8th International Conference on Advanced Computational Intelligence
Chiang Mai, Thailand; February 14-16, 2016

Stock Price Manipulation Detection using a Computational Neural Network Model

Teema Leangarun and Poj Tangamchit
Dept. of Control System and Instrumentation Engineering
King Mongkut's Univ. of Tech.
Thonburi, Bangkok, Thailand
Email: teema.lea@mail.kmutt.ac.th, poj.tan@kmutt.ac.th

Suttipong Thajchayapong
National Electronics and Computer Technology Center
NSTDA Bangkok, Thailand
Email: suttipong.tha@nectec.or.th

***Abstract*— We investigated the characteristics of stock price manipulation. Two manipulation models were studied: pump-and-dump and spoof trading. Pump-and-dump is a procedure to buy a stock and push its price up. Then, the manipulator dumps all of the stock he holds to make a profit. Spoof trading is a procedure to trick other investors that a stock should be bought or sold at the manipulated price. We constructed mathematical models that use level 2 data for both procedures, and used them to generate a training set consisting of buy/sell orders within on order book of 10 depths. Order cancellations, which are important indicators for price manipulation, are also visible in our level 2 data. In this paper, we consider a challenging scenario where we attempt to use less-detailed level 1 data to detect manipulations even though using level 2 data is more accurate. We implemented feedforward neural network models that have level 1 data, containing less-detailed information (no information about order cancellation), but is more accessible to investors as input. The neural network model achieved 88.28% for detecting pump-and-dump but it failed to model spoof trading effectively.**

Keywords—stock price manipulation; neural network; pump-and-dump; spoof trading

I. INTRODUCTION

Stock market gathers participations from all kinds of investors. Millions of buy/sell orders enters the market every day. Stock price fluctuates due to several factors, mainly from the profit that the company can make. However, there are some investors who attempt to get benefits from the stock market using irregular trade behaviors that affect the stock price. Some of these attempts are illegal. The control of these irregular trade behaviors is difficult due to the large amount of trade data.

Automatic computer algorithms for detecting price manipulation are the solution to this problem. It can scan large amount of price data in a short time. Price manipulation can be divided into three categories: trade-based, information-based, and action-based. This research discusses an ad-hoc approach to detect trade-based price manipulations in stock market. Two scenarios are investigated: pump-and–dump and spoof trading. Pump-and-dump is the action of buying stock, making the price to go higher, and then selling to others in a short period of time. Spoof trading is an action of sending passive orders in high volume to trick others that the stock should be sold at that price. After the manipulators secure benefits from that fake price, they cancel their passive orders. These actions allow the manipulators to sell their stock at a price higher than usual. These actions are illegal in some countries, and should get under control.

This research was supported by the National Science and Technology Department Agency (NSTDA), Thailand Graduate Institute of Science and Technology (TG-44-20-58-027M).

The effectiveness of manipulation detection depends on how much the information we have. We rely on using the price data that buyers and sellers sent to the market. The trade data can be classified into two levels. Level 1 data consists of buy/sell orders that are successfully executed. It has a format of open, high, low, close price and volume within a specific time period. Level 1 data is usually accessible by the public, thus easy to obtain. Level 2 data consists of all information from Level 1 data plus buy/sell orders that are not matched. It shows each particular order that is entered, cancelled, or matched. Sometimes, level 2 data shows an order ID, or buyer/seller ID, which can be an important clue to show that the actions originate from the same person. In general, level 2 data will not be opened to the public. It can only be accessible by market authorities. This is because an investor can lose his benefits if his ID can be identified. This makes other people know what he is doing and perform a counter-action to gain the benefit from him.

In this paper, we consider a challenging scenario where we attempt to use less-detailed level 1 data to detect manipulations even though using level 2 data is more accurate. The results showed that this can be done in pump-and-dump, in which price data reflects the intention of the manipulator. However, the spoof trading cannot be identified using only level 1 data, because its trace is not noticeable in level 1 data.

II. RELATED WORK

Allen F. and D. Gale [1] analyzed price manipulation models from asymmetric information. Price manipulation activities were categorized into three categories. Information-based manipulation attempts to release market rumors, which affects the fair price. Action-based manipulation controls demand and supply, which influences the real value of assets. Trade-based manipulation uses buy/sell orders to manage equity price.

Stock price manipulation model in Tehran stock exchange was forecasted by F. Rahnamay Roodposhti et al. [2]. The capability to forecast price manipulation of three models: logit model, artificial neural network, and multiple discriminant analysis were compared. The efficiency of these models were similar. The models used a dependent variable, which was the position of price manipulation. It can be divided into two sets: non-manipulated and manipulated companies. Size of company, P/E ratio, stock liquidity rank, status of information clarity, and structure of shareholder were defined as impact factors of price manipulation.

According to H. Ögüt *et al.* [3], stock price manipulation in Istanbul stock exchange, which is an emerging market, was detected. Index's average daily return, average daily change in trading volume, and average daily volatility were used as interpretative variables. Data mining techniques (Artificial neural network and support vector machine) were superior methods to detect stock price manipulation than multivariate techniques (logistic regression and discriminant analysis).

In Y. Cao and Y. Li's work [4], Adaptive Hidden Markov Model with Anomaly States (AHMMAS) was applied with level 2 data from NASDAQ and London stock exchange to detect intraday price manipulation activities. The model was tested with simulated data and real market data. The performance evaluation of the model was compared to standard algorithms such as: Gaussian Mixture Models (GMM), K-Nearest Neighbors Algorithm (kNN), and One Class Support Vector Machines (OCSVM).

However, most of investors have limited access to level 2 data. This paper explores neural network models to recognize stock price manipulations using only level 1 data, which is easier to acquire. Price manipulations have different forms. In this paper, we focused on detecting "pump-and-dump" and "spoof trading", because they are the most common form of price manipulations.

III. Stock Price Manipulation and Its Models

A. Stock Price Manipulation

A stock market is a place where investors can buy and sell ownership of companies. An investor places orders to the exchange market system. The orders represent buy or sell volume and price. Price matching occurs when a buyer and a seller offer the same price for buying and selling a security respectively. The highest price that a buyer will pay to buy number of shares of a security is also called the bid price. The ask price is the lowest price that a seller will receive to sell number of shares of a security. Price manipulation is an attempt to control these bid and ask prices. We focus on trade-based manipulations, in which manipulators submit crafted bid and ask orders to control the stock price. Trade-based manipulations can have many forms, but they all use similar tactics. A manipulator enters non-bona fide buy orders into the exchange market system, the price of the security is increased, and some investors add buy orders, joining the rising price. When the manipulator secures enough profit, non-bona fide orders and holding position are reduced.

B. Data Level, Depth of Market, and Cancellation Order

The market data can be divided in two levels [5]: level 1 market data and level 2 market data. Level 1 data contains current basic market information, which is composed of open, high, low, close, and volume (OHLCV). In general, most investors use level 1 data for trading. Level 2 data has more trading information than level 1 data, such as the depth of market data, order book, and market participant ID (MPID). Order book is known as depth of market (DOM). The order book displays bid and ask price that have not been matched. Higher market depth provides more resolution data. MPID is the most important information that indicates the one who places orders in the system. Irregular size of cancellation or deletion order and faster cancellation time can be used as important indicators for detecting price manipulation. A system overview is shown in Fig.1. Level 2 data was transformed to level 1 data, which used as inputs and manipulated class. We implemented price manipulation models that have level 1 data as inputs. The model forecasted an output value in probability of stock price manipulation.

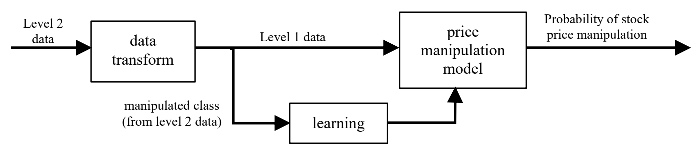


Fig. 1. System Overview.

C. Stock Price Manipulation Detection Model: Pump-and-Dump

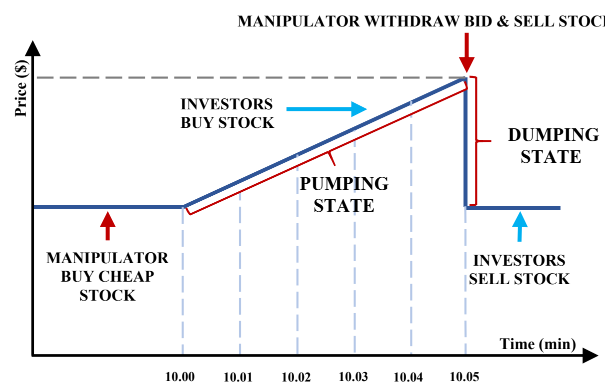


Fig. 2. Stock Price Manipulation Detection: Pump and Dump.

A manipulator starts by sending buy orders, pumping up the price to reach a desired level as shown in Fig.2. In this period, most of other investors think that the price is going up and add buy orders. Then, in the dumping phase, profits are made by cancelling all remaining buy orders and rapidly executing sell orders at the manipulated prices. Therefore, the other investors who are not cautious have bought the stock at high price.

This paper established a model to forecast stock price manipulation for detecting pump-and-dump. Level 2 data with 10 depth was used as a training set. It was processed and separated into two parts. The first part was level 1 data, which we used as inputs. The second part was the points that pump-and-dump are detected, which we used as desired output. We defined the length of data set for detecting pump-and-dump by a sliding window as below.

$V_{buy}^{cancel}(t)$	cancellation volume of buy orders at time t
$E[V_{buy}^{matched}(t)]$	average volume of buy orders that have been matched at time t

$P_{sell}^{max}(t)$ — the highest price of sell orders that have been matched at time t

$P_{sell}^{min}(t)$ — the lowest price of sell orders that have been matched at time t

$P_{bid}^{max}(t)$ — the highest bid price that have been matched at time t

$P_{bid}^{min}(t)$ — the lowest bid price that have been matched at time t

Step 1: In the dumping period, the amount of cancellation orders and price matching orders were verified.

For the first condition, we classified the event as a dumping action when the number of the withdrawal of buy orders $V_{buy}^{cancel}(t)$ is more than threshold_1, where threshold_1 is 50% of the average volume of buy orders $E[V_{buy}^{matched}(t)]$ during that period.

$$\text{dump\_cond1} = \begin{cases} 1\,; & V_{buy}^{cancel}(t) > (\text{threshold\_1})(E[V_{buy}^{matched}(t)]) \\ 0\,; & \text{otherwise} \end{cases} \quad (1)$$

For the second condition, the different between the high price of sell orders $P_{sell}^{max}(t)$ and the low price of sell orders $P_{sell}^{min}(t+1)$ is more than threshold_2 = 0.15%.

$$\text{dump\_cond2} = \begin{cases} 1\,; & \frac{P_{sell}^{max}(t) - P_{sell}^{min}(t+1)}{P_{sell}^{max}(t)} > \text{threshold\_2} \\ 0\,; & \text{otherwise} \end{cases} \quad (2)$$

Then, we treated that the dumping action occurs when both condition 1 and condition 2 are met.

$$\text{dump} = \text{dump\_cond1} \land \text{dump\_cond2} \quad (3)$$

When dump activities (step 1) are detected, the condition in step 2 is tested. If dump activities are not detected, we identify it as normal activities or non-manipulated of a stock.

Step 2: After the dumping points are detected, we test whether the price was pumping up before. We treated the event as a pumping activity when the highest bid price $P_{bid}^{max}(t-1)$ increases more than threshold_3 = 0.2% compared to the lowest bid price $P_{bid}^{min}(t-4)$ at the starting point.

$$\text{pump} = \begin{cases} 1\,; & \frac{P_{bid}^{max}(t-1) - P_{bid}^{min}(t-4)}{P_{bid}^{max}(t-1)} > \text{threshold\_3} \\ 0\,; & \text{otherwise} \end{cases} \quad (3)$$

$$\text{pump and dump} = \text{dump} \land \text{pump} \quad (4)$$

When the conditions of both step 1 and step 2 are satisfied, we treated them as pump and dump points. They will be used as a training data in our neural network model.

D. Price Manipulation Detection Model: Spoof Trading

Spoof trading is a strategy to trick investors that the stock should be bought or sold at the manipulated prices. This is done by inserting spoof bid or ask orders, which will be cancelled when it is about to be matched. Since the volume of the spoof orders is high, other investors usually think that this price comes from the consensus and is reasonable. In Fig.3, a manipulator wants to buy the stock at the price lower than the current ask price. In Fig. 3(a), he then sends large orders (large volume at passive price) as shown in the dotted bar, which is lower than the current ask price. Other investors join into this spoof orders in Fig. 3(b), expecting the best ask price to go down. Then, a manipulator will withdraw the large spoof orders and buy all the remaining orders from other investors at this manipulated price in Fig. 3(c).

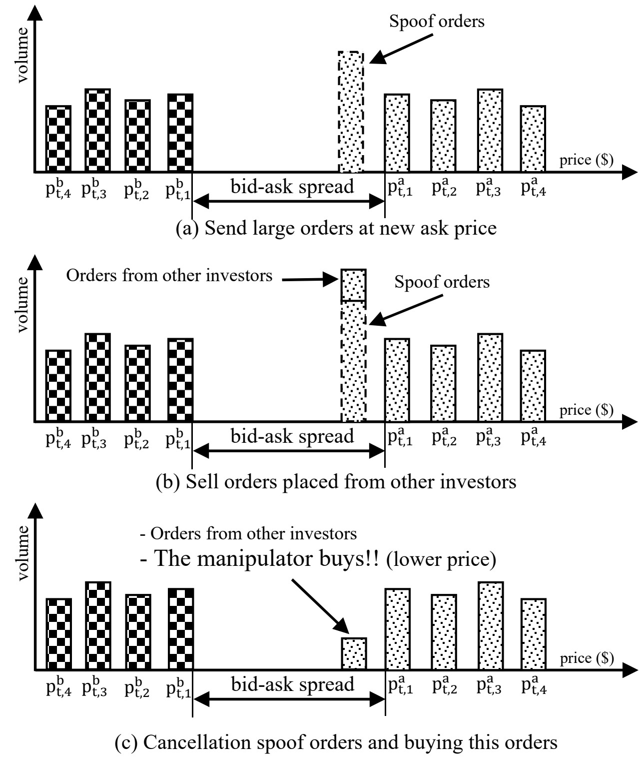


Fig. 3. Stock Price Manipulation Detection: Spoof Trading. (a) Send large orders at new ask price (b) Sell orders placed from other investors (c) Cancellation spoof orders and buying this orders

We used 10 depth of level 2 data to create a training set for the spoof trading detection model. The data was processed to extract inputs (level 1 data) and labels (the point where spoof trading occurred) in the same way as the pump-and-dump model. We defined the length of data set for detecting spoof trading by a sliding window as below.

We defined three conditions for checking spoof trading: cancellation order has its price close to the current bid or ask price, cancellation volume is high enough, and there are high volume for the last buy order. For the first condition, we treated the event as a spoof trading action when the cancellation sell orders has a price $P_{sell}^{cancel}(t)$ within 0.5% (threshold_4) of the current ask price $P_{buy}^{matched}(t)$.

$P_{sell}^{cancel}(t)$ — price of sell orders that has been cancelled at time t

$P_{buy}^{matched}(t)$ — price of the last buy order that has been matched at time t

$V_{sell}^{cancel}(t)$ — cancellation volume of sell orders at time t

$V^{matched}(t)$ — volume of buy and sell orders that have been matched at time t

$V^{matched}_{buy}(t)$ volume of buy orders that have been matched at time t

$$spoof\_cond1 = \begin{cases} 1\,; \left|\frac{P^{cancel}_{sell}(t) - P^{matched}_{buy}(t)}{P^{cancel}_{sell}(t)}\right| < threshold\_4 \\ 0\,;\ otherwise \end{cases} \quad (4)$$

For the second condition, the amount of the cancellation sell orders $V^{cancel}_{sell}(t)$ is more than threshold_5, where threshold_5 is five times of the summation of matching orders $\sum_{n=1}^{t-1} V^{matched}(n)$ since the starting point.

$$spoof\_cond2 = \begin{cases} 1\,;\ V^{cancel}_{sell}(t) > (threshold\_5)(\sum_{n=1}^{t-1} V^{matched}(n)) \\ 0\,;\ otherwise \end{cases} \quad (5)$$

For the third condition, the amount of matching buy orders $V^{matched}_{buy}(t)$ is more than 50% (threshold_6) of the summation of matching orders $\sum_{n=1}^{t-1} V^{matched}(n)$ since the starting point.

$$spoof\_cond3 = \begin{cases} 1\,;\ V^{matched}_{buy}(t) > (threshold\_6)(\sum_{n=1}^{t-1} V^{matched}(n)) \\ 0\,;\ otherwise \end{cases} \quad (6)$$

Therefore, we interpreted that the spoof trading action occurs when three conditions are satisfied. We will evaluate the spoof trading in our neural network model.

$$spoof\ trading = spoof\_cond1 \wedge spoof\_cond2 \wedge spoof\_cond3 \quad (7)$$

IV. Neural Network Model for Detecting Price Manipulation

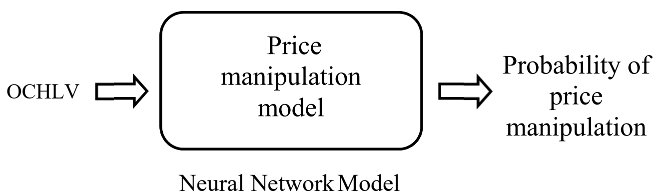


Fig. 4. Overall of Neural Network Model for Detecting Stock Price Manipulation.

In Fig. 4, we created a feedforward neural network model to detect stock price manipulation using open, high, low, close, volume (OHLCV) of five time steps as inputs. The model consisted of 25 nodes in the input layer, 3 nodes in the hidden layer, and one node in the output layer. We used the back propagation algorithm for supervised training. The model forecasted an output value in a binary variable. We represented '0' and '1' for non-manipulated point and manipulated point respectively. The model is used to pinpoint the position that pump-and-dump and spoof trading occurred. We expected to predict stock price manipulation using only level 1 data. Finally, the performance was evaluated by a statistical method.

A. Data Preparation

From NASDAQ (LOBSTER [6]), we collected stock data of three companies: Amazon, Intel, and Microsoft in intraday period on 21st June, 2012. The sample data set in this research was level 2 data, which showed cancellation orders. This data set was assumed to indicate price manipulation, because there were high trading volume and large price fluctuation.

1) Pump-and-dump: We started by defining a sliding window with 5 points t = {1, 2, …, 5}. The time frame from each period was 1 minute. Fig. 5 shows a pump-and-dump position of Intel company from level 2 data. In general, an exchange market cannot reveal public cancellation orders. Thus, we used neural network for training a pump-and-dump model from level 1 data. The data sample set was divided into 2 parts: the training set for training a neural network model and the test set for verifying the training results. The model was tested by leave-one-out cross validation. Mean square errors were utilized for performance evaluation.

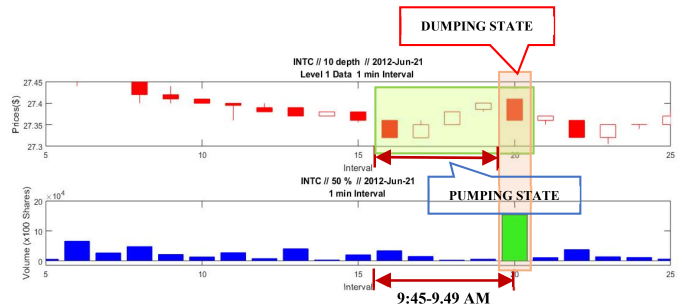
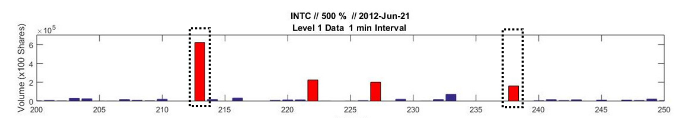
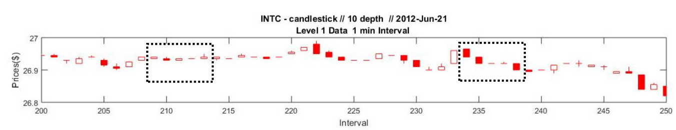


Fig. 5. Price Manipulation Detection of Intel Company: Pump-and-Dump.

2) Spoof trading: The length of a sliding window was defined as 5 points t = {1, 2, …, 5}. Time frame of 1 minute was used. Spoof trading positions of Intel company are shown in Fig.6. Although spoof trading positions were found in Fig. 6(a), we could not identify them using price pattern from level 1 data as shown in Fig. 6(b).



(a) Spoof trading positions (red bar)



(b) Level 1 data

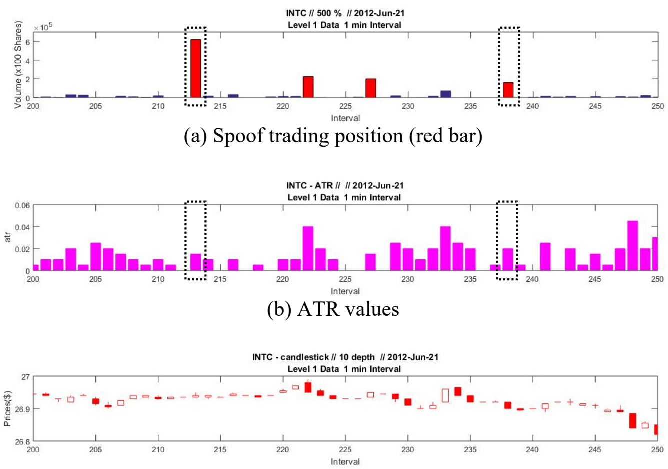
Fig. 6. Price Manipulation Detection of Intel Company: Spoof Trading. (a) Spoof trading positions (red bar) (b) Level 1 data

V. Experiment and Result

We used level 1 data (OHLCV) as inputs for forecasting stock price manipulation. The data was sliced and selected as sample for 22 sets. Each set has a tag whether it has manipulation signal or not. Total data training sets (22 sets) were split into 2 groups: non-manipulated stock 50%, and manipulated stock 50%. For performance evaluation, the model was trained 10 times per model and chose the one that had the smallest error. We used the maximum number of training epochs and the target error weight as 500 and 0 respectively. After the training is completed, we tested the model with the test set using leave-one-out cross validation method. We selected leave-one-out cross validation, because there were no many manipulations compared to normal trading. The results of pump-and-dump and spoof trading model are shown in table. 1. The neural network model for detecting spoof trading has more average mean square error, false positive error, and false negative error than pump-and-dump model. Thus, level 1 data is not sufficient to indicate spoof trading, because its price pattern is not distinct.

TABLE I. RESULTS USING NEURAL NETWORK MODEL FOR DETECTING PUMP-AND-DUMP AND SPOOF TRADING.

	Pump-and-dump	Spoof trading
Average mean square error	0.0641	1.3992
False-positive error	0.0634	1.3479
False-negative error	0.0649	1.4505



(a) Spoof trading position (red bar)

(b) ATR values

(c) Level 1 data

Fig. 7. Price Manipulation Detection of Intel Company: Spoof Trading with ATR values. (a) Spoof trading position (red bar) (b) ATR values (c) Level 1 data

For spoof trading, we investigated it further more to understand why the neural network training has a poor result. The average true range (ATR) was selected and compared with the cancellation volumes. The ATR can be computed from a moving average of true ranges, which indicates price volatility. We assumed that the spoof trading positions have high ATR values. The true ranges can be calculated in three conditions: distance from current high to current low, distance from previous close to current high, and distance from previous close to current low. The period of moving average was 14. In Fig. 7(b), the ATR values are shown. The points where spoof trading occurs can have either high or low ATR values. This means that spoof trading does not imply price fluctuation. Thus, the price does not show a distinct pattern. Therefore, price pattern from level 1 data cannot be an effective indicator for pinpointing spoof trading.

VI. CONCLUSION

This paper aimed to construct neural network models to detect trade-based stock price manipulations. We chose pump-and-dump and spoof trading cases to be our main study. Level 2 data was used to pinpoint price manipulation events, because we can see cancellations of orders in level 2 data. Cancellations of orders are important indicators to identify price manipulation. A set of less-detailed level 1 data was used as inputs of neural network models, which were trained to recognize these manipulation events. We evaluated the models by leave-one-out cross validation. Performance evaluation was based on the mean square errors. The pump-and-dump model can indicate the intention of a manipulator from price pattern in level 1 data. The results of price manipulation detection using level 1 data and level 2 data were close, thus showing the effectiveness the pump-and-dump model. Nevertheless, level 1 data has less information. It could not indicate spoof trading positions from the price pattern of level 1 data. We compared ATR values to the cancellation volumes. ATR implied price volatility. At the points where spoof trading occurs, ATR values are not consistent. This means that the price pattern of level 1 data is not obvious, and thus ineffective for predicting spoof trading.

ACKNOWLEDGMENT

This research was supported by the National Science and Technology Department Agency (NSTDA), Thailand Graduate Institute of Science and Technology (TG-44-20-58-027M).

REFERENCES

[1] Allen F. and D. Gale, "Stock-Price Manipulation," The Review of Financial Studies, 1992. 5(3): pp. 503-529.

[2] F. Rahnamay Roodposhti, M. Falah Shams, and H. Kordlouie "Forecasting Stock Price Manipulation in Capital Market," World Academy of Science, Engineering and Technology 56, 2011, pp 151-161.

[3] H. Ögüt, M. M. Doganay, and R. Aktas, "Detecting stock-price manipulation in an emerging market: The case of Turkey," Expert Syst. Appl., vol. 36, no. 9, 2009, pp. 11944-11949.

[4] Y. Cao and Y. Li, "Adaptive Hidden Markov Model with Anomaly States for Price Manipulation Detection", IEEE Trans. on Neural and Learning Systems, Vol. 26, no. 2, 2015, pp.310-330.

[5] Adam Milton, NYC, USA. (2015). Order Book, Level 2 Market Data, and Depth of Market [Online]. Available: http://daytrading.about.com/od/daytradingglossary/a/OrderBook.htm

[6] LOBSTER, Atlanta, GA, USA. (2012). Limit Order Book System [Online]. Available: http://www.lobster.wiwi.hu-berlin.de