

Detecting Insider Trading

Instructor: Prof. Kay Giesecke

TA: Benjamin Armbruster

Group Members:

Manabu Kishimoto

Xu Tian

Li Xu

Abstract

In this project, we first find the salient statistical patterns of insider trading

- 1: Call-put imbalance is large.
- 2: Total option volume is high.
- 3: Slightly in-the-money or out-of-the-money option is preferred (by insiders).
- 4: Near-term option is preferred (by insiders).
- 5: Near-term implied volatility is high.

Then we develop methodologies to detect possible insider trading events using information from the market. Then, we design trading strategies to profit from the detection and back-test these strategies. Finally we developed an automatic processing program using PERL language. We coded the criteria into the program and test it in three databases. Our results show that the methodologies are very successful.

Contents

1. Introduction
 2. Previous work
 3. Litigation Case Studies
 4. Detecting Strategy
 5. Automation and Optimization
 6. Performance Evaluation
 7. Future Work
 8. Conclusion
- References

1. Introduction

Insider trading occurs when individuals with “inside” information gained from their position in the company illegally use this knowledge to trade and make money off uninformed investors. Insider trading can be seen in both stock and options markets. Insider trading is a crime; it hurts market makers who have to give bid/ask quotes for options, for instance; thus, it is prohibited by law. However, if one is able to detect insider trading before the news release, excess returns can be legally generated.

In this project, we first find the salient statistical patterns of insider trading and develop methodologies to detect possible insider trading events using information from the market. Then, we design trading strategies to profit from the detection and back-test these strategies. Although insider trading could occur in both stock and options markets, we choose the options markets to analyze, because for insider traders, options markets are more attractive than stock markets because they are leveraged. For example, McMillan [3] points out that to exploit their information, insider traders would be most likely to buy stock options because those markets results in the most highly-leveraged gains. In addition, options markets are more informative; options market data provide many clues about insider trading. Stock markets are too noisy and yield only price and volume data.

We also focus on *good news scenarios*, such as M&As, because they are typical occurrences that will affect a stock’s price predictably. However, our methodology could also be applied to the *bad news scenarios* with slight modifications.

2. Previous work

There are a lot of papers published in this research area. Based on these papers and our case study, we design our criterions. Donoho [1] developed a scoring system in his paper which was inspired by McMillian’s hypothesis in [3] that people with inside information leave evidence in option trading data that might predict news. McMillian’s method was mostly manual and required a large amount of human intuition and manual analysis. Donoho use data mining technology which is a good fit for this problem because it is able to automatically pick up faint signals from noisy data. It is able to discover correlations that experts may not have been aware of. Data mining has also proven useful at finding trends in human behavior, and markets are the synthesis of the behaviors of many individuals (some with inside information, most without). Lipkin [2] analyzed the statistic patter of the take over case, he pointed out that implied volatility can be used as an indicator of insider trading in take over case and discussed several examples.

Cao [4] examines the information embedded in both the stock and option markets prior to takeover announcements. In time-series regressions they find that during the benchmark period, lagged stock volume imbalances are more informative of next-day returns and that lagged call volume imbalances are not related to returns. In the pre-announcement period, option imbalances become significant predictors of next-day stock returns. They find that this strong relation between pre-announcement call imbalances and returns is concentrated in successful takeover targets. They compare firms with and without options and find that when both options and stocks are available for trading, calls displace information in the pre-announcement period that might otherwise be reflected in stock imbalances. In the cross-sectional analysis, they find that large pre-announcement increases in call imbalances are associated with higher takeover premiums, while pre-announcement increases in share imbalances are not related to future returns. Thus, ahead of major announcements the options market plays an important role in information revelation, whereas during normal market times the stock market is the primary place of price discovery. Among option characteristics, short-term OTM calls (which are also the most profitable) experience the largest increase in volume and buyer-initiated volume. They find that post-announcement trading activity does not predict the future success or failure of a deal. To examine the scope of their conclusions, they have included in out-of-sample exercise all firms that had options traded on the CBOE. Extremely high call-volume trigger rules lead to significantly higher returns. On the other hand, for signals based on share volume, the higher the volume threshold, the lower the average returns. An implication of these results is that the options market can be particularly informative ahead of extreme material events, while the stock market may be more suitable for disseminating ordinary information flow.

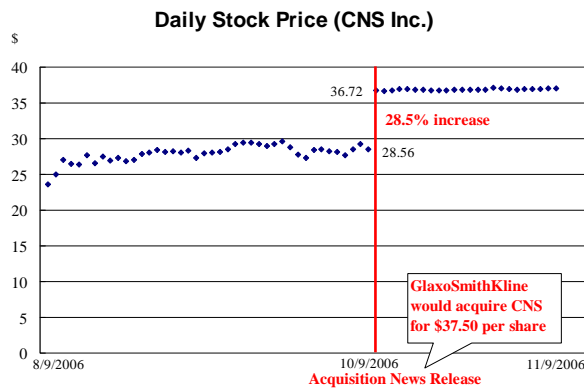
Pan [5] also presented strong evidence that option trading volume contains information about future stock prices. Taking advantage of a unique data set, they construct put-call ratios from option volume initiated by buyers to open new positions. Moreover, they were able to partition the signals obtained from option volume into various components and to investigate the process of price adjustment at a greater depth than previous empirical studies. Their findings indicate that it takes several weeks for stock prices to adjust fully to the information embedded in option volume. The main economic source of this predictability, however, does not appear to be market inefficiency. Rather than a disconnection between the stock and the option markets, the predictability that they document appears to be driven by valuable nonpublic information which traders bring to the option market. They found that, in accordance with the theoretical models, the predictability is increasing in the concentration of informed traders and the leverage of option contracts. Applying the same predictive analysis to the index option market, however, yielded no evidence of informed trading. This is indeed consistent with the view that informed traders tend to possess firm-specific rather than market-wide information.

3. Litigation Case Studies

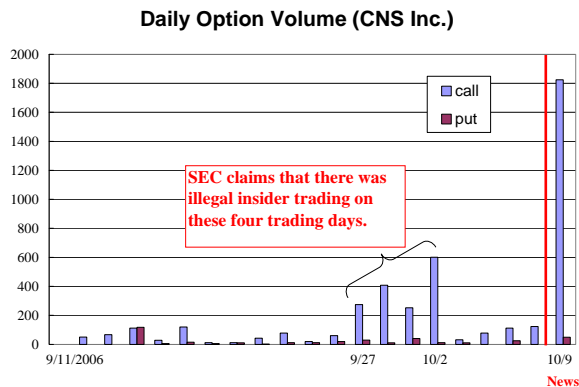
In the U.S. Securities and Exchange Commission (SEC) website, “Litigation releases – Federal Court Actions” [6] provides many litigation cases that focus on insider trading. In this section, we examine two typical litigation cases in options markets: the CNS Inc. case and the InVision Technologies case, to find salient statistical patterns of insider trading in options markets.

CNS Inc.

On October 9, 2006, there was a news release that GlaxoSmithKline would acquire CNS Inc. for \$37.50 per share. On that day, the stock price jumped up by 28.5%.

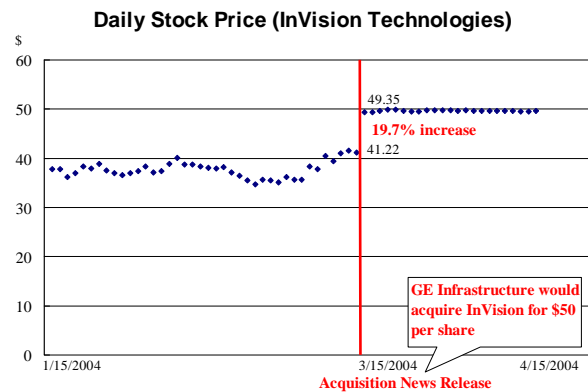


SEC claims that there was illegal insider trading from September 27 to October 2, 2006. During that period, call option volume jumped up and stayed high.

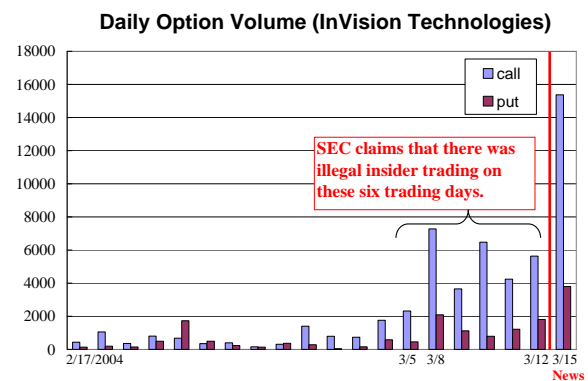


InVision Technologies

On March 15, 2004, there was a news release that GE Infrastructure would acquire InVision Technologies for \$50 per share. On that day, the stock price jumped up by 19.7%.



SEC claims that there was illegal insider trading from March 5 to March 12, 2004. During that period, call option volume jumped up and stayed high.

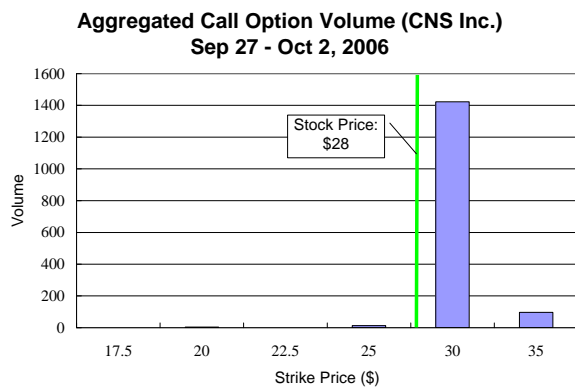


Salient statistical patterns 1: Call-put imbalance is large.

Salient statistical patterns 2: Total option Volume is high.

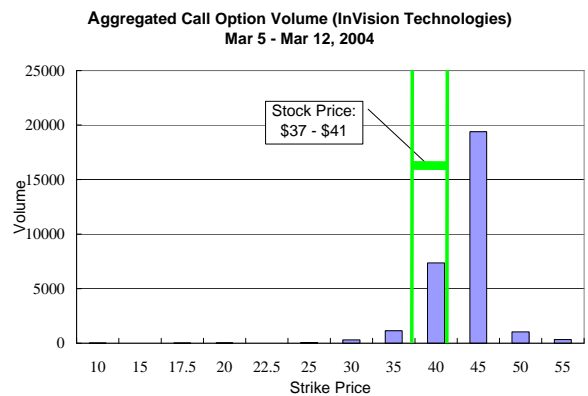
CNS Inc.

We analyzed the call option volume during four insider trading days. The strike price of \$30 was particularly preferred. During that period, the stock price was around \$28, which means that the call option with a strike price of \$30 is Out-of-the-money (OTM).



InVision Technologies

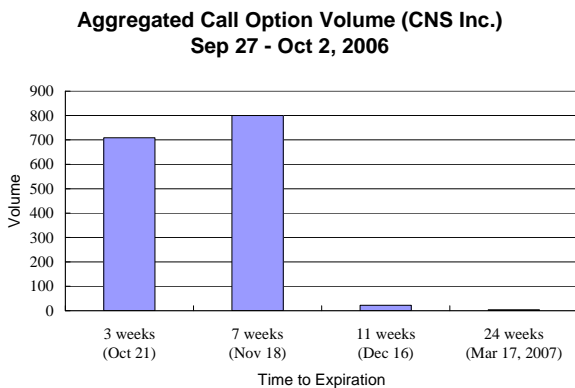
We analyzed the call option volume during six insider trading days. The strike prices of \$40 and \$45 were particularly preferred. During that period, the stock price was around \$37 - \$41 which means that the call option with a strike price of \$40 is slightly in-the-money (ITM) and that of \$45 is Out-of-the-money (OTM).



Salient statistical patterns 3: Slightly in-the-money or out-of-the-money option is preferred (by insiders).

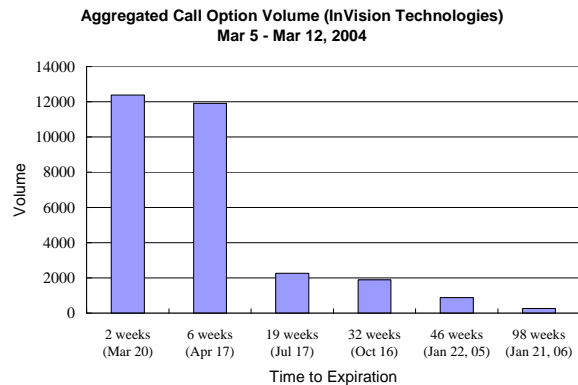
CNS Inc.

The expiration date, October 21 (3 weeks later) and November 18 (7 weeks later) were particularly preferred.



InVision Technologies

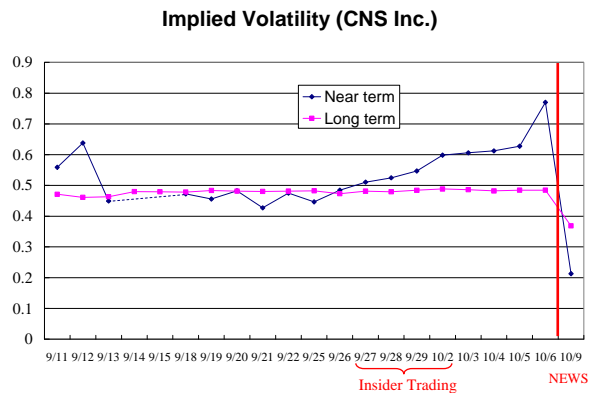
The expiration date, March 20 (2 weeks later) and April 17 (6 weeks later) were particularly preferred.



Salient statistical patterns 4: Near-term option is preferred (by insiders).

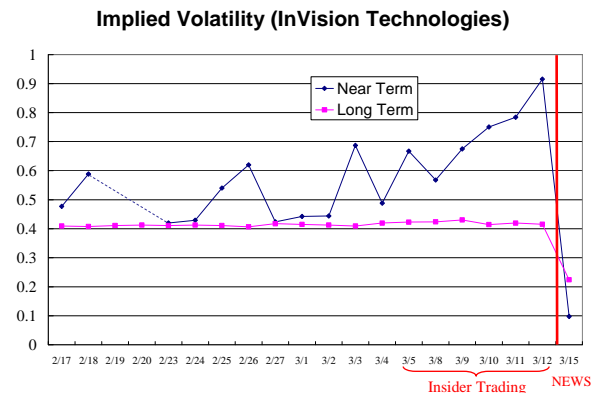
CNS Inc.

Regarding the implied volatility, near-term implied volatility increased before the news release; on the other hand, long-term implied volatility remained the same.



InVision Technologies

Regarding the implied volatility, near-term implied volatility increased before the news release; on the other hand, long-term implied volatility remained the same.



Salient statistical patterns 5: Near-term implied volatility is high.

In conclusion, after investigating these two litigation cases, we found that insider trading in options markets have the following salient statistical patterns.

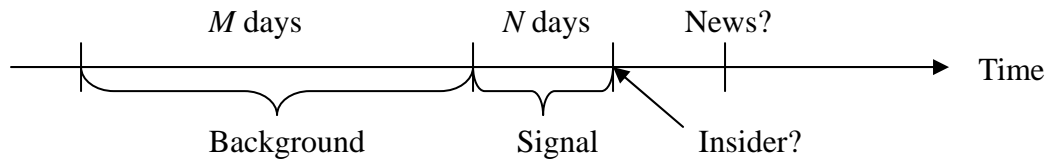
(Salient statistical patterns of insider trading)

- 1: Call-put imbalance is large.
- 2: Total option volume is high.
- 3: Slightly in-the-money or out-of-the-money option is preferred (by insiders).
- 4: Near-term option is preferred (by insiders).
- 5: Near-term implied volatility is high.

These are the “fingerprints” of insider traders, in the good news case. It makes sense because insider traders with confidential good news, such as M&As, tend to buy call options. Such call options should not be deep-in-the-money because insider traders expect the stock price to rise, even though they do not know by how much. In addition, insider traders prefer near-term call options because the news is released sooner; near-term call options are much cheaper than long-term call options.

4. Detecting Strategy

According to the statistical patterns we found, we build a detecting strategy. First, we use two moving windows: consecutive M trading days as background data, and the subsequent consecutive N trading days as a signal. M and N are the parameters we will determine later.



Next, since we are interested in slightly in-the-money or out-of-the-money, and near-term call options, we filter the data; we focus on the data which satisfy the following two conditions:

(Strike Price Filter Criterion)

$$\frac{\text{Stock price} - \text{Strike price}}{\text{Stock price}} < r$$

(Expiration Date Filter Criterion)

$$\text{Expiration date} - \text{Current date} < n \text{ days}$$

Parameters r and n will be determined later. Then, we apply the following two criteria:

(Call Ratio Criterion)

$$\frac{\text{Call volume}}{\text{Call volume} + \text{Put volume}} > \alpha \%$$

(Total Volume Criterion)

$$\frac{\text{The daily average of the volume in signal data}}{\text{The daily average of the volume in background data}} > R$$

Again, parameters α and R will be determined later. Theoretically, it is also possible to formulate the Implied Volatility criterion, but due to the lack of data (many of the near-term implied volatility data are missing in the options database), we could not apply the criterion.

5. Automation and Optimization

Statistically, insider events exhibit salient patterns as mentioned in the previous section. In order to test whether these criteria will help detect the insider trading events in a large portfolio, we developed an automatic processing program using PERL language. We coded the criteria into the program and test it in three databases.

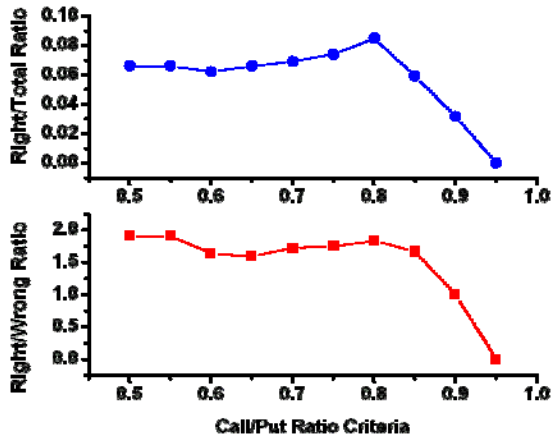
The first database, which we called “litigation database”, has three securities. They are Dow Jones Company (DJ), CNS Inc. (CNXS), and InVision Technologies (INVN). We know for sure the exact time of the illegal insider trading for these securities according to SEC’s litigation filing. This database is great for us to initialize the guess for insider trading patterns. The second database, we called “training database”, has 99 securities. This database is full of corporate events such as earning announcement, upgraded/downgrade, and most importantly, merger and acquisition events. This database potentially has more insider events than the market portfolio. Thus it is great for training the model and optimizing the key parameter in detection criteria. The third database, referred as “testing database” has 3068 securities. It is essentially the market portfolio except some securities which we don’t have the option data. After we optimized the parameters by the training database, we fixed the values, and test it in testing database to get the final output. The table below shows a summary of these three databases and number of events that our program determined as insider trading events:

	Litigation Database	Training Database	Testing Database
	CNXS, DJ, INVN	Event Database	2007 First Half OptionMetrics Database
# of Tickers	3	99	3068
Year	2002/01-2004/06 2005/01-2007/06	2005/01/01-2007/06/30	2007/01/01-2007/06/30
# of events	15	474	1902

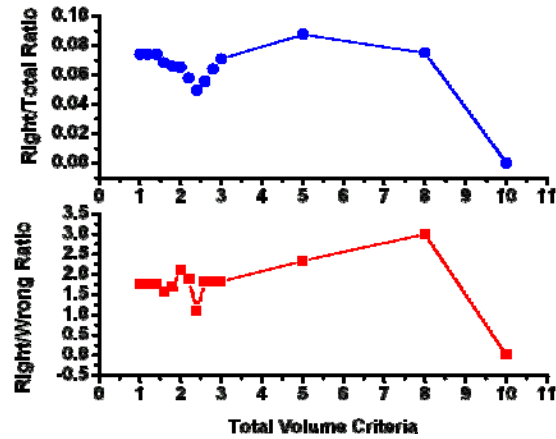
In order to get the optimized value for detection criteria, we changed only one parameter at a time, and run the program to find out the Right/Total ratio and Right/Wrong ratio. Here, a “right detection” is defined as those that after the insider events are detected, the stock do jump more than 10%. And correspondingly, a “wrong detection” is defined as those that after the insider events are detected, the stock actually tumbles more than -10%. The Right/Total ratio tells us the

ability of the program to detect insider events rather than normal trading noise. On the other hand, the Right/Wrong ratio tells us the program's ability to avoid serious loss. The following figures show how changing the detection criteria would change the Right/Total ratio and Right/Wrong ratio.

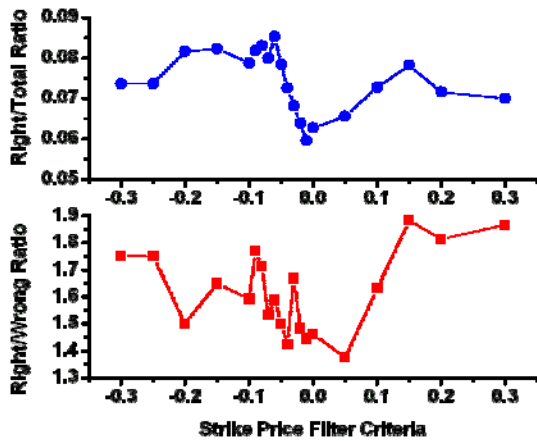
Call/Put criteria:



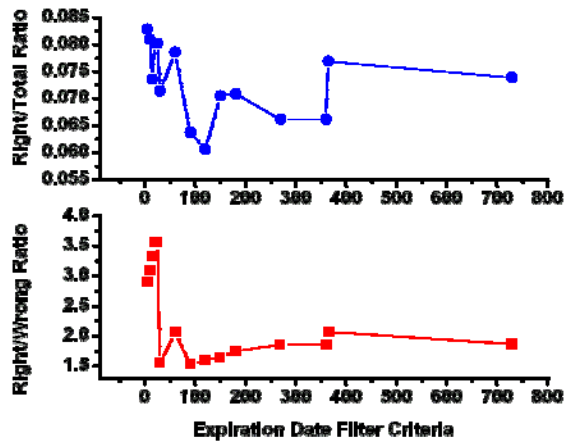
Total Volume criteria:



Striking price filter:



Expiration data filter:



One may think that choosing the value that gives both high Right/Wrong ratio and Right/Total ratio will be good enough. However, it is important to note that the sensitivity analysis has to be considered. For example, when choosing the call/put ratio criterion value, 0.8 gives the highest value of Right/Total. However, as shown in the figure, the Right/Total ratio decreased significantly once the call/put ratio goes beyond 0.8. So it may be safer to use other values such as 0.75 to enhance the detection capability, because these criteria will be used towards other general databases also. The following is the tentative values we selected for detection. (Note: these values are not the absolute choice. Small variation is allowed to fit different database and trading strategy.)

Optimized value:

	Background Length	Signal length	Call/Put Ratio	Total Volume Ratio	Strike Price Filter	Expiration Date Filter
Value	100	10	0.75	1	0.15	6months

Ideally, optimization should be done in the whole parameter space. That way, we can find the global maxima in this 6 dimensional space, and do sensitivity analysis on that 6-dimension space, and choose proper value according to specific trading strategy. Due to time constraint, here we only did a local search of the optimum value.

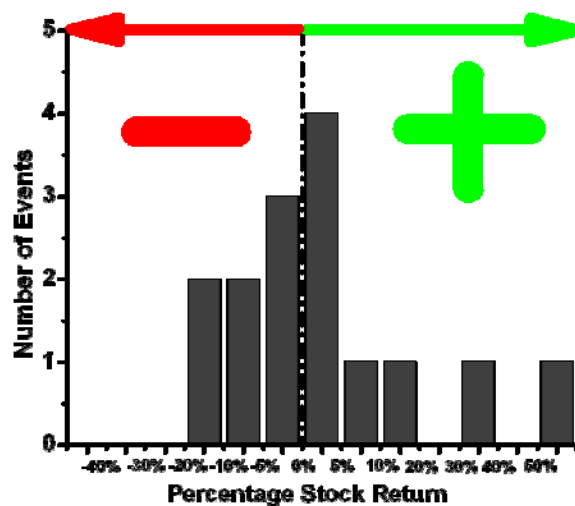
6. Performance Evaluation

To evaluate how well the program can predict insider trading, we consider the following four benchmarks.

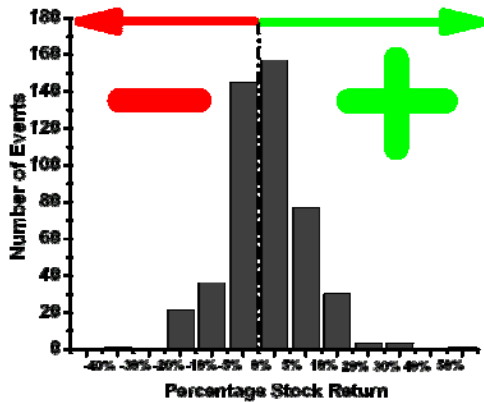
Benchmark#1: Histogram of Stock Return

We examine how the stock performs in a 10 trading day period when an insider event is detected by the program. For example, in the following histogram, we can see in the litigation database, one event gives us more than 50% jump in stock market, one event gives 30%-40% jump in the stock market, one event gives 10%-20% jump, one gives 5%-10% jump, and 4 events give 0%-5% jump, etc.

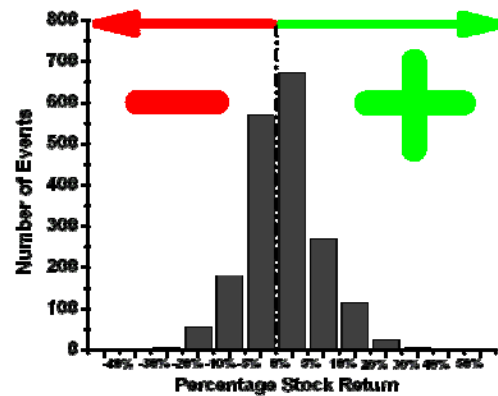
Litigation Database:



Training Database:



Testing Database:



Histogram on training database and testing database show more events on the positive side than negative side, this is a clear evidence of ability to detected insider trading.

Benchmark #2: Percentage Return of Non-leveraged Simple Trading Strategy

For this benchmark, we consider the following simple trading strategy:

First, we allocate \$1 in every security in the database. Then we run the program everyday just before the market closes. If the program tells us that there is insider trading on that day, then we use the balance of the fund to buy the stock, and sell the stock after 10 trading days; otherwise, we do nothing and wait for the signal on next day. At the end of the period, we calculate the total annualized return for all the funds allocated over the entire database, and compare this return with buy-and-hold return.

The results on this simple strategy are summarized in the following table:

	Litigation Database	Training Database	Testing Database
NSTS Return	+15%	+5.7%	+7.47%
Buy and hold Return	+39% (Acquisition rich)	+28% (Acquisition rich)	+2.82%

Note that buy-and-hold return in litigation database and training database are very high. This is because these databases have more acquisition events than average market portfolio. And a buy-and-hold strategy certainly captures all the acquisition events, thus gives unbeatable returns.

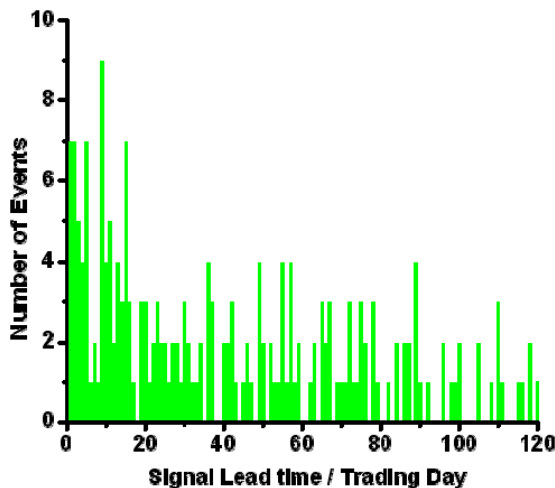
In the testing database, we beat the market return by around 5%. Remember this strategy is not leveraged! And it is not even best non-leveraged strategy! For example, if there is only one insider event for a security, then after capturing the event, that \$1 will just wait until the end of the period without reinvesting. Ideally, if the events don't overlap each other in time, we could trade all of them and make a huge profit.

Benchmark #3: Histogram of Signal Lead Time

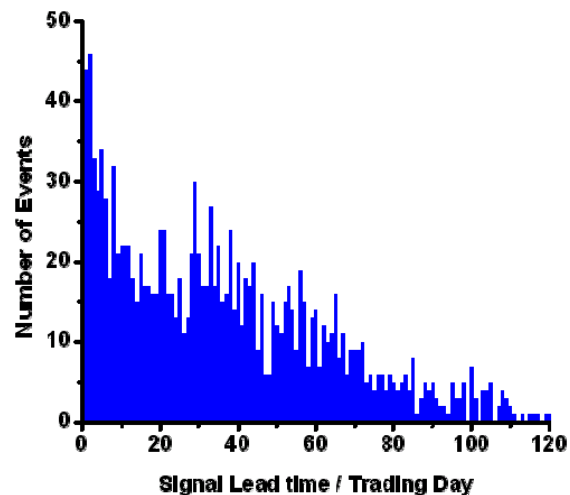
There is a certain lead time between the detection of an event and actual jump ($\geq 5\%$) in the stock. The lead time varies from event to event, and from security to security. It is worthwhile to examine the distribution of the lead time. A good detection method will have well defined lead time. And the maximum of the lead time, limited by the short term nature of insider trading, should be about 6 months.

In the following figures, we can see the lead time distribution for training database and testing database. In the testing database, it is clearer that most probable lead time is within 5 trading days, and long lead time of around 6 months (120 trading days) is not likely to happen.

Training Database:



Testing Database:



Benchmark #4: Prediction Errors

A final benchmark is about prediction errors. In this part, we will answer the questions such as

1. If the program predicts insider events, how likely it will become true?
2. If the stock jumps more than 5% (potential insider trading), how likely our program can predict this?

The answers to these questions are summarized in the following tables:

Litigation Database:

# of events		Stock jump more than 5%?	
		Yes	No
Detected?	Yes	4	11
	No	55	

Training Database:

# of events		Stock jump more than 5%?	
		Yes	No
Detected?	Yes	114	360
	No	828	

Testing Database:

# of events		Stock jump more than 5%?	
		Yes	No
Detected?	Yes	417	1485
	No	18108	

To interpret these tables, we give an example. For testing database, if the program predicts insider trading, there are 417 times the prediction is right, and 1485 times the prediction is wrong. And if the stock do jump more than 5% relative to previous day's price, there are 307 times we can predict it, and the other 18108 times, we are not able to predict. This may due to the reason that most jumps in stock price don't relate to insider events at all.

7. Future Work

There are many directions the project could be taken. We did not include implied volatility in our detecting criterions. The main reason is that our database has many blanks for implied volatility, so we can not apply this dataset into our project. Moreover, if we can obtain the dataset which was used by Pan [5] in their paper, we can construct more robust detecting criterions. Based on

the finds of their paper, we can reasonably expect more promising detecting results. Another important problem is how to take advantage of the detecting results. In our project, we only use very simple trading rule to test our detecting strategy. Once we received signal, we can use more complicated trading strategy to obtain more profit.

8. Conclusion

In this project, we first find the salient statistical patterns of insider trading which are

- 1: Call-put imbalance is large.
- 2: Total option volume is high.
- 3: Slightly in-the-money or out-of-the-money option is preferred (by insiders).
- 4: Near-term option is preferred (by insiders).
- 5: Near-term implied volatility is high.

Then we develop a moving window strategy to detecting the insider trading. First, we filter the dataset and only keep the slightly in-the-money or out-of-the-money, and near-term call options, next we apply call ratio criterion and total volume criterion these filtered dataset. Finally, we developed an automatic processing program using PERL language. We coded the criteria into the program and test it in three databases.

References

- [1] Donoho, S., “Early Detection of Insider Trading in Option Markets,” 2004.
- [2] Lipkin, M., “Sherlock Trader: Take-Over Sleuth,” Talk at Stanford, Dec. 1, 2006.
- [3] McMillan, L., *McMillan on Options*, 2nd ed., John Wiley & Sons, 2004.
- [4] Cao, C., Z. Chen, and J. M. Griffin, 2005, “Informational Content of Option Volume Prior to Takeovers,” *Journal of Business*, 78, 1073–1109.
- [5] Pan, J., and A. M. Poteshman. 2006. The Information in Option Volume for Stock Prices. *Review of Financial Studies* 19:871–908.
- [6] SEC website (litigation): <http://www.sec.gov/litigation.shtml>
 - CNS Inc. case (Litigation Release No. 19867):
<http://www.sec.gov/litigation/litreleases/2006/lr19867.htm>
 - InVision Technologies case (Litigation Release No. 18627):
<http://www.sec.gov/litigation/litreleases/lr18627.htm>