Sri Lankan Journal of Applied Statistics, Vol.10, 2009, p.237-250

Estimating Privately held Information Using Trading Volume: Predicting the Future Price Movements to Reduce Investor Risks

C. FONSEKA AND L. LIYANAGE

School of Computing and Mathematics, University of Western Sydney, Locked Bag 1797 Penrith South DC NSW 1797, Australia

ABSTRACT

Traders involved in the stock market use information available to them to make decisions on trading. Two kinds of information are available to traders – public and private. The market reacts and adjusts itself almost immediately to the publicly available market information. A number of papers have studied the impact of the trading volume on the predictability of future price movements of a stock. A large number of these use traditional ARCH and GARCH models to forecast future returns based on lagged returns and autocorrelation. In this context this paper seeks to learn by investigating empirically, a novel perspective of the basic economic question of how market information is incorporated into the price of a stock. The data has in it hidden patterns of information generated by the behaviour of those who hold private information. The algorithm discussed in this paper reveals the existence of private information and makes it usable by a wider audience. The predictability established by our analysis correspond to the gradual process the prices incorporating to the privately held information available to traders. We suggest evidence that under certain conditions the volume of stock traded provides information about the likely future movements in the underlying stock prices. We show that stocks having high levels of volatility in the volume traded together with a marginal increase in stock price are more likely to go up in price further provided that the volume shows a sharp increase. Our analysis is capable of locating potential stocks that are more likely to have privately held information.

Keywords: Data mining, Predictive modelling, Stock market, Stock Market strategy, Stock Market algorithm, Trading volume

INTRODUCTION

Stock market investors whose primary objective is to gain high yield for their investments are always looking for a superior system which will improve the predictability of the future value of a share of a listed company. Stock market prediction is the attempt of estimating the future value of a financial instrument

traded on a stock exchange. One of the best known concepts in finance is that markets are efficient. An efficient market adjusts prices without delay to reflect all available information. This is a concept based on that stocks take a random and unpredictable path. The random walk theory states that it's impossible to outperform the market without assuming additional risk. Many financial economists now contend that stocks do maintain price trends over time - in other words, that it is possible to outperform the market by carefully selecting entry and exit points for equity investments.

Since the idea of the random walk research have attempted to find ways to improve the predictability of the market. In "Beyond the Random Walk", Singal (2004) discussed the concept of market efficiency and anomalies to the Random Walk hypothesis including the frequency of the mispricing, the financial instruments that can be used, and the number of transactions per year. Suk-Joong and McKenzie (2006) considered the relationship between stock market autocorrelation, the presence of international investors and the stock market volatility. Chui and Kwok (1998) evaluated the Cross-autocorrelation between A (the shares owned by Chinese citizens only) Shares and B Shares (shares owned by foreigners only) in the Chinese Stock Market. Lam and Lam (2000) attempted to improve trading results by forecasting a key summary statistic of future prices using neural network for the generation of trading signals. Bettini *et al.* (1998) used multiple granularities in time sequences in mining temporal relationships to predict the stock market. Morena and Olmedal (2007) found that Artificial Neural Networks do not provide superior performance than the linear models. Brooks (1998) emphasizes that and researchers are still uncertain as to the precise role of volume in the analysis of financial markets as a whole. Lo and Wang (2006) found the joint behaviour of price and quantity reveals more information about the relation between asset prices and economic factors than do prices alone.

The volume of stock traded is a measure of supply and demand that is independent of price. Investors often look at trading volume for confirming evidence of price trend and price reversal patterns. The general view among the investor population is that when volume begins to increase, it helps to identify the relative enthusiasm of buyers and sellers and direction of the price trend. Falling price accompanied by falling volume speaks of a lower relative enthusiasm to sell. When price goes to a new high on increased volume, traders often compare volume with that which occurred during previous rally's in prices. If the current volume is less than the previous rally's volume, it is generally accepted that there is a potential for a price trend reversal. There are several indicators and volume gauges commonly used by investors.

One such trading volume gauge includes the low-price activity ratio which is a comparison of the volume of high risk speculative stocks to blue chips. Net member buy sell ratio measures and compares the volume of shares purchased to the volume of shares sold by members of the stock exchange. The odd-lot theory claims that the ratio of odd lot purchases to odd lot sales gives indication of the uninformed investor and is used as a contrary indicator. Other indicators include the Volume Rate of Change, Volume Oscillator and Positive Volume Index.

When stock markets move away from random behaviour this creates an opportunity for a smart trader to make an excessive profit. These anomalies cannot last long because in a stock market a large number of traders who move in promptly to take advantage of non random patterns. Institutional investors clearly have an advantage over the small investor when it comes to smart investing. Recent empirical findings suggest that institutional investors play the role of smart traders on stock markets and, therefore, may have an impact on stock market anomalies. Institutional investors are the large fund managers and corporations. They have vast resources, information systems, researching capabilities which give them a clear informational advantage. This enables them to exploit economies of scale in information acquisition and processing. The finance professionals employed by the large institutional investors have the necessary qualifications, experience and the training well above the capabilities of an individual investor. In addition there is the financial industry practice of companies providing information to financial analysts in advance so that the good relationships can be maintained between the company and the analysts.

This fact has been confirmed by the research conducted by Dennis and Weston (2001). They found evidence for U.S. stock exchanges that institutions are better informed than individual investors. Another feature of the institutional investors is that they quickly move the stock prices towards correct fundamental valuation. Cohen *et al.* (2002) point out that this fact and confirm that they do this at the expense of individual traders. Barber and Odean (2008), point out "Individual investors display attention-based buying behaviour, whereas institutions do not exhibit this kind of trading mentality."

Investors invest money based on the available information. Normally the investment schema includes information flowing into the investors and investor activity changing the markets. The changed conditions in the markets are also a significant part of the information that is used by investors to make market related decisions. However in practice the investment procedure has another source of information;

that is privately held information. Figure 1 illustrates the investment procedure in practice.

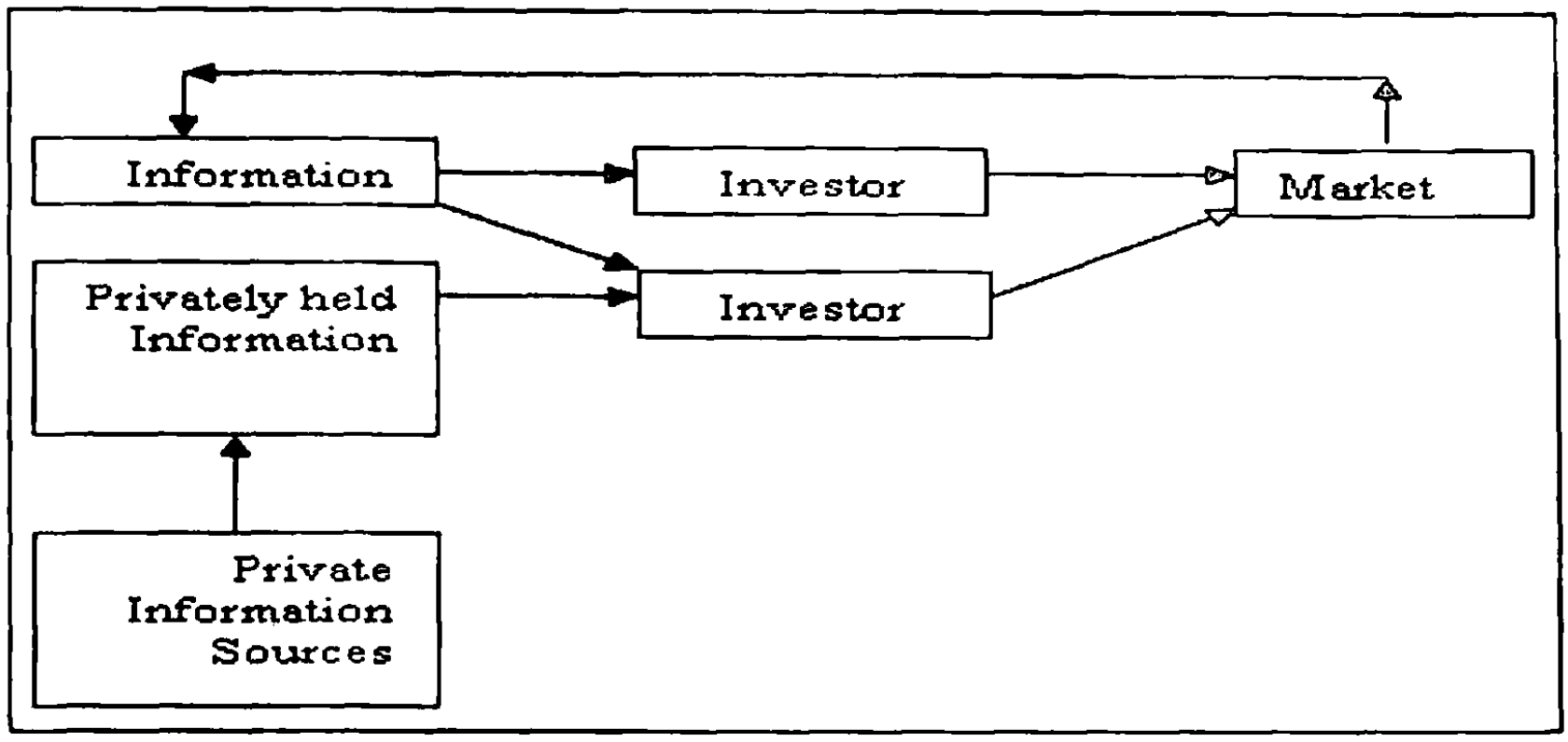


Figure 1: Investment Procedure – in practice

This privately held information is at best comes about due to extensive research conducted by large investors and at worst comes from sometimes illegally receiving information about major events of a stock which has an impact on its price. Large investors have the resources to allocate to continuously collect and evaluate market rumours, speculations and other information. This is an environment where the small investor certainly is at a disadvantage. One of the options taken by many small investors is that to entirely stay away from the stock market. This deprives the small investor of a valuable option of investment – the stock market.

This research uses available data to find ways to identify the effects of privately held information and reveal them to the small investor before it is too late, thus removing the undue advantage some large investors have over the small investor.

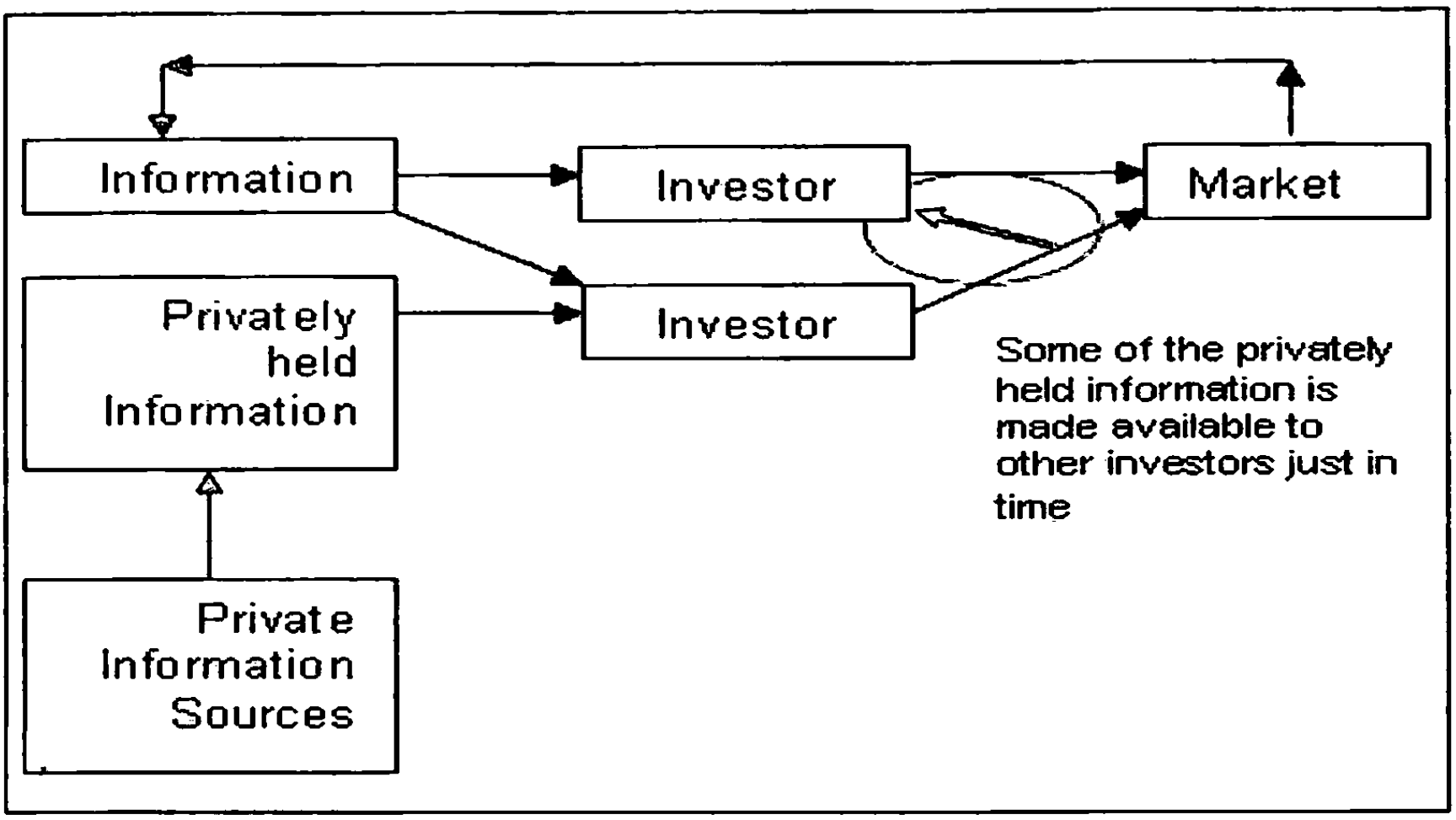


Figure 2: Investment Procedure – in practice with revealed privately held information

This also attempts to develop an algorithm which will analyse market-wide daily data and identify opportunities for the small investor to reduce the risk of investment in the short term. There is no published research which analyses the large increases in trading volume with little changes in the price of the stock. This algorithm studies the market-wide trade volume data to identify and isolate the companies which potentially hold private market related information as a way of removing a major disadvantage the small investor has over the large investment firms. This algorithm helps the small investor by reducing the information gap between the small and large investor.

The primary motivation of this paper is threefold. First, this research develops an algorithm to find recent excessive increases in the trading volume of stocks through a market-wide analysis. The analysis is extended to list the stocks with most excessive change in trading volume. When this is combined with marginal increase or decrease of stock price this is an indication of existence of privately held

information. Considering the direction of the price movement during the short period of the excessive change in trading volume, a prediction is made for the direction of the future value of the stock. Second, this paper endeavours to determine the feasibility and practicality of using this analysis as a forecasting tool for the individual investor. As a part of this research to achieve this purpose, a software tool was developed which has a simple graphical user interface. This can be used to perform the required analysis by an investor. Thirdly, accuracy of results of the model is validated empirically by performing the predictions on 100 different trading dates.

Overall, the paper makes three major contributions. First, it identifies the stocks with most increases in trading volumes in recent times at the Australian Stock Exchange and using the tool an investor can repeatedly generate the information every time the market data is updated -which is daily for the purpose of this research. Secondly the investor has an easy way of establishing the direction of the price trend. Finally this paper specifies the general algorithm for the process so that it can be automated in a data mining technique.

In summary, the paper finds ways for the investor to reduce the short term risk of investing in the share market.

METHODOLOGY

Data

The daily Australian Stock Exchange, stock market data was downloaded from Netquote (http://asx.netquote.com.au/about.asp) for the period of 01/07/2005 and 29/07/2008 for the initial analysis of the privately held information. Data from 01/09/2001 to 01/08/2009 was downloaded from the same source and used for the detailed analysis and evaluation of the algorithm. For this research end of the data feed was used. The data for each day comprises of Stock name, Date, Open price, Close price, Daily highest price, Daily lowest price and the volume of stock traded on that day.

Considering the high volume data and the complexity of analysis to be performed a software tool was specifically developed as a part of this research. This software tool – Stock Strategy Analyser (SSA) can upload the files in text format, store historical stock market data in a relational database, perform the necessary data transformation and analyse the data according the algorithms developed for this research.

Selection of Data

Initial analysis, predictions and evaluations were performed using daily stock market data for all companies which had daily data between 01/07/2005 and 29/07/2008. If a company had started after 01/07/2008 that company was excluded from the analysis for lack of historical data. Since the investment strategy being researched is a short term strategy, if the company was not trading on the 29/07/2008 (the date the analysis was carried out) that company was also excluded. For the detailed prediction on 100 days and the subsequent portfolio analysis, data from 01/09/2001 to 01/08/2009 was used.

ANALYSIS

The first step in the algorithm is to identify the stocks having unusually high recent trading volumes by scanning all the stocks in the market and their trading volumes. For the purpose of this research Partition 1 was created by inputting 01/02/2009 as the start date and 27/07/2009 as the end date. The SSA then separated the trading volume for each company within the date range period and calculated the average trading volume for each company within the user input date range. This is the partition 1. For the purpose of this research Partition 2 was created by inputting 28/07/2009 as the start date. SSA by default picked 29/07/2009 as the end date for partition 2 since it was the latest date where trading volume data were available. The SSA then separated the Trading volume for each company within the date range period and calculated the average trading volume for each company within the user input date range for the partition 2. The difference between Partition 1 and Partition 2 shows the change in volume in Partition 2 identifying recently hyper active stocks. An analysis of the average volume of trading stocks by the stock code revealed the number of days the trading volume exceeds 500% over the average trading volume, is less than 2% of the trading days. Therefore the cut off was set to 5 times the average trading volume. The stocks picked up have at least a 500% increase in trading volume in the partition 2 of the data set. Sorting the difference in descending order clearly identifies the stocks in the market with unusually high volumes of trading during the recent past – in this case from 28/02/2009 to 29/02/2009. SSA provides a facility to browse such companies graphically representing the closing stock price and trading volume for the period the data is available in the database – since 01/01/2005. It also lists the daily transaction file for each company showing the transaction date, open, high, low and close price for that stock together with the trading volume sorted in descending order of the date of transaction.

The second step is to predict the direction of the movement of the stock price within the next few days. The facts considered in this research are that if the volume has increased and the price has also increased significantly the increase can be attributable to favourable information regarding the company becoming public. Opportunity to make a short term gain in this case is limited since the available information is already reflected in the price. On the other hand if the volume has increased and the price has decreased significantly the decrease can be attributable to adverse information regarding the company becoming public. Opportunity to make a short term gain in this case is limited since the available information is already reflected in the price.

This research pays particular attention to situations other research does not consider – that is where the trading volume increases significantly (over 500%) while the price only shows a marginal increase or decrease. This supports the idea of the existence of privately held information.

The hypothesis tested in this research is that under such conditions, there is a good chance of privately held information which has not yet filtered through to the public arena. For the evaluation of the hypothesis, the programme was set to locate stocks with more than 500% increase in the trading volume in the partition 2 of the data. Simultaneously the price change in the partition 2 of the data set to lie within -5% to +5% from day 1 to day 2. In this situation the trading volume has increased over 5 times without a significant change in the price of the stock. This is hypothesised to be a pre runner to a rapid move in price.

This research attempts to verify this theory using market-wise ASX data. An astute investor can look for further confirmation of the existence of the privately held information by checking the market depth for the listed companies having unusually high trading volumes. If the market depth having buying pressure this is further confirmation of impending increase in price in the short term.

In traditional trading systems in addition to buy signals, sell signals are also generated. While this algorithm clearly identifies the buy signal, it is more difficult to generate the sell signal due to the nature of the patterns of trading this system is trying to predict. It is not possible to predict how long the advantage of privately held information would last. Since the purpose of this algorithm is to provide a way of making a capital gain in the short term, the evaluation period was set at 3 weeks.

Therefore rather than giving a sell signal, this research analyses the immediately following trading days to establish whether the price movements in the following 3 weeks since the prediction creates opportunities for the investor to make a profit. Depending on the direction of the prediction and the actual direction of the price movement the prediction accuracy is measured. Table 1 summarises the determination of the selling price for evaluation purposes.

Table 1: Determination of the Selling Price and the Profit/Loss

Direction			
Predict	**Actual**	**Selling Price**	**Comments**
Up	Up	Maximum closing price during the next 3 weeks	The maximum profit that can be made during the period of evaluation
Up	Down	Minimum closing price during the next 3 weeks.	The maximum loss that can be made during the period of evaluation.
Down	Up	Minimum price during the next 3 weeks	The maximum loss that can be made during the period of evaluation.
Down	Down	Maximum price during the next 3 weeks.	The maximum profit that can be made during the period of evaluation.

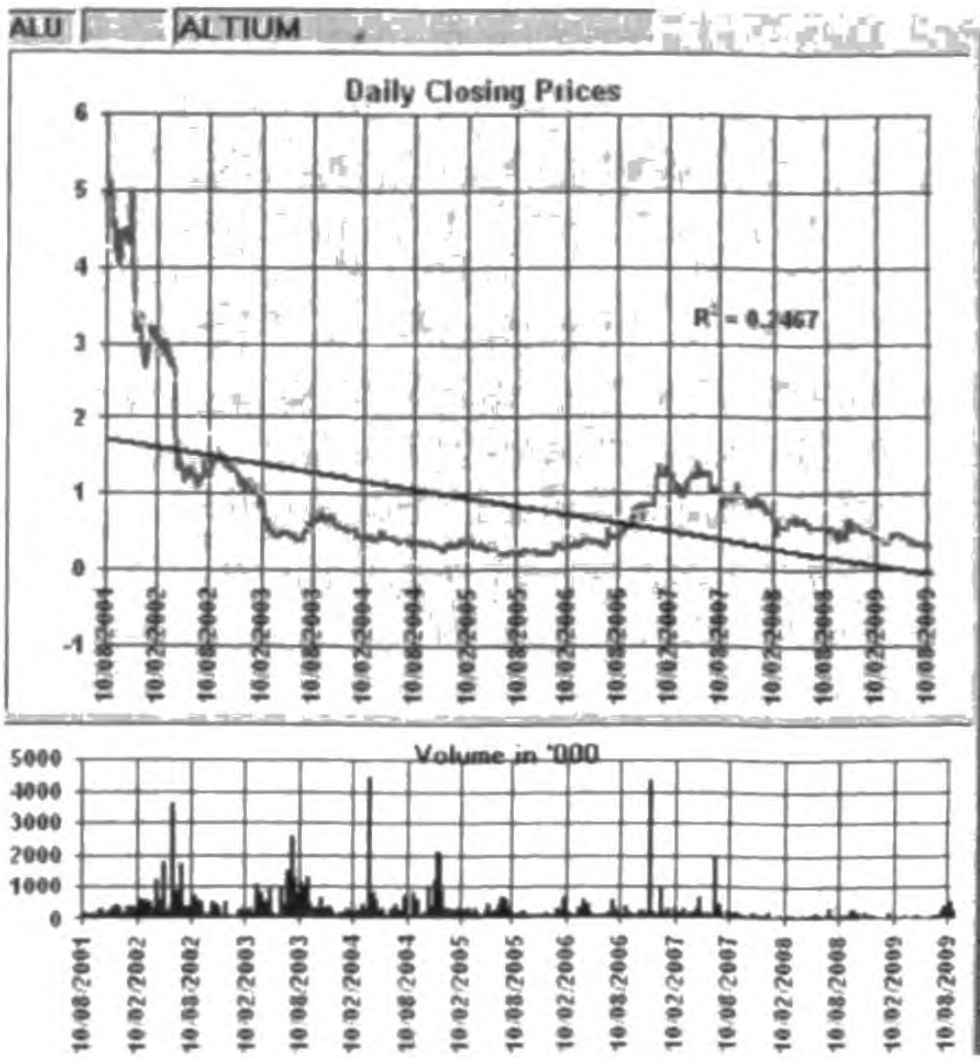
The initial analysis and the evaluation for the detection privately held information was carried out on the 29/07/2009. The software was provided with the data up to this date and the analysis was performed. The evaluation was carried out during the subsequent 3 weeks until 21/08/2009. The following figure shows the partition 1, partition2 and the evaluation period.

ALU Altium

ALU Altium was picked up by the software as having unusually high average trading volume (see figure 3) on the 28/07/2009 and 29/07/2009.

Closing price on 28/07/2009	31.5 c
Closing price on 29/07/2009	32.5 c
Predicted direction of price movement	Up
Purchase price on the 30/07/2009	32.5 c
Maximum possible Selling price until 21/08/2009	34.0 c
Maximum possible profit/loss until 21/08/2009	1.5 c
Accuracy	Correct
Profit % for 21 days	4.61%
Projected annual profit %	80.2%

This graph plots the daily closing prices for ALU for the period of 01/01/2005 to 21/08/2009.



Date	Open	Close	Low	High	Volume	Value
21/08/2009	0.31	0.31	0.31	0.31	27,500	8,525
20/08/2009	0.3	0.3	0.3	0.3	2,368	710
19/08/2009	0.305	0.305	0.31	0.34	554,679	169,177
18/08/2009	0.3	0.305	0.3	0.31	278,605	84,975
17/08/2009	0.32	0.31	0.31	0.32	75,000	23,250
14/08/2009	0.33	0.33	0.33	0.33	20,000	6,600
13/08/2009	0.32	0.34	0.32	0.34	183,748	62,474
12/08/2009	0.305	0.3	0.3	0.31	15,600	4,680
11/08/2009	0.31	0.31	0.31	0.32	154,763	47,977
10/08/2009	0.32	0.305	0.31	0.32	192,700	58,774
7/08/2009	0.32	0.32	0.32	0.33	231,208	73,987
6/08/2009	0.32	0.32	0.32	0.32	15,754	5,041
5/08/2009	0.33	0.32	0.32	0.33	79,500	25,440
4/08/2009	0.33	0.32	0.32	0.33	134,986	43,196
3/08/2009	0.34	0.34	0.34	0.34	4,522	1,537
31/07/2009	0.325	0.335	0.33	0.34	11,761	3,940
30/07/2009	0.335	0.325	0.32	0.34	103,000	33,475
29/07/2009	0.33	0.325	0.32	0.34	105,000	34,125
28/07/2009	0.32	0.315	0.32	0.33	94,062	29,630
24/07/2009	0.35	0.33	0.32	0.36	417,644	137,823
21/07/2009	0.32	0.32	0.32	0.32	2,000	640
20/07/2009	0.315	0.315	0.32	0.32	12,750	4,016
15/07/2009	0.31	0.32	0.31	0.33	20,500	6,560
14/07/2009	0.32	0.32	0.32	0.32	225,686	72,220
13/07/2009	0.33	0.31	0.31	0.33	39,462	12,233
8/07/2009	0.335	0.335	0.34	0.34	3,750	1,256
7/07/2009	0.34	0.34	0.34	0.34	4,301	1,462
3/07/2009	0.35	0.35	0.35	0.35	20,000	7,000

Figure 3: ALU Altium

Overall Evaluation – Portfolio Performance, Privately held information

For the purpose of verifying the claim high volume trading activity creates an opportunity for capital gains, 100 different dates over a period of 8 years were initially selected on which high volume trading activity to be located.. These dates were picked up at an interval of 20 trading days.

The high volume was measured by comparing the average trading volume during the partition 1 of the data set with the partition 2 of the data set for all listed companied in the ASX. For this analysis stocks having more than 500% increase in the trading activity in the partition 2 of the data set were used. The other is the movement of the price. If the price has moved significantly, the information is already included in the price. If the price is not showing significant movement during the period of the partition 2 there is the possibility of privately held information. The direction of the prediction is based on the direction of the price movement during the partition 2 period, usually the last 2 days.

SUMMARY RESULTS

The overall accuracy based on the number of prediction is 79%. Out of a total of 1620 cases 1287 cases were predicted correctly

Table 2 provides a summary of predictions made from data from 2001 to 2009 of privately held information indicated by sudden increase in trading volume by over 500%.

Table 2: Predictions made from 2001 – 2009 overall Performance

Predicted Up/ Actual Up	228
Predicted Down/ Actual Down	1059
Total Correct Predictions	1287
Predicted Up/ Actual Down	198
Predicted Down/ Actual Up	135
Total Incorrect Predictions	333
Total Predictions Made	1620
Accuracy %	79
Overall Yearly Profit %	71.93

Summary Results by Year
The following is a summary of predictions made from data from 2001 to 2009 of privately held information (Volume X 5) analysed by year (Table 3).

Table 3: Predictions made from 2001 – 2009 yearly performance

Year	Number of Predictions	Projected Profit % Per Year
2001	180	63.68
2002	542	77.05
2003	432	4.79
2004	199	65.25
2005	96	73.53
2006	75	57.75
2007	43	76.31
2008	18	100.19
2009	35	43.79
Total/Overall	1620	79

Generic Algorithm for Automating the Technique for Data Mining
(For all stocks in the market from 1 to N)

For J = 1 to N
    Partition data file to Training and Test datasets partitions
    Partition the training dataset into 2 subsets A and B such that d1 < A < d2 and d3 < B d4.
    d1, d2, d3, d4 are training dates for stock i. d1<d2<d3<d4
    Diff = Average Volume i = d3 to d4 - Average Volume i = d1 to d2
    If Diff > threshold save Diff in a file
    Repeat fro next stock until all completed
Next J
Sort diff values in descending order

The market-wide dataset is partitioned into 2 sub datasets. The partition 1 date range is defined by the user input. The second partition is the last few days where the trading volume is displaying increased activity. The algorithm identifies the stocks

with the highest difference between the average trading volume and the last few days trading volume.

CONCLUSION

This paper developed an algorithm for predicting the market short term stock price direction more accurately when unusually high volumes of trading data indicates privately held information. A generic algorithm is specified so that an automated data mining process can be programmed. This will identify the stocks that have unusually high volumes of trading data, display the daily price movements and its direction side by side and forecast the direction of the price movement for the short term. This can produce more accurate results over the general trend of a given stock which is more suitable for long term investment horizon.

FURTHER RESEARCH

Currently the automatic algorithm does not consider the market depth information for stocks identified as having unusually high volumes of trading data and potentially privately held information. Further research can incorporate this aspect into the algorithm to include additional information that can be useful increasing the accuracy of the prediction. Buying pressure in the market depth indication increase in price and selling pressure indicating decrease in price. During the period this research performed the analysis the market had a general downward trend due to the financial crisis. The analysis still gave good predictions in the direction of the movement in stock price. Further research can be carried out performing the analysis when the market has a general upward trend. This research was limited to generating the buy signals. The effectiveness of the buy signals was evaluated based on the opportunities for selling with a profit in the short term investment horizon. A further improvement would be to improve the algorithm to generate a clear sell signal. As part of this project it is planned to conduct further research into closely related stock prices with a lag of $K = 2,3,\ldots$ trading days to predict the short term direction of the stock price.

REFERENCES

Barber, B.M. and T. Odean (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Rev. Financ. Stud.*, 21: 785–818.

Bettini, C., X.S. Wang and S. Jajodia (1998). Mining temporal relationships with multiple granularities in time sequences. *Data Engineering Bulletin*, 21:32-38.

Bollerslev, T., R.Y. Chou and K.F. Kroner (1992). ARCH modelling in Finance: A review.

Brooks, C. (1998). Predicting Stock Index volatility: Can market volume help? *Journal of Forecasting*, 17: 59-80.

Chui, A.C.W. and C.C.Y. Kwok (1998). Cross-autocorrelation between A Shares and B Shares in the Chinese Stock Market. Journal of Financial Research, September 1998. Available at SSRN< http://ssrn.com/abstract=89628>

Cohen, R.B., P.A. Gompers and T. Vuolteenaho (2002). Who underreacts to cash-flow news? Evidence from trading between individuals and institutions, *Journal of Financial Economics*, 66: 409–462.

Dennis, P.J. and J.P. Weston (2001). Who's Informed? An Analysis of Stock Ownership and Informed Trading. <http://ssrn.com/abstract=267350>.

Suk-Joong, K. and M.D. Mckenzie (2006). Conditional Autocorrelation and Stock Market Integration, November.

Lam, K. and K.C. Lam (2000). Forecasting for the generation of trading signals in financial markets. *Journal of Forecasting*, 19: 39-52.

Lo, A.W. and J. Wang (2006). Trading volume: Implications of an inter-temporal capital asset pricing model. *The Journal of Finance*, LXI.

Morena, D. and I. Olmeda1 (2007). Is the predictability of emerging and developed stock markets really exploitable? *European Journal of Operational Research*, 182(1): 436-454.

Singal, V. (2004). *Beyond the Random Walk: A Guide to Stock Market Anomalies and Low-Risk Investing*. Oxford University Press, USA.