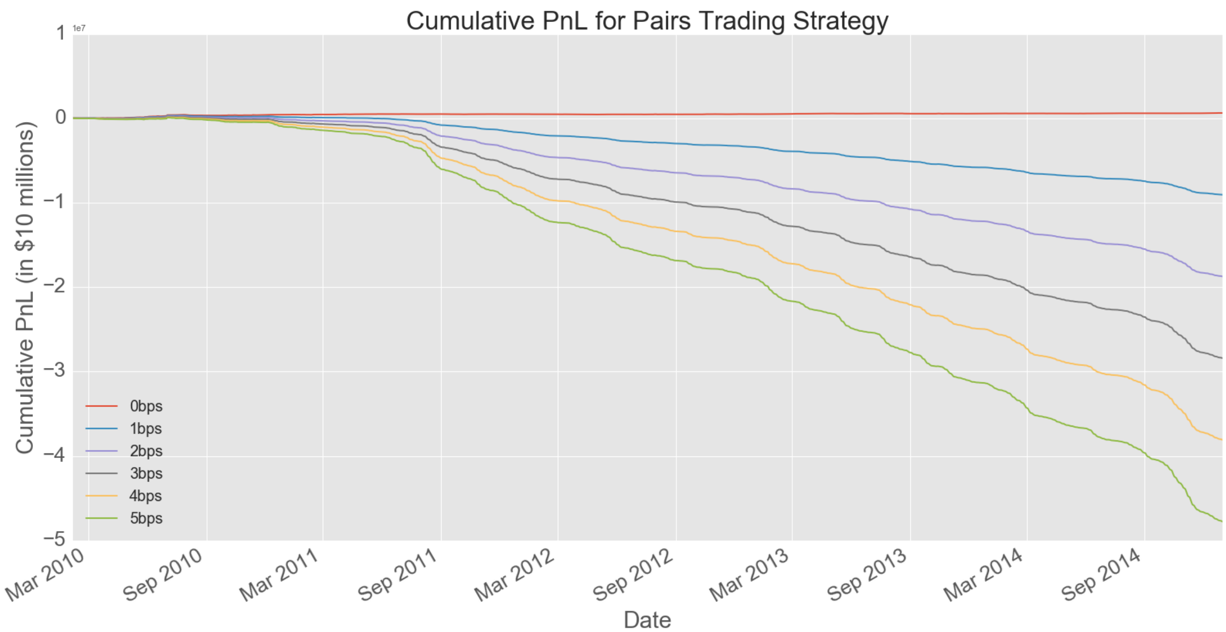


Figure 7: In Sample Pairs Trading Strategy Performance at 0.97 Cutoff With Transaction Costs

Because the 0.97 correlation strategy was profitable without transaction costs, I decided to test it out using the out of sample dataset to see if it were also robust to a different time period. Figure 8 shows how the pairs trading strategy performed in each year.

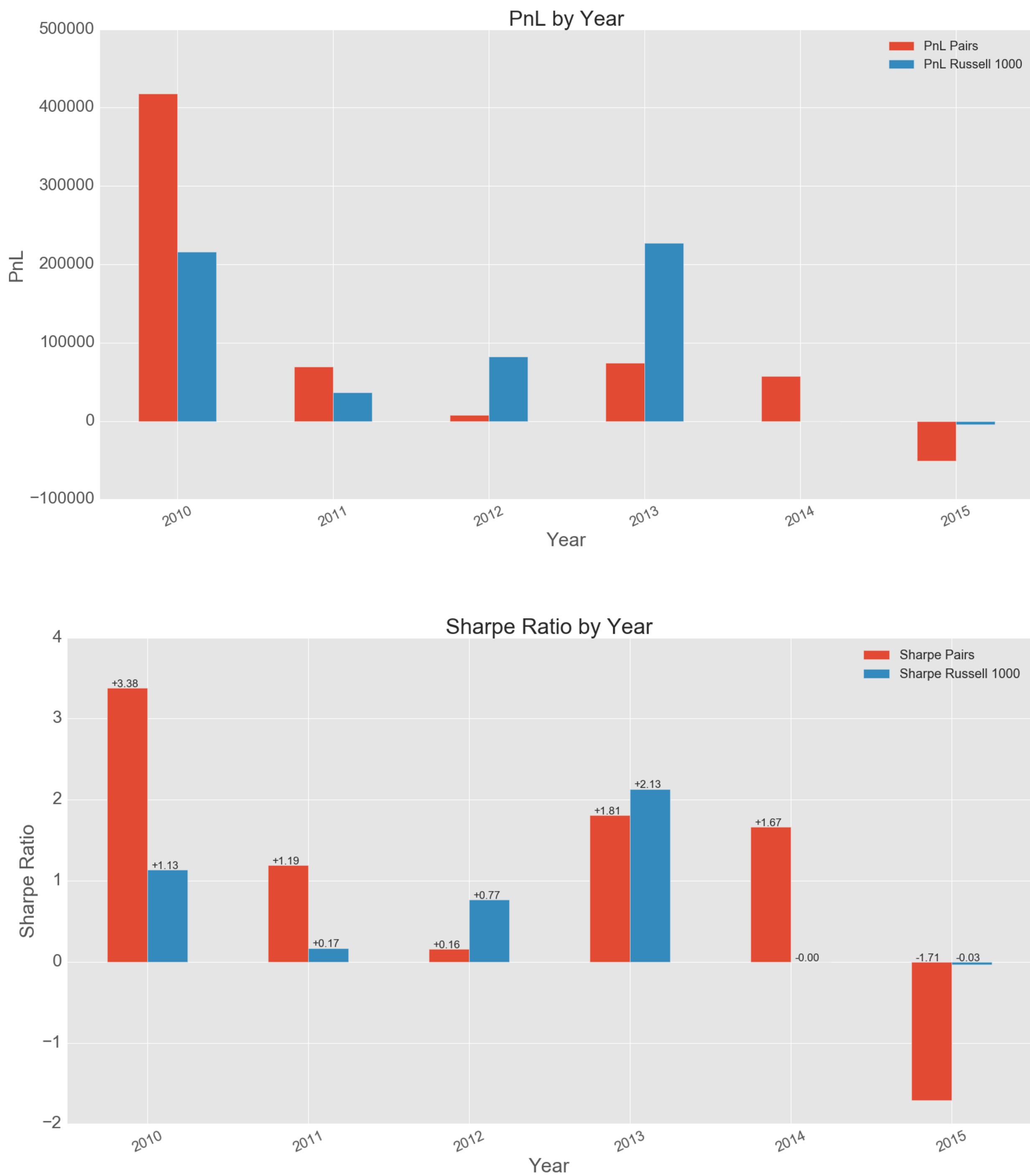


Figure 8: Pairs Trading Strategy Performance by Year

Similar to the yearly performance in the simple long short strategy, Figure 8 suggests that higher Sharpe ratios corresponded to higher profits in that year. With the exception of 2013, the fact that this strategy consistently beats the Russell 1000 returns supports the fact that this strategy may be better at modeling the true alpha. Moreover, if since the Sharpe ratios for the strategy are also consistently higher than the Russell 1000, it is fair to say that this pairs trading strategy has a higher overall return per unit of risk than the Russell 1000.

Ignoring the performance of 2010 and 2015, the overall profitability of the strategy was relatively stable, with the exception of 2012. It is also reassuring that despite the Tesla IPO trades in 2010, the Sharpe ratios for the strategy are above one and fairly steady. While it is hard to determine why 2012 was not a strong year for the pairs trading strategy, any of the following reasons could have contributed. First, 2012 was an election year. The Twitter crowd could have been focused more on the election than particular stocks. Given this possibility, the strength and power of the sentiment signals may have diminished. Second, the Federal Reserve continued their quantitative easing program. Because of this, many people may have shifted from tweeting about stocks to tweeting about US Treasury yields and what the Federal Reserve will do next. Third, the profits from 2012 may have just been random. There were several instances in which the sentiment signal on a stock would be negative and the return on the stock that day would be positive. The signal may have just not worked in 2012.

The best performing year was 2010, which had a series of incredibly lucrative trades as discussed above. Though, as we saw earlier, even without those particular trades, the strategy seemed to have performed better than the simple long short. Figure 9 shows the performance of the 0.97 pairs trading strategy throughout 2015.

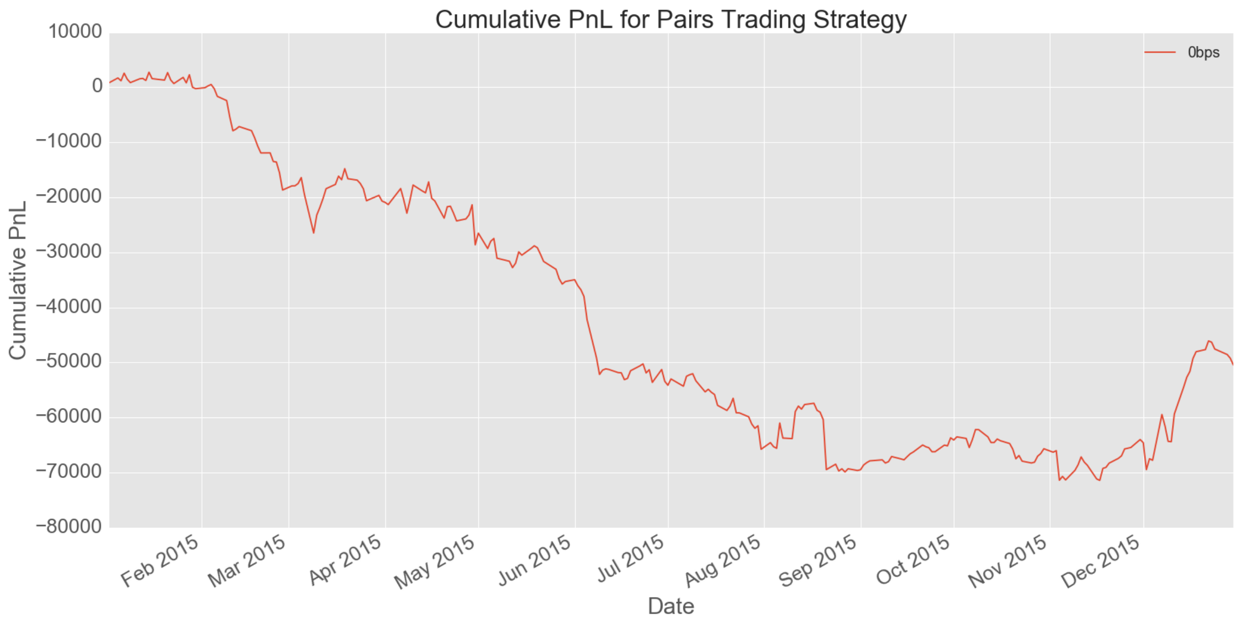


Figure 9: Out of Sample Pairs Trading Strategy Performance 0.97 Cutoff (0bps Transaction Cost)

Unfortunately, the strategy does not perform well in 2015. In addition to the reasons discussed in chapter 3 about poor strategy performance out of sample, one large factor could be the increase in the number of Twitter users. With more Twitter users tweeting about various stocks, the voice of an individual Tweeter is quieted. Consequently, there might inherently be more noise in the Twitter sentiment signal, and the signal may not accurately reflect the views of investors who are going to act in the market. Moreover, Twitter users saw a shift from using it as a source of news and self-expression to a means of following and getting updates on celebrities.

Another reason for this change may be the introduction of Symphony Communication, a cloud based communication platform created by a group at Goldman Sachs whose setup was extremely similar to Twitter. This platform was initially intended to be an internal messaging platform that banks and hedge funds would be able to use to communicate with each other. However, in 2015, the platform was made public in a partnership with Dow Jones, McGraw Hill Financial, and Selerity ("Symphony Launches Wall Street Network", 2015). This deal gave Symphony users the ability to view live news and breaking news stories directly on the platform, cutting back on the time it took a trader to open the news vendor's website and find an article about the story. This move made it more attractive for professionals using Twitter to voice their opinion on stocks to use Symphony instead, since it already had an established professional base of users. So, rather than discussing Apple earnings with a crowd that may not be focused or experienced, a serious finance Tweeter would be able to discuss his opinions with full-time traders and portfolio managers. This, coupled with the shift in the focus of Twitter, likely made the Twitter sentiment signals not as reliable in 2015. Nonetheless, understanding some of the changes in the markets and Twitter sentiment signals helps in further quantifying the impact of the signal on the stock market.

Chapter 5: Forward Guidance

Despite the data indicating that both strategies lost money with the addition of transaction costs, the fact that the strategies both made some money without costs suggests that there is valuable information in the data that can be advantageously used by financial institutions to make a profit. This chapter will discuss how a bank can use the strategies without drastically changing their day-to-day regimes. In addition to that, this chapter will discuss potential problems of the two strategies' models and provide further methods to analyze the data.

Implementation

A typical investment bank's trading floor has several large institutional clients who pay transaction fees to buy or sell stocks on their behalf for the best available price. One way that a bank can profit from the strategies discussed in this paper is through market making—the act of making both a bid to buy and offer to sell for a particular stock at a given time. By quoting both sides of a trade, the bank ultimately hopes to make a profit from the spread between the two. For example, if a stock were trading at $100 a share, a trader could make a market buy bidding for the stock at $99 a share and offering the stock at $101 a share. If someone is willing to sell at $99 a share and another client is willing to buy at $101 a share, the trader just made a profit of $101-99 = $2 for every share he or she sold/bought. The closer bid and offer to the current market price, the more likely it is that a client will be willing to buy from and sell the stock to the trader. Thus, a market of $99.99 bid and $100.01 offer would get more hits than the market of $99 bid and $101 offer.

Market making is one of the largest revenue generating paths of an equities trading floor at a large investment bank. Relatedly, one problem traders generally face when market making is balancing the trade-off between a wider market and the number of transactions they execute. A trader can market make with a $5 spread between the bid and the ask, but only have 5 clients take her up on the offer, generating a profit of 5*5 = $25 per share each client bought or sold. Or, the trader can market make with a $2 spread between the bid and ask and have 100 clients take her up on the offer. In the second scenario, the trader would have made $2*100 = $200 per share each client bought or sold. Often times, with markets being as unpredictable as they are, it is not entirely clear how tight or wide a trader should make the bid-ask spread.

One way to take advantage of the two strategies discussed here is to use the information in the signals to make wider or tighter markets. For example, if a trader sees that the signal on AAPL is a strong buy, he can adjust his bid to buy AAPL stock at a price closer to the market price of AAPL. By foregoing some of the profits the trader could have made by bidding to buy AAPL at a lower price, the trader is more likely to get her entire bid order filled. So, in actuality, the trader makes up for the higher price by purchasing more shares and selling them at a higher price at the close, according to either strategy discussed in this paper.

Another way the trading floor of a bank makes money is through taking on inventory for securities that their clients want to unload from their portfolios. When a client wants to do this, they will pay the bank to take the stocks they want to unload, on top of the bid-ask spread. Given this, if a bank knows that the signal on AAPL from Twitter sentiment is a strong buy, then the bank can take on the inventory for one day. Not only will this cut down on transaction costs for the strategies, but it will also allow the bank to buy AAPL at a slight discount, thus increasing the potential profits of the trade. If a bank is able to do this for even a fraction of the stocks in the Russell 1000 whose signals meet the standards of either strategy discussed, the strategies' cumulative profits and performance would look a lot different.

Finally, a third way large investment banks can profit from these strategies is use them alongside other alpha-generating strategies. A trading desk can quantify the relationship among multiple strategies to see if including one of the two strategies discussed in this paper makes the signal for another strategy stronger. They can also pair this strategy with another strategy that is uncorrelated to take advantage of the two different signals for inefficiencies in the market. By doing so, the transaction costs will be spread out among multiple strategies, and perhaps they will all individually have a net positive pnl. This same logic and implementation ideas can be applied to brokers, financial advisors, and market makers. Nevertheless, regardless of how and if these strategies are used in practice, it is important to note that the results of this study, while limited, suggest that considering investor sentiment on Twitter may be useful in informing the one-day performance of the stocks in the Russell 1000. However, before any of these strategies are implemented, and before I can personally recommend one of them, I strongly believe that more research on this topic needs to be conducted.

Model Biases & Further Research

All of the results discussed in this paper relied only on one sentiment signal provider's data, but perhaps Psych Signal is not the best at capturing the true investor sentiment from Twitter. Psych Signal did not provide insight on how many of the tweets scanned by their algorithm were from the company that the tweets were about. It would have been interesting to see how a company's presence on social media changed when there were a lot of negative tweets about the company. Did the company itself tweet a lot of positive news to counteract the negative press? Naturally, no company wants to have a negative image in the press, so they may actually be incentivized to attempt to bias the sentiment on social media, and as a result, the sentiment signals from Twitter may be very different if the algorithm excluded the twitter handles of the companies in the Russell 1000. Before making any solid conclusions about the impact of Twitter sentiment on the stock market, it would only be wise to repeat this study with data from other sentiment providers.

Additionally, since this paper only focused on the members of the Russell 1000, there may have been an implicit bias in the performance of large capitalization US stocks. Maybe if another index were to have been used, the results would have differed. It may be the case that the signal is the strongest for small-cap stocks rather than large cap stocks that are tweeted about more frequently.

Repeating the same strategies for other indices and then comparing them may be worthwhile.

It would have been interesting to look at the data on a minute-by-minute basis, since many stocks react very quickly to news, positive or negative. Since the signal is received at 7pm, the strategy is forced to trade the next day. This is an issue because a lot can change from 7pm to 9:30am the next day. News that might have been relevant at 6pm may not be as relevant the next day, and so the signal may not be accurate. As a result, any return captured using the signal as an indicator may be random and not meaningful.

Another issue is that it is hard to determine whether the high sentiment signal was the one that caused the price movement in the stocks, or whether it was the price movement in the stocks overall that caused the high sentiment signal. It would be fascinating to see how the price of a stock reacted immediately after a signal is received. Unfortunately, since the signal was provided on a daily basis, and because data on after hours trading is incredibly difficult to obtain, this idea could not be tested further. Moreover, if each tweet that went into the sentiment signal was weighed according to the time it was written, the daily sentiment signal may have been a better indicator of 1-day performance. This is because markets may have already reacted to tweets intraday, but did not yet have a chance to react to the tweets after the close. If there is a sentiment provider that weighs the tweets by the time of day that they were written, then perhaps that study would be able to more confidently make recommendations as to how to trade sentiment signals.

Frequently, sentiment is used as a contrarian signal. Because of this, a stock with a strong buy signal may not result in as significant an increase in price as it would if no one sold short the stock when the signal was to buy. Given this idea, a person using this signal as an indicator of stock price or return would not be able to capture the true relationship between the two. So, rather than using Twitter sentiment as a signal for stock market performance, perhaps it makes more sense to use it to predict trade volume for particular stocks.

To help the future researchers of this topic, I am also releasing part of the Python code used to run the analysis in **Appendix F**. Since a lot of the setbacks I had when writing this thesis related to writing the most efficient code to make the calculations I wanted, I attempted to write functions that can be modified easily to cater to what you may need. Given the results of this study, I would advise investment banks and hedge funds considering purchasing this data to look deeper into the factors described above, perhaps with a different data vendor that shares the scores for individual tweets. Moreover, this study is still in its elementary stage, and I do believe that there is a lot more to be learned by exploring Twitter sentiment.

References

"Twitter Turns Six." Twitter. Twitter, Inc., 21 Mar. 2012. Web. 14 Dec. 2016.

Ahern, Kenneth R., and Denis Sosyura. "Who Writes the News? Corporate Press Releases during Merger Negotiations." *The Journal of Finance* 69.1 (2015): 241-91. Web.

Baker, Malcolm, and Jeffrey Wurgler. "A Catering Theory of Dividends." *The Journal of Finance* (2004): n. pag. Web.

Baker, Malcolm, and Jeffrey Wurgler. "Investor Sentiment in the Stock Market." *Journal of Economic Perspectives* (2007): n. pag. Web.

Barber, Brad M., and Terrance Odean. "All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors." *Review of Financial Studies* 21.2 (2008): 785-818. Web.

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny. "A Model of Investor Sentiment." *Journal of Financial Economics* (1997): n. pag. Web.

Bary, Andrew. "How Investors Should Play AT&T Bid for DirecTV." *Barron's*, 19 May 2013. Web. 17 May 2016.

Bender, Jennifer. "Foundations of Factor Investing." MSCI, Dec. 2013. Web.

Brown, Gregory W., and Michael T. Cliff. "Investor Sentiment and Asset Valuation." *The Journal of Business* 78.2 (2005): 405-40. Web.

Bollen, Johan, Huina Mao, and Xiaojun Zeng. "Twitter Mood Predicts the Stock Market." *Journal of Computational Science* 2.1 (2011): 1-8. Web.

Bordino, Ilaria, Stefano Battiston, Guido Caldarelli, Matthieu Cristelli, Antti Ukkonen, and Ingmar Weber. "Web Search Queries Can Predict Stock Market Volumes." *PLOS ONE* 7.7 (2012): n. pag. Web.

Curme, C., T. Preis, H. E. Stanley, and H. S. Moat. "Quantifying the Semantics of Search Behavior before Stock Market Moves." *Proceedings of the National Academy of Sciences* 111.32 (2014): 11600-1605. Web.

Da, Zhi, Joseph Engelbert, and Pengjie Gao. “The Sum of All FEARS: Investor Sentiment and Asset Prices.” *Review of Financial Studies*, Vol 28, 1-32 (2015). Web.

Dang, Ha V., and Mi Lin. "Herd mentality in the stock market: On the role of idiosyncratic participants with heterogeneous information." *International Review of Financial Analysis* 48 (2016): 247-260.

Dimpfl, Thomas, and Stephan Jank. "Can Internet Search Queries Help to Predict Stock Market Volatility?" *European Financial Management* 22.2 (2015): 171-92. Web.

Dougal, Casey, Joseph Engelberg, Diego Garcia, and Christopher A. Parsons. "Journalists and the Stock Market." *SSRN Electronic Journal* (2012): n. pag. Web.

Drew, Christopher. "Boeing Speaks in Trump Terms on Iran Deal: It's About Jobs." *The New York Times.* The New York Times, 11 Dec. 2016. Web. 14 Dec. 2016.

Evan Gatev & William N. Goetzmann & K. Geert Rouwenhorst, 2006. "Pairs Trading: Performance of a Relative-Value Arbitrage Rule," *Review of Financial Studies*, Oxford University Press for Society for Financial Studies, vol. 19(3), pages 797-827.

Fama, Eugene F., and Kenneth R. French. "The Cross-Section of Expected Stock Returns." *The Journal of Finance* 47.2 (1992): 427. Web.

Fidler, Stephen. "Referendum Result Takes Greece, and Eurozone, Into the Unknown." *The Wall Street Journal*. Dow Jones & Company, 05 July 2015. Web. 27 Jan. 2017.

Gervais, Simon, Ron Kaniel, and Dan H. Mingelgrin. "The High-Volume Return Premium." *The Journal of Finance* 56.3 (2001): 877-919. Web.

Hou, Kewei, Wei Xiong, and Lin Peng. "A Tale of Two Anomalies: The Implications of Investor Attention for Price and Earnings Momentum." *SSRN Electronic Journal* (2009): n. pag. Web.

Huberman, Gur (2005): Arbitrage pricing theory, Staff Report, Federal Reserve Bank of New York, No. 216.

Jorion, Philippe. "Risk Management Lessons from Long-Term Capital Management." *SSRN Electronic Journal SSRN Journal* (n.d.): n. pag. Web.

Jussa, Javed. *Signal Processing*. New York: Deutsche Bank, n.d. 24 Apr. 2013. Web.

Kaminsk, Katheryn. "Return Dispersion, Counterintuitive Correlation: The Role of Diversification in CTA Portfolios." *CME Group* (n.d.): n. pag. Feb. 2015. Web.

Khan, Ronald N. "What Practitioners Need to Know About Backtesting." *Financial Analysts Journal* 46.4 (1990): 17-20. Web.

Kumar, Alok, and Charles M.c. Lee. "Retail Investor Sentiment and Return Comovements." *The Journal of Finance* 61.5 (2006): 2451-486. Web.

Lee, Seong. "Long-only Trading Strategy with NLP Derived Social Media Sentiment - Tear Sheet Attached." *Quantopian*. Quantopian, 18 May 2016. Web.

Luo, Yin. "QWAFAFEW Presentation." (n.d.): n. pag. Deutsche Bank, Jan. 2015. Web.

McLannahan, Ben. "Symphony Launches Wall Street Network" *Financial Times*, 17 Oct. 2015. Web. 21 Feb. 2017.

Mossin, Jan. "Equilibrium in a Capital Asset Market." *Econometrica* 34.4 (1966): 768. Web.

O'Brien, Timothy and Laura. "A Hedge Fund's Stars Didn't Tell, and Savvy Financiers Didn't Ask." *New York Times* (1923-Current file): 2. Oct 23 1998. ProQuest. Web. 16 May 2016.

Rennie, Shannon. "Do Tweets Matter: Assessing the Value of Social Media Marketing." *Drew University.* May 2016. Web. 14 Dec. 2016.

Ritholtz, Barry. "Twitter Is Becoming the First and Quickest Source of Investment News." *The Guardian*. Guardian News and Media, 23 Apr. 2013. Web. 14 Dec. 2016.

Ross, Stephen A. "The Arbitrage Theory of Capital Asset Pricing." *Journal of Economic Theory* 13.3 (1976): 341-60. Web.

"Rubber Faqs." *Rubber Manufacturers Association*. N.p., n.d. Web.

Sharpe, William F. "Capital Asset Prices: A Theory Of Market Equilibrium Under Conditions Of Risk." *The Journal of Finance* 19.3 (1964): 425-42. Web.

Smales, Lee A. "Impact Of Macroeconomic Announcements On Interest Rate Futures: High-Frequency Evidence From Australia." *Journal of Financial Research* 36.3 (2013): 371-88. Web.

Smith, Bradley H. "Investor Relations Study: 1 in 4 of Institutional Investors Use Social Media for Research." *PR Newswire*. PR Newswire, 02 Dec. 2016. Web. 27 Jan. 2017.

Stone, William. "Factor Analysis: What Drives Performance?" *Investment and Portfolio Strategy* (n.d.): n. pag. PNC Bank, Feb. 2014. Web.

Tetlock, Paul C. "Giving Content to Investor Sentiment: The Role of Media in the Stock Market." *The Journal of Finance* 62.3 (2007): 1139-168. Web.

Treynor, Jack L. "Market Value, Time, and Risk." *SSRN Electronic Journal* (1961): n. pag. Web.

Vidyamurthy, G. Pairs Trading. *John Wiley & Sons*, 2004.

Vieira, Elisabete SimÃµes. "Investor Sentiment and the Market Reaction to Dividend News: European Evidence." *Managerial Finance* 37.12 (2011): 1213-245. Web.

Wei, Lingling. "China Moves to Devalue Yuan." *The Wall Street Journal.* Dow Jones & Company, 11 Aug. 2015. Web. 7 Feb. 2017.

Zhao, Yiwei, Zheng Yang, and Xiaolin Qian. "Investor Sentiment and Chinese A-Share Stock Markets Anomalies." *International Journal of Economics and Finance* 7.9 (2011): n. pag. Web.

Appendix A

Justin Pulitzer
@JustinPulitzer

Follow

Why shouldn't $TSLA have a higher Market Cap than $F? Tesla actually builds products people want… not junk to feed union labor contracts.

RETWEETS 8 LIKES 26

11:48 AM - 3 Apr 2017

4 8 26

Voenix Harmonics
@VoenixHarmonics

Follow

$AAPL #AAPL #Apple Bullish White Swan Forming on the Daily
mql5.com/en/market/prod…

RETWEET 1 LIKES 2

5:37 PM - 3 Apr 2017

1 2

Barron's @barronsonline

Follow

#EmergingMarkets Roundup: Record Fund Inflows, Russia Rate Surprise on.barrons.com/2n1pblZ $EEM $BHP $PEK $CAT $RSX via @DeFotis

RETWEET
1

9:33 AM - 24 Mar 2017

1

Causal Nexus @CausalNexus

Follow

President Trump just announced full speed ahead on tax reform. $AAPL $MSFT $IBM $GOOGL $CSCO $ORCL $C $GE $AMGN $PFE

> **Causal Nexus** @CausalNexus
> @realDonaldTrump Let's expedite repatriation tax reform so companies can bring $ home, invest locally, increase jobs & boost economic growth

LIKE
1

2:35 PM - 24 Mar 2017

1

Appendix B

Potential Drawbacks to Pairs Trading: A Case Study on Long Term Capital Management

This seemingly market-neutral strategy can be extremely risky and detrimental to the profits of the firm if not executed properly. One prime example of the riskiness of this strategy is with a hedge fund that started and failed in four years, Long Term Capital Management (LTCM). The fund, which was founded by John Meriwether in 1994 after he left the bond-arbitrage group at Salomon Brothers, used pairs trading as its core strategy. Rather than applying the strategy to stocks, though, LTCM used it to find a mispricing in closely related fixed income securities.

Take, for instance, an off-the-run Treasury bond that has a yield of 5.5%, and an on-the-run Treasury bond that has a yield of 5%. The two bonds have almost exactly the same characteristics, with the only difference being that the off-the-run Treasury bond has higher liquidity risk than the on-the-run Treasury bond because it is not traded as frequently. The key thought process regarding a pairs trade is that securities with similar characteristics will converge to the same relative price. Therefore, since bond prices and yields move in opposite directions, a simple strategy to take advantage of this "mispricing" would be to long the off-the-run 5.5% Treasury bond and short the on-the-run 5% Treasury bond. That way, as the two prices begin to converge, the trader would make a profit of 50 basis points per dollar invested into the trade. These types of strategies were not only used by LTCM for Treasury bonds, but also for pairs like mortgage-backed securities and government bonds, high-yield and low-yield bonds, and convertible bonds (Jorion, 2000, 3).

One major downside of using pairs trades to make a profit is that since financial markets are extremely efficient, the amount of mispricing in the two securities one can find is often small, so the profit, given that the fund invests using no leverage, that one can extract from it is also small. Leverage is often used to make an unattractive-looking profit worth taking a risk for. By borrowing money to make a trade, or trading on margin, traders can take on leverage to invest more money into each trade. For example, if a firm has $10,000 of assets to allocate to a single trade, the most the firm can make from the Treasury bond pairs trade is 0.005*10,000 = $50. On the other hand, if the firm was levered up and had $1,000,000 to allocate to a single trade, the most the firm can make from the Treasury bond pairs trade is 0.005*1,000,000 = $5,000. As a result, the leverage on the pairs trades at LTCM had to be large, and often ended up with assets that were four times the assets of the next largest hedge fund, representing "an astonishing leverage ratio of 25-to-1" (Jorion, 2000, 3). This means that for every $1 asset the firm had on hand, it would borrow $25 of assets.

It is important to mention here that taking on leverage is not without risk. A higher capital investment in a trade could lead to higher profits, but also higher losses. A 10% loss on $10,000 is $1,000, but a 10% loss on a $1,000,000 is a hefty $100,000. Moreover, if a highly leveraged firm cannot pay back the money it owes, or if it takes a loss that wipes out its initial capital, the firm receives a margin call and is forced to liquidate its positions and pay back its borrowers immediately. Thus, if LTCM had \$10,000 initial cash under management, and it was leveraged 25-to-1, the firm would have a total of \$250,000 of assets to manage. However, if on any given day, LTCM were to take a loss of more than 4%, the firm would be forced to liquidate its assets and any of its investors would not be paid back before their debts were paid. While being leveraged 25-to-1 initially was not a problem for the hedge fund returning 40% after fees in the first couple years, it began to become a problem for LTCM when its assets grew since it is harder to achieve a 40% growth on \$1,000,000 than on \$10,000. Consequently, as the fund's assets grew, so did its leverage. Figure 10 below depicts the growth of LTCM's leverage and asset over its lifespan. The amount of leverage they took on throughout the years is shown along the solid line and reported on the left y-axis, and the amount of assets they had under management is shown along the dotted line and reported on the right y-axis. At one point in 1997, LTCM forced out some of its investors so that they could grow their leverage ratio and amplify returns to the remaining investors (Jorion, 2000, 6). It is certainly clear that towards the end of the fund's lifespan, leverage was more than twice as high as it was initially.

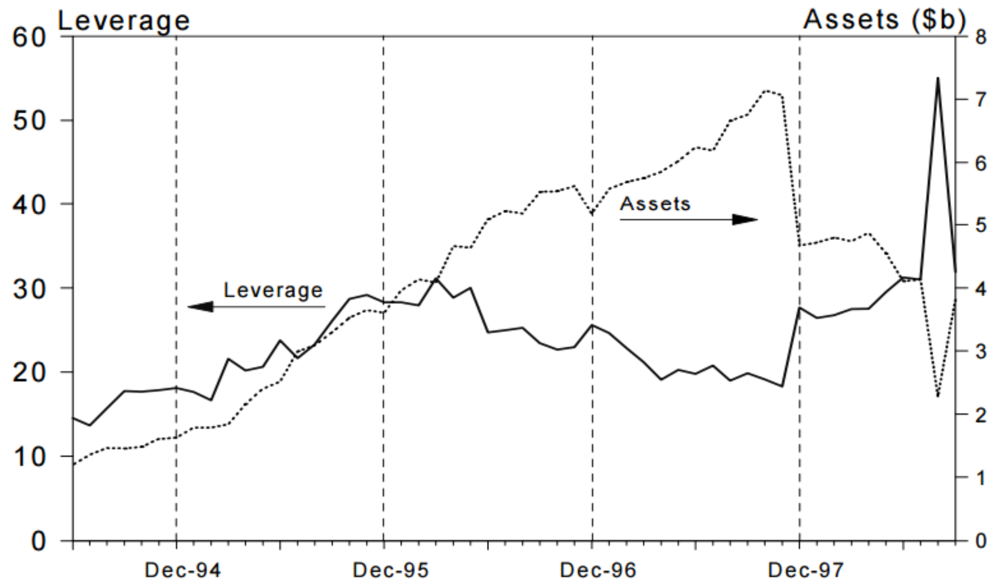


Figure 10: LTCM Leverage and Asset Growth Over Time (Jorion, 2000, 6)

With this higher leverage, the fund was open to larger losses on the chances that markets fluctuated suddenly, like in 1998 when Russia defaulted on its debt. When this happened, "credit spreads, risk premia, and liquidity spreads jumped up sharply. Stock markets divided. LTCM lost $550 million on August 21 alone" (Jorion, 2000, 7). Even the least volatile of spreads moved by several basis points each day, and by the end of August, LTCM had lost over 52% of its assets. Markets continued to be volatile and by September 21st, the fund received a margin call. Since LTCM had billions of dollars under management, liquidating out of these positions would have wreaked havoc on already volatile financial markets, the Federal Reserve actually encouraged big banks to help facilitate the bailout of LTCM. Unfortunately, this bailout only helped to cushion the fall of LTCM. By September 28th, the value of the firm dropped to only $400 million, and it had lost 92% of its year-to-date investments. In a New York Times article by Timothy O'Brien and Laura Holson, $3 billion of the $4.4 billion lost by the fund came from pairs trades involving interest rate swaps and volatile equities (“A Hedge Fund's Stars Didn't Tell, and Savvy Financiers Didn't Ask.”, New York Times, October 1998).

Several lessons can be learned from this hedge fund’s failure, though the most important one concerns the leverage they took on their trades. Although the trades were based on statistical reasoning, it is not wise to assume that the trade will play out forever. In fact, one crucial flaw in a pairs trading strategy is that the correlations on the trades change at different points in time because of fundamental changes in the underlying security. Therefore, if the correlation during one year between two securities is 0.97 and it drops to 0.80 in the following year, the strength of the pairs trade is diminished and there is more room for an error since the two securities’ prices do not move as closely together as before. When a firm is as leveraged as LTCM was, this change in the correlation can adversely affect profits for the company because, as pointed out earlier, even a small drop in a security’s price can bear down on the firm’s losses. One of the larger trading strategies LTCM used was a pairs trade between corporate and treasury yields. Historically, the correlation between the two securities’ yields had been above 0.94, but this changed as markets began to change in the late 80’s and 90’s. Figure 3 shows a graph of the correlations between corporate and treasury yields over time.

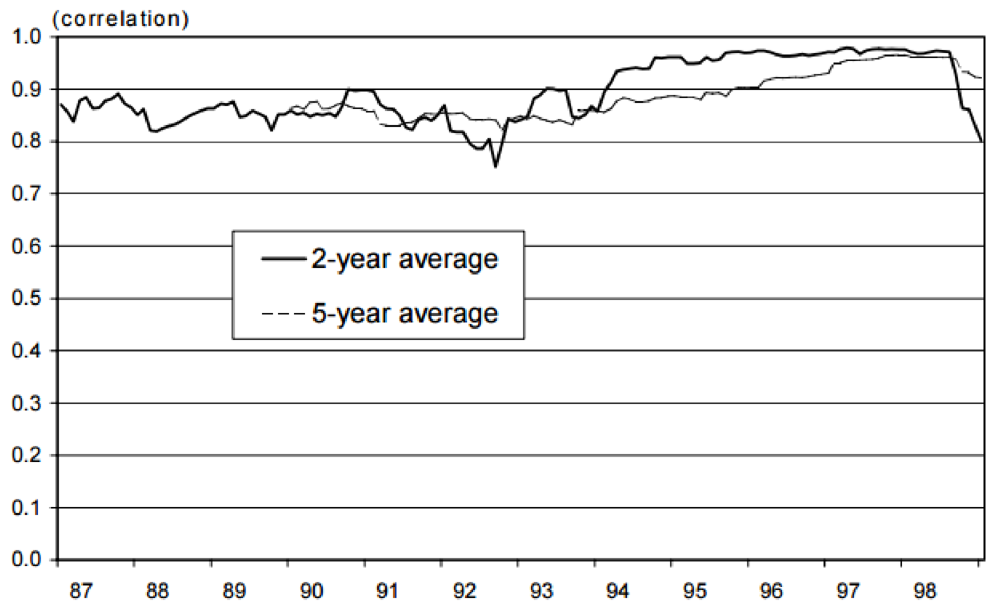


Figure 11: Correlation Between Corporate and Treasury Yields (Jorion, 2000, 19)

The solid line shows the 2-year rolling correlation while the dotted line shows the 5-year rolling correlation. It is clear that in 1988, the 2-year correlation had dropped to roughly 0.80, which is considerably different from the 0.94 that it used to be. Since a decrease in correlation leads to an increase in the error and loss, this drop could be a leading statistical reason explaining why LTCM lost so much of their capital that year.

Appendix C

EOD Database

This database, published by Quote Media, provides daily open and close prices for all publicly traded US stocks as well as their respective volumes. It covers all stocks trading on NASDAQ, AMEX, NYSE, and ARCA. All prices and volumes are also adjusted for dividends, stock splits, and spinoffs.

For dividends, the Adjustment Ratio = (Close Price + Dividend Amount) / (Close Price).

For stock splits, the Adjustment Ratio = Split Ratio.

For spin-offs, the Adjustment Ratio = 1 + (Spinoff Open Price * Spinoff Shares) / (Parent Open Price * Parent Shares).

All stock prices are reported at 5pm EST on trading days. Moreover, for increased accuracy, the database prices are cross-checked daily with other data sources such as Yahoo Finance and Bloomberg.

Psych Signal

Psych Signal is a provider of real time Trader Mood data and analytics from Twitter and Stock Twits. The company launched on October 1st, 2011 with the goal of quantifying the sentiment between investors on social media. The reason for choosing the psych signal database over other sentiment score providers was twofold. First, the data was the most economic option to purchase for a student. Second, Psych Signal uses natural language processing and machine learning to calculate the sentiment on each tweet. This is better than a simple "bag-of-words" dictionary approach because it is able to pick up on subtle nuances in the English language (Mass 2011, Spice 2016). A snippet of the final dataset is shown in Table 7 and the process by which the signal is extracted is further described below.

Table 7:Psych Signal Sample Dataset

Date	Ticker	Bullish Score	Bearish Score	Bull – Bear Score	Total Tweets Scanned
01/01/2010	AAPL	0.67	2.1	-1.43	35
01/01/2010	MSFT	1.37	1.93	-0.56	15
01/02/2010	MSFT	0	1.89	-1.89	19

First, each tweet on Twitter or Stock Twits is filtered into a bucket unique to the stock it is referring to using identifiers such as cash tags and hash tags. Only Tweets about the particular stock are taken into consideration; tweets about the market as a whole are not used in the analysis.

Next, the tweeter's individual Twitter username is analyzed and further categorized by level of significance. For example, a tweet from credible news sources such as the WSJ or Bloomberg will have a higher level of significance than a tweet by a student like myself.

After this, each tweet is read and scored through Psych Signal's proprietary algorithm using natural language processing. This method essentially allows a computer program to read and understand text like a human would, since words can sometimes have multiple meanings depending on how they are used in a sentence. Phrases like "Apple killed earnings today" are correctly marked as positive news for the stock, rather than negative due to the word 'killed'.

Then, two weighted average sentiment scores are calculated every day after markets close using the level of significance scale mentioned previously. The "bullish score" represents one score for all of the positive tweets, while a "bearish score" represents one score for all of the negative tweets. Both of these scores range from 0-4, with zero representing an instance in which there were no positive or no negative tweets. The final score is calculated by subtracting the bearish score from the bullish score.

Appendix D

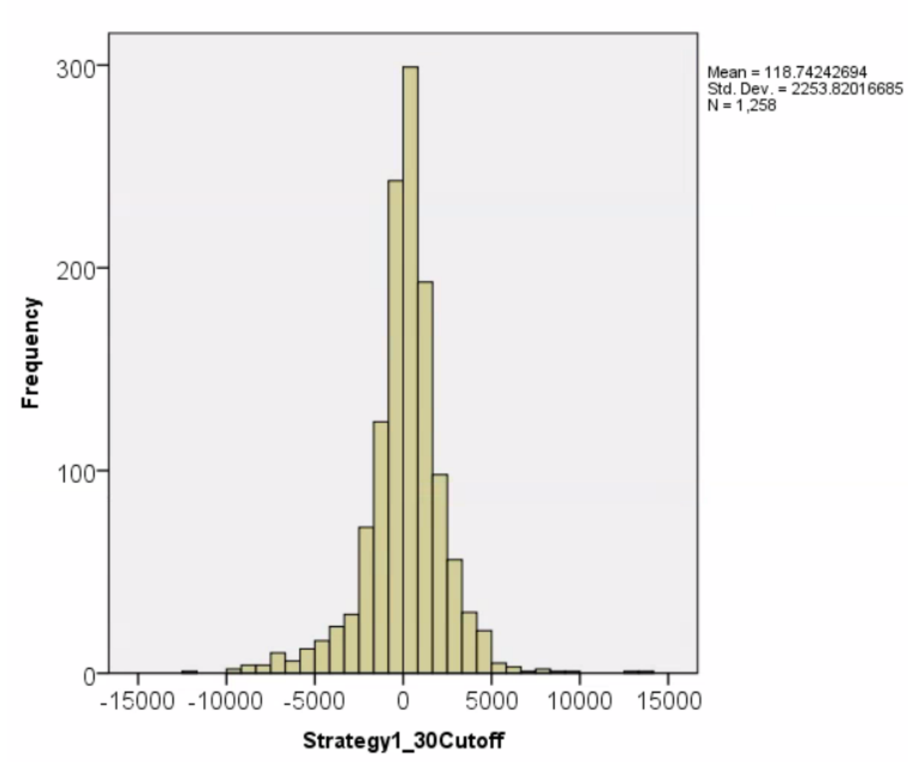
One Sample t-test

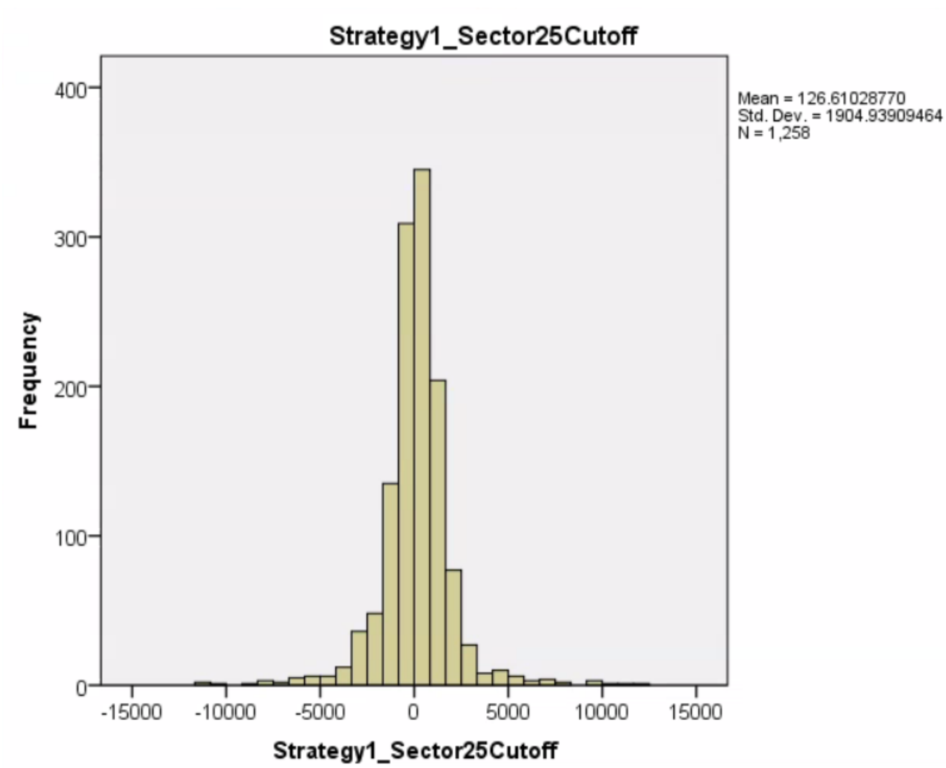
A one sample t-test is often used to determine whether a sample mean is statistically significantly different from a known or hypothesized population mean. In order to use this test, the following criteria must be made:

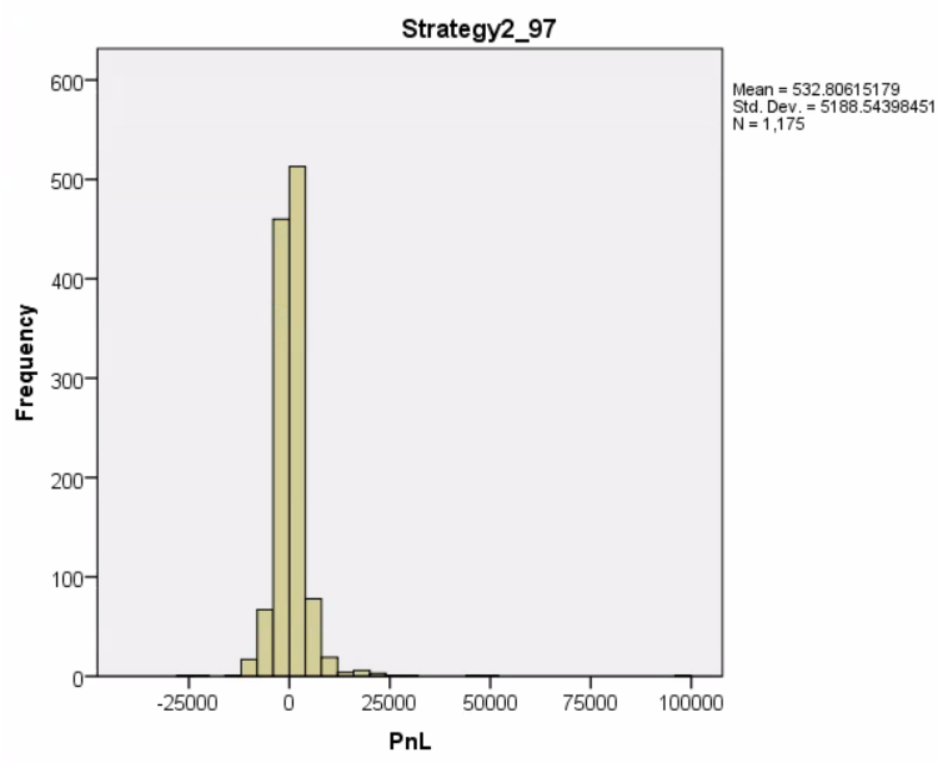
1. The data is continuous.
2. The sample being tested is a random sample from its population.
3. The data has a roughly normal distribution.

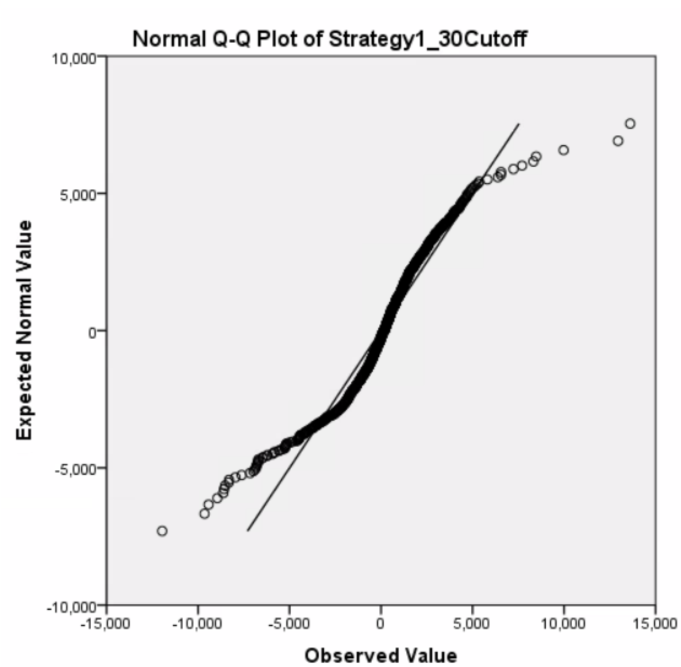
In the context of this thesis, the t-test was run on the returns of the portfolio for each strategy. I did this because it was a uniform way of comparing the significance of each strategy. One disadvantage of doing it this way is that by repeatedly conducting t-tests, the probability of type I error, that is, incorrectly rejecting the null hypothesis, increases. For that reason, while I would caution against using just the p-values to verify significance, I believe if all three factors point to a similar result, then I am more likely to believe that the returns generated by that strategy share are significant. For example, if the t-test showed that one of the strategies was statistically significant at a 5% level, but that strategy had very little cumulative profits and a very low Sharpe ratio, then I would be less confident in the results of that t-test. However, if it happened to be the case that the p-value, the cumulative profits, and the Sharpe ratio all indicated that the strategy performed well, then I would be more inclined to believe that there is merit in that strategy.

We will now assess each assumption. The first assumption that the data is continuous is met, since the return of a stock can be any real number. The second assumption is also met. This is because I am assuming that the population is the return of the Russell 1000. Since each strategy that I am testing is a subset of the Russell 1000, chosen "at random" by the strength of the sentiment signal, it can be assumed that the return data for each strategy is a random sample from the population. Finally, the third assumption of the data being *roughly* normally distributed is shown pictorially in the following figures in both histogram and Q-Q plot form. While the Q-Q plots indicate that the data might be a bit more long tailed than a standard normal distribution, the standard deviation of the portfolio's returns is also higher because of these outliers, and as a result, the t-statistic is smaller. As a result of the smaller t-statistic, it is harder to get a statistically significant result. Thus, even though this assumption of normality is tentatively met, the outcome of the t-test is actually robust to the longer tails of the portfolios.

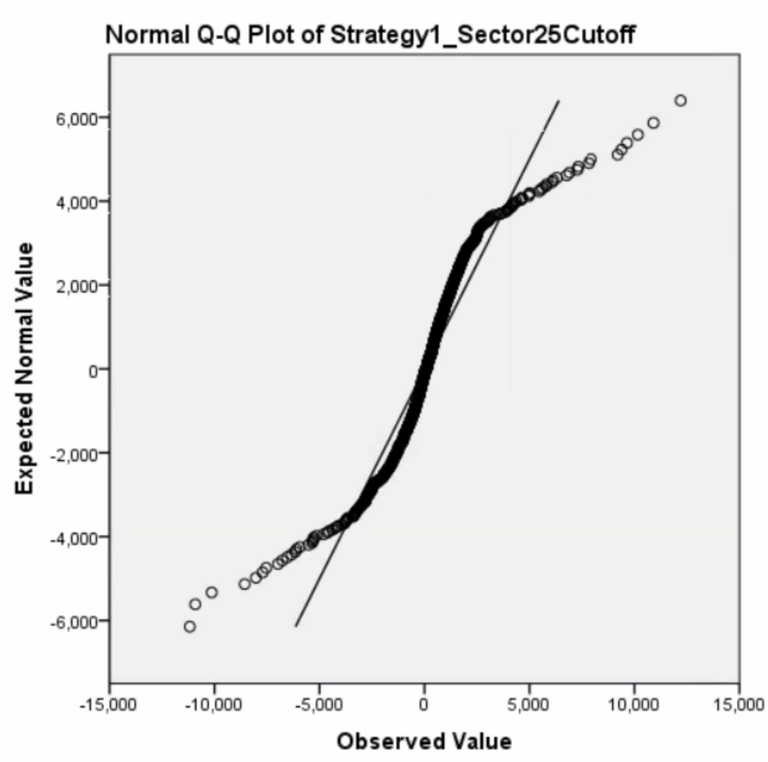

Mean = 118.74242694
Std. Dev. = 2253.82016685
N = 1,258
Frequency
300
200
100
0
-15000
-10000
-5000
0
5000
10000
15000
Strategy1_30Cutoff


Strategy1_Sector25Cutoff
Mean = 126.61028770
Std. Dev. = 1904.93909464
N = 1,258
Frequency
400
300
200
100
0
-15000
-10000
-5000
0
5000
10000
15000
Strategy1_Sector25Cutoff

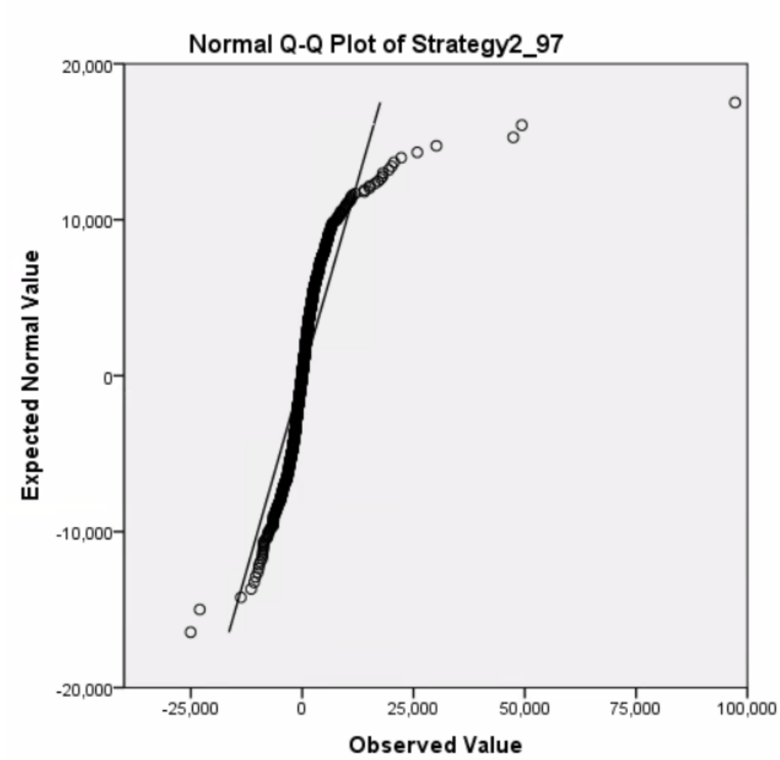




This q-q plot suggests that the data is longer tailed than a standard normal distribution.



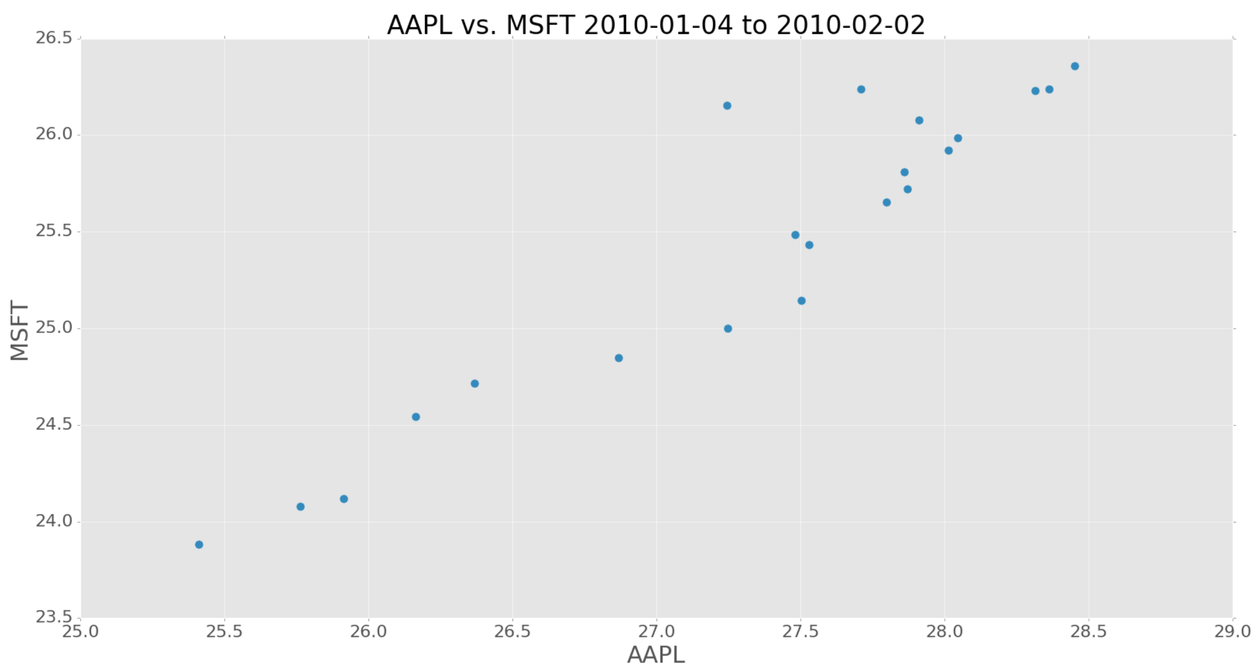
Relative to the first q-q plot depicting the portfolio pnl without sector grouping, this q-q plot suggests that the tails are even longer. This is seen by the number of points to the left and to the right of the 45-degree line.



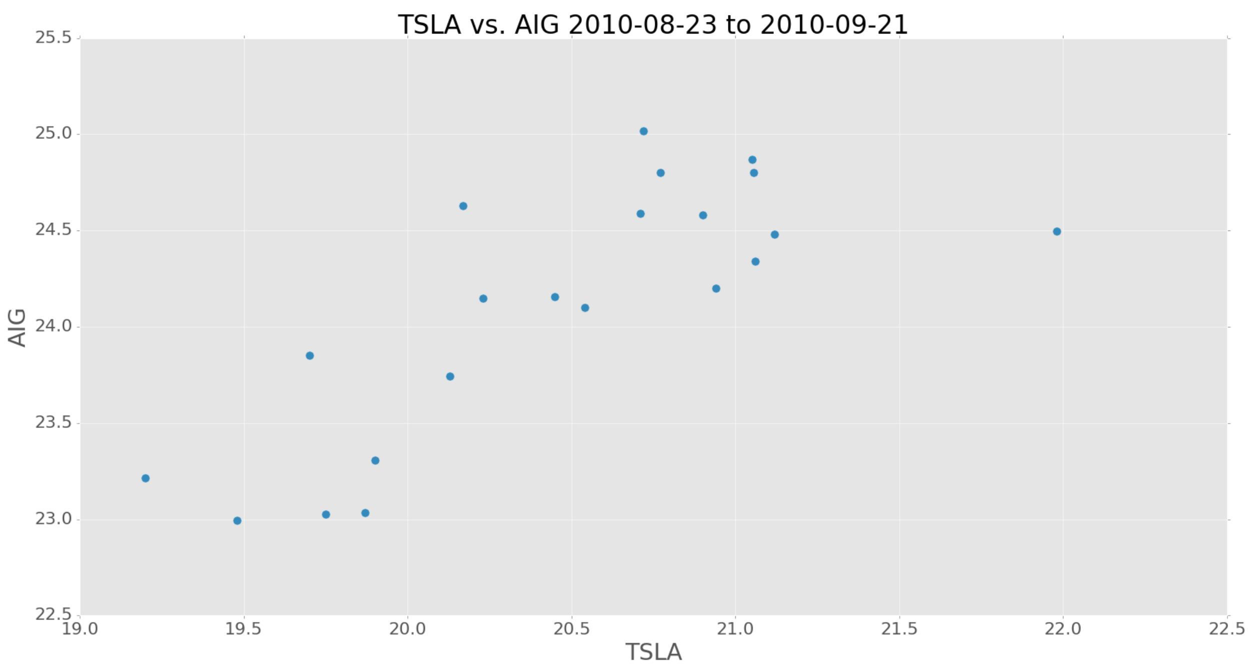
In this case, the right hand side of the pnl distribution is longer than expected for perfectly normally distributed data. As discussed earlier, these longer tails will cause the standard deviations to be large, leading to a smaller t-statistic. With a smaller t-statistic, it is more difficult to conclude statistical significance. Thus, I believe that the t-test is indeed a robust way of testing the hypothesis that the mean return for each portfolio is statistically significantly higher than zero.

Appendix E

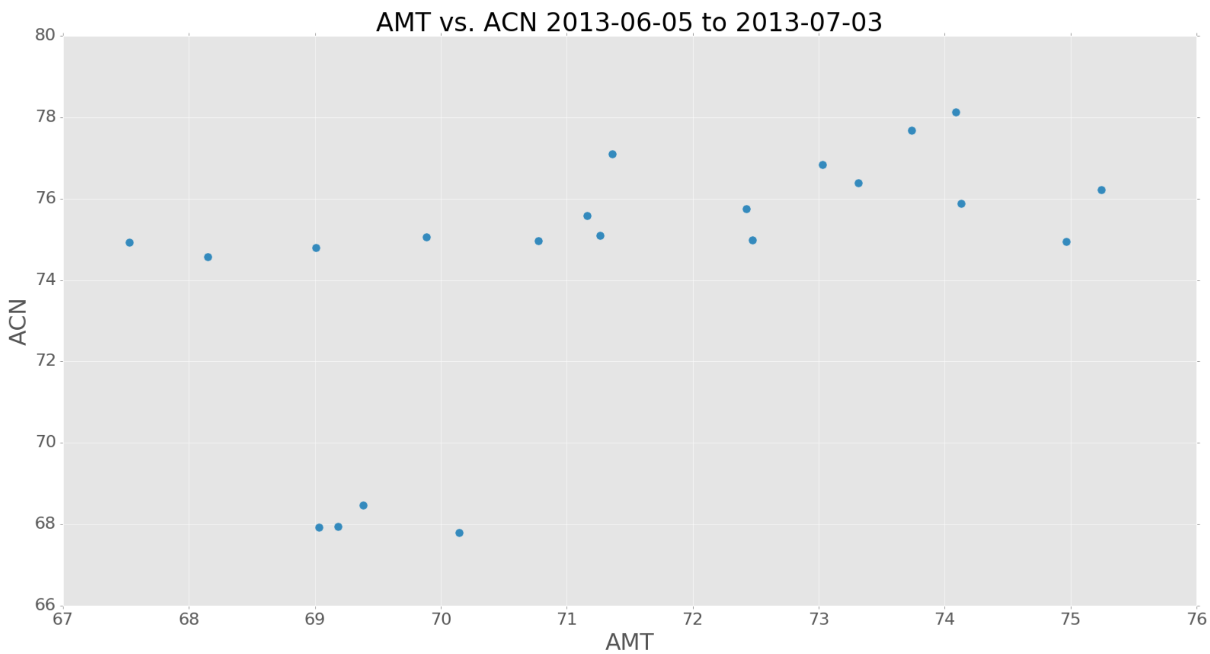
A large part of the second pairs trading strategy relies on 21-day pairwise correlations of the stocks in the Russell 1000. While I discuss the practical reasons for using a 21-day pairwise correlation, the following few figures demonstrate the pros and cons of using this statistic. All of the figures below have a correlation of at least 0.90. In the first figure, we notice that the relationship between AAPL and MSFT stock is very linear between January 4th, 2010 to February 2nd, 2010. There appear to be no extreme outliers in the scatterplot that would make the correlation between the two stocks appear to be stronger than it actually is. This is an ideal correlation.



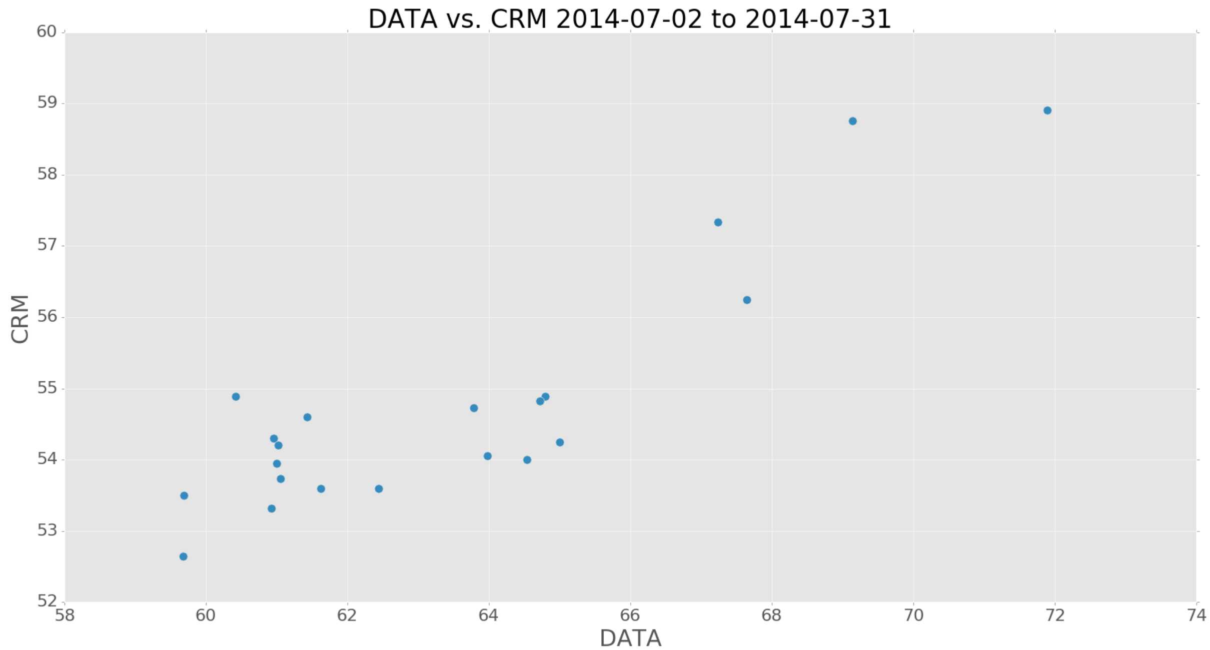
The next graph shows the relationship between TSLA and AIG stock between August 23rd, 2010 and September 21st, 2010. While this relationship is not like as perfectly aligned as the AAPL MSFT relationship, it seems that overall there is a fairly linear relationship between the two stocks. There is, however, an outlier that causes the regression line's slope to decrease. As a result, the residuals increase and the correlation decreases. Perhaps without the outlier, the correlation of TSLA and AIG stock would be above 0.95, but with the point, the correlation is lower and thus, TSLA and AIG is not traded in the strategy that has a 0.97 cutoff.



Another example of how the correlation is affected can be seen in the relationship between AMT and ACN between June 5th, 2013 and July 3rd, 2013. It is clear that there is some grouping going on in the two stocks' prices. If one were to ignore the four outliers at the bottom left portion of the scatterplot, then it looks like there is a very weak relationship between the two stocks. However, once you add those points back in, the slope of the regression increases, and ultimately increases the correlation between the two stocks.



One final example of how the correlation is affected can be seen in the relationship between DATA and CRM between July 2nd, 2014 and July 31st, 2014. Again, it seems like there is some grouping going on in the two stocks' prices, but after a certain price point, the clusters become scarce. Looking at the very top right corner point in the graph below, because that data point is on the same trajectory as the grouped points on the bottom left corner of the graph, the value of the correlation increases because the residuals decrease. Though this example is not as extreme due to the points in between, one can imagine a scenario in which there is an obvious clustering and one outlier point along the linear regression line. In that example, that outlier is especially problematic because without that one point, the correlation between the stocks may not have been very strong, but the strategy would have traded the pair anyway.



While I recognize these setbacks to using a statistic that is influenced by outlier points, I still believe that 21-days is a good time frame to use. The first reason for this is that I am assuming that since I have several thousand 21-day correlations that the resulting trades that I consider for the strategy will have a majority of good correlations to base my analysis on. The second reason I am comfortable using this statistic is that I am using the correlations as a binary signal to filter out the stocks that have a high correlation and the ones that do not. If it were the case that I was trying to predict a particular stock's exact return based on its historical correlation with another stock, I would be more careful in selecting the timeframe for each correlation. Finally, even after I have the list of stocks that are highly correlated, I am not using all of them in my portfolio. I am only selecting the ones that have conflicting signals. Hence, with all of these filters in place, I am hoping to have a robust enough dataset to conduct my analysis on.

Appendix F

Below are some lines of code that were central to my thesis[3].

```
#This function calculates the pairwise correlation between the 1000 stocks over the 6-year time period. Each day's data is appended to a csv file, and each day (1000 stocks correlations) takes 7 seconds to run.

import time
import pandas as pd
import dask.dataframe as dd
from dask.dataframe import rolling

df = dd.read_csv("unstacked.csv")
data = pd.read_csv("unstacked.csv")

n=0
while n!=1491:
  new = df.loc[n:n+20]
  a = new.corr().compute()
  flat = a.stack().reset_index()
  flat.columns = ['Ticker 1', 'Ticker 2', 'Correlation']
  flat = flat.loc[flat.Correlation < 1, ['Ticker 1', 'Ticker 2', 'Correlation']]
  flat = flat.loc[flat['Ticker 1'] < flat['Ticker 2']]
  flat["Date"] = data.loc[n+21]["Date"]
  flat = flat.loc[flat["Correlation"] >= 0.90]
    flat.to_csv("cor_matrices/dailycor_21day.csv", index=False, header=False,
    mode='a')
    n+=1
```

```
import pandas as pd
import numpy as np
import time
import matplotlib.pyplot as plt
import matplotlib
matplotlib.style.use('ggplot')
```

[3] The full code (over 1,000 lines) will be made available to the reader upon request. If interested, the reader should email rpatel3@drew.edu.

```
#STRATEGY 1
def returndf(cutoff):
    data = pd.read_csv("psysig_insample.csv")

    #TO SELECT THE TIMEFRAME OF THE DATA
    data = data.loc[data["Date"] > "2010-01-01"]
    data = data.loc[data["Date"] < "2011-01-01"]
    data["Date"] = pd.to_datetime(data["Date"])

    colnames = list(data.reset_index().columns.values)
    ret = pd.DataFrame(columns=colnames)

    #RANKS THE DATA AND ASSIGNS SIGNALS BASED ON CUTOFF
    data["ranksector"] =
data.groupby(["Date","Sector"])['Signal_1_psych'].rank(ascending=False)
    data['rank_pctsector'] =
data.groupby(["Date","Sector"])['Signal_1_psych'].rank(ascending=False,
pct=True)
    data['BuySell'] = np.where(data["rank_pctsector"] <= cutoff, 1,
np.where(data["rank_pctsector"] >= (1-cutoff), -1, 0))
    data = data.reset_index()


    #SEPERATES DATA BY SIGNAL
    a = pd.DataFrame(data.loc[data["BuySell"] == 1])
    b = pd.DataFrame(data.loc[data["BuySell"] == -1])

    #CREATES A DATAFRAME OF ALL OF THE BUYS AND SELLS
    total = pd.DataFrame()
    total['buys'] = a.groupby(["Date"])["BuySell"].sum()
    total['sells'] = b.groupby(['Date'])["BuySell"].sum()*-1
    total['total_trades'] = total['buys']+total['sells']
    total.reset_index(inplace=True)
    total.dropna(inplace=True)

    ###MERGES ALL THE DATAFRAMES, CALCULATES PNL PER TRADE
    ret = ret.merge(a, how='outer')
    ret = ret.merge(b, how="outer")
    ret = ret.merge(total, how='outer')
    ret['notional'] = 1000000
    ret['amountpertrade'] = ret['notional']/(ret['total_trades'])
    ret['pnl0'] = ret['amountpertrade']*ret['return']*ret['BuySell']

    ret = ret.groupby("Date").sum()
    listc = ["index", "Unnamed: 0"]
    ret = ret.drop(listc, axis=1)
```

```
#TRANSACTION COSTS
ret['pnl1'] = ret['pnl0'] - (0.0001 * ret['notional'] * 2)
ret['pnl2'] = ret['pnl0'] - (0.0002 * ret['notional'] * 2)
ret['pnl3'] = ret['pnl0'] - (0.0003 * ret['notional'] * 2)
ret['pnl4'] = ret['pnl0'] - (0.0004 * ret['notional'] * 2)
ret['pnl5'] = ret['pnl0'] - (0.0005 * ret['notional'] * 2)

#CUMULATIVE PNL
ret['0bps'] = ret['pnl0'].cumsum()
ret['1bps'] = ret['pnl1'].cumsum()
ret['2bps'] = ret['pnl2'].cumsum()
ret['3bps'] = ret['pnl3'].cumsum()
ret['4bps'] = ret['pnl4'].cumsum()
ret['5bps'] = ret['pnl5'].cumsum()
return ret
```
