Stock Price Manipulation Detection Based on Mathematical Models

Teema Leangarun, Poj Tangamchit, and Suttipong Thajchayapong

Abstract—We investigated two popular scenarios of stock price manipulations: 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 proposed mathematical definitions based on level 2 data for both scenarios, and used them to generate a training set consisting of buy/sell orders in an order book of 10 depths. Order cancellations, which are important indicators for price manipulation, are also visible in this level 2 data. In this paper, we considered a challenging scenario where we attempted to use less-detailed level 1 data to detect manipulations even though using level 2 data is more accurate. First, 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 an input. The neural network model achieved 88.28% accuracy for detecting pump-and-dump but it failed to model spoof trading effectively. Therefore, we further investigate the two-dimensional Gaussian model and show that it can detect spoof trading using level 2 data as input.

Index Terms—Stock price manipulation, pump-and-dump, spoof trading, neural network.

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 and spot manipulations in a short time. Price manipulation can be divided into three categories: trade-based, information-based, and action-based. This research discusses a mathematical model that classifies trade-based price manipulations from normal trades in stock markets. Two types of manipulations are investigated: pump-and–dump and spoof trading. Pump-and-dump is an action of buying stock, making the price to go higher, and then selling to others for a profit. Spoof trading is an action of sending passive orders in large volume to trick others that the stock should be sold at that price. After the manipulators secure enough benefits from that artificial price, they cancel their passive orders. These actions allow the manipulators to sell their stock at a price higher than usual.

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 our work [1], we considered a challenging scenario where we attempt to use less-detailed level 1 data to create a neural model for detecting 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. Therefore, we had to use level 2 data and created a 2-dimensional Gaussian model for the spoof trading cases.

II. LITERATURE REVIEW

Allen F. and D. Gale [2] investigated price manipulation models from asymmetric information, in that the financial market agents have reasonable expectations and maximizing expected utility. Price manipulation activities were classified into three categories. Information-based manipulation tries to publicize false information, which influences the fair price. Action-based manipulation, which affects the price of a stock, is an action other than trading that can manipulate demand/supply of the stock. In trade-based manipulation, a manipulator creates non-bona fide buy/sell orders to control

Teema Leangarun and Poj Tangamchit are with the Department of Control Systems and Instrumentation Engineering, King Mongkut's University of Technology Thonburi, Bangkok, Thailand (e-mail: teema.lea@mail.kmutt.ac.th, poj.tan@kmutt.ac.th).

Suttipong Thajchayapong is with the National Electronic and Computer Technology Center, National Science and Technology Development Agency (NSTDA), Thailand (e-mail: suttipong.tha@nectec.or.th).

an equity price. R. A. Jarrow [3] analyzed price momentum theory that was identified to be a result of price manipulation activities. Large investors have high influence to the stock price. He indicated that increasing stock price during manipulation activities produces the same trend of the stock price in the future. A. K. Rajesh, and G. J. Wu [4] extended Allen F. and D. Gale model. Stock price manipulation in U.S. stock market from 1990 to 2001 was studied. They found that an investor who has insider information of the company has high possibility to be a manipulator. Stock price usually increase during the manipulation process and decreased after the manipulation process has completed. Stocks with low liquidity have high probability to be manipulated. After manipulation process, the stock liquidity was increased. H. Öğüt, M. M. Doganay, and R. Aktas [5] studied trade-based manipulations. Stock price manipulation in Istanbul stock exchange, which is an emerging market, was investigated. Index's average daily return, average daily change in trading volume, and average daily volatility were used as information related to their price. They showed that data mining techniques (Artificial neural network and support vector machine) were superior methods for detecting stock price manipulation than multivariate techniques (logistic regression and discriminant analysis). According to J. Mongkolnavin and S. Tirapat [6], association rules were applied to detect mark-the-close in intraday trades from the Thai Bond Market Association. Price variation in the market and behavior of investors were integrated to analyze warning signals in real time. The method can produce a list of investors, who perhaps are manipulators. F. Rahnamay Roodposhti, M. Falah Shams and H. Kordlouie [7] used logit model, artificial neural network, and multiple discriminant analysis to create stock price manipulation models in Tehran stock exchange. The performances of three aforesaid models were effective. The selected data were thoroughly studied by runs test, skewness test, and duration correlative test. The events of price manipulation were indicated. The selected data can be divided into two sets: manipulated and non-manipulated companies. The factors that were related to stock price manipulation were defined such as: size of company, P/E ratio, liquidity of stock, status of information clarity, and structure of shareholders. In F. Yang, H. Yang, and M. Yang work [8], logistic regression model was chosen to detect stock price manipulation activities in Shanghai and Shenzhen market that were published as manipulated stocks. They analyzed independent variables based on primary component analysis, which increased performance for forecasting the model. The model was better than the linear regression model. Y. Cao and Y. Li's [9] proposed the Adaptive Hidden Markov Model with Anomaly States (AHMMAS) to detect intraday stock price manipulation activities. The stock tick data were level 2 data from NASDAQ and London stock exchange. The model was tested with simulated data and real market data. The performance evaluation of AHMMAS outperforms other benchmark algorithms such as: Gaussian Mixture Models (GMM), K-Nearest Neighbors Algorithm (kNN), and One Class Support Vector Machines (OCSVM).

This paper differs from the mentioned work in that only level 1 data was used as an input to our neural network models. Level 1 data is accessible by the public, and easy to be acquired. However, we have succeeded only for the pump-and dump cases. For the spoof trading cases, we had to model them with the 2-dimentional Gaussian distribution and use level 2 data.

III. STOCK PRICE MANIPULATION AND ITS MODELS

A. Stock Price Manipulation

In a stock exchange market, many transactions are executed, in which investors can buy or sell ownership of companies. Limit orders describe buy or sell at specific volumes and specific prices. If a buyer and a seller offer the same price, order matching will happen. The bid price is the highest price that a buyer will pay to buy a number of shares of a security. The lowest price, which a seller will receive when he sells a number of shares of a security is also called the ask price. Bid-ask spread [10] represents the difference between the bid price and the ask price. Price manipulation is an action by some investors to control the stock price. This research studied trade-based manipulation, in which manipulators enter crafted buy/sell orders to control the stock price. Trade-based manipulations have many techniques, but are based on the same tactic. A manipulator enters non-bona fide orders into the exchange market system. If the price of the security is going up, some investors will add buy orders and join the rising price. When the manipulator acquires sufficient profit, he withdraws the non-bona fide orders and takes an opposite position immediately.

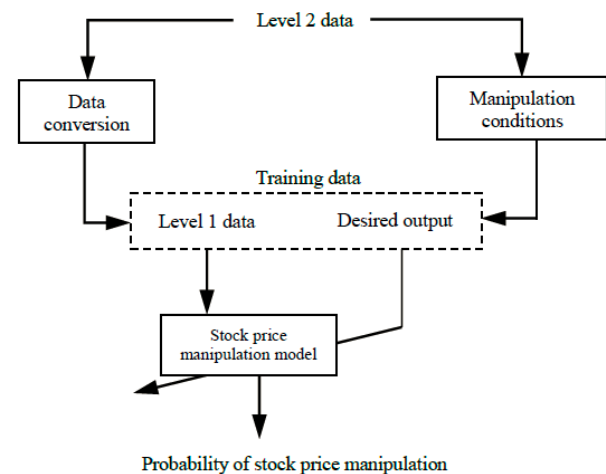### B. Data Level, Depth of Market, and Order Cancellation

Fig. 1. System overview.

The market data contains several information for investors. It can be classified into two levels [11]: level 1 market data and level 2 market data. The difference of the levels of market data depends on what kind of information contained in the data. Normally, most of investors utilize level 1 data for trading, which has less information than level 2 data. Level 1 data is known as current basic market information, which is composed of open, high, low, close price and volume (OHLCV) at different timeframe. More trading information appears in level 2 data, such as the depth of market data (order book), and market participant identifier (MPID). Order book is known as the depth of the market. It displays bid and ask orders that have not been matched. Higher the depth of the market means more details of bid and ask data. MPID is an important information that can refer the one who places

orders in the market system. Order cancellations display the price and the size of orders that investors withdraw. Irregular size of order cancellations in a short period of time can be used as an important indicator for detecting stock price manipulation. The system overview is shown in Fig. 1. First, 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 gives an output value as the probability of stock price manipulation.

C. Pump-and-Dump Model

In pump-and-dump [12], a manipulator generates buy orders, increasing the price and volume of an equity as shown in Fig. 2. In this period, other investors think that the price is going up and join the buy orders. In the dumping state, a manipulator make profits by cancelling all remaining buy orders and executing sell orders at the higher prices than the price before the manipulation. Therefore, the other investors who are not cautious about the non-bona fide orders from the manipulator have bought the equity at higher price than usual.
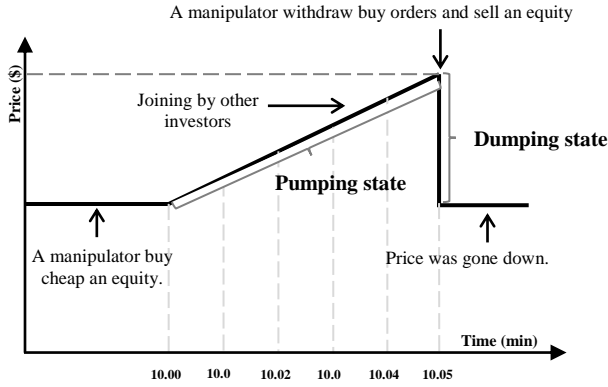

Fig. 2. Pump-and-dump scheme.

We created a pump-and-dump model for forecasting stock price manipulation. Our training set used level 2 data with 10 depth of market, which was processed and separated into two parts. The first part was level 1 data that was used as inputs to our model. The second part was the pump-and-dump events that corresponded to the first part. It was used as desired outputs coupled with level 1 data of the first part to train our neural network model.

In order to give a mathematical definition of pump-and-dump, we defined symbols as follows. *P, V,* and *t* represented price, volume, and time index of a sliding window. Cancellation or deletion volume of buy orders was denoted by $V_{buy}^{cancel}(t)$. $E[V_{buy}^{matched}(t)]$ was the average volume of buy orders that have been matched. $P_{sell}^{max}(t)$ and $P_{sell}^{min}(t)$ were defined as the highest price of sell orders that have been matched, and the lowest price of sell orders that have been matched respectively. Moreover, the highest bid price that have been matched was defined as $P_{bid}^{max}(t)$. The lowest bid price that have been matched is represented by $P_{bid}^{min}(t)$.

We defined three conditions to detect pump-and-dump as below: one condition for the pumping state, and two conditions for the dumping state. First, we checked two conditions in the dumping state: the amount of the order cancellations and the matched orders.

For the first dumping condition that is defined in (1), we classify the activity as a dumping position when the amount of the cancellation and deletion of buys orders $V_{buy}^{cancel}(t)$ is more than threshold1 = 50% of the average volume of buy orders $E[V_{buy}^{matched}(t)]$.

$$dump\_1 = \begin{cases} 1 \; ; \; V_{buy}^{cancel}(t) > (E[V_{buy}^{matched}(t)] \cdot threshold1) \\ 0 \; ; \; otherwise \end{cases} \quad (1)$$

In the second dumping condition, the difference between the highest price of sell orders $P_{sell}^{max}(t)$ and the lowest price of sell orders $P_{sell}^{min}(t+1)$ is more than threshold2, where threshold2 = 0.15%.

$$dump\_2 = \begin{cases} 1 \; ; \; \frac{P_{sell}^{max}(t) - P_{sell}^{min}(t+1)}{P_{sell}^{max}(t)} > threshold2 \\ 0 \; ; \; otherwise \end{cases} \quad (2)$$

We found a dumping state when the first dumping condition and the second dumping condition were true.

$$dump = dump\_1 \wedge dump\_2 \quad (3)$$

If the dumping states are not detected, we identify it as normal states or non-manipulated events. On the other hand, when dumping events are detected, the pumping condition is next verified.

For the pumping condition, after the dumping positions are found, we tested whether the price was going up before that dumping period. Then, we defined the price rising activity as the pumping state when the difference between the highest bid price that have been matched $P_{bid}^{max}(t-1)$ and the lowest bid price that have been matched $P_{bid}^{min}(t-4)$ at the starting period is more than threshold3 = 0.2%.

$$pump = \begin{cases} 1 \; ; \; \frac{P_{bid}^{max}(t-1) - P_{bid}^{min}(t-4)}{P_{bid}^{max}(t-1)} > threshold3 \\ 0 \; ; \; otherwise \end{cases} \quad (4)$$

$$pump - and - dump = dump \wedge pump \quad (5)$$

Therefore, when the aforesaid three conditions are satisfied, we interpreted that the pump-and-dump event has taken place. Its positions are used as a desired output in the training set of our model.

D. Spoof Trading Model

Spoof trading is one of the most popular techniques that a manipulator uses to make a profit. A manipulator starts placing large spoof ask or bid orders into the market to trick other investors that there are high demands for it. The manipulator has no intention for these orders to be matched. The orders will be cancelled when it is about to be matched. These orders are known as passive orders. The volume of these passive sell or buy orders is usually large. The spoof orders can be implemented in two ways: the passive sell price is lower than the current ask price, or the passive price is higher than the current bid price. In Fig 3, a manipulator intended to buy an equity at the price lower than the current ask price. He placed a large-volume order at a passive price, which is lower than the current ask price as shown in the dashed bar in Fig. 3(a). Then, other investors joined into this spoof orders in Fig. 3(b), and they expected that the current

ask price will decrease. Afterwards, a manipulator withdrawed the large spoof sell orders and rapidly bought all remaining sell orders from other investors who were not cautious about this manipulated price as shown in Fig. 3(c).

We constructed a spoof trading model to predict stock price manipulation. We used 10-depth level 2 data to create a training set. The data was processed to separate inputs from level 1 data, and labels (the point where spoof trading occurred) in the same way as the pump-and-dump model. We defined *P*, *V*, and *t* as price, volume, and time index of a sliding window. $P_{sell}^{cancel}(t)$ was described as the price of sell orders that has been cancelled. The price of the last buy order that has been matched was denoted by $P_{buy}^{matched}(t)$. Cancellation volume of sell orders was represented by $V_{sell}^{cancel}(t)$. $V^{matched}(t)$ and $V_{buy}^{matched}(t)$ were defined as the volume of buy and sell orders that have been matched, and the volume of buy orders that have been matched respectively.

(a) Place large orders at new ask price

(b) Joining from other investors

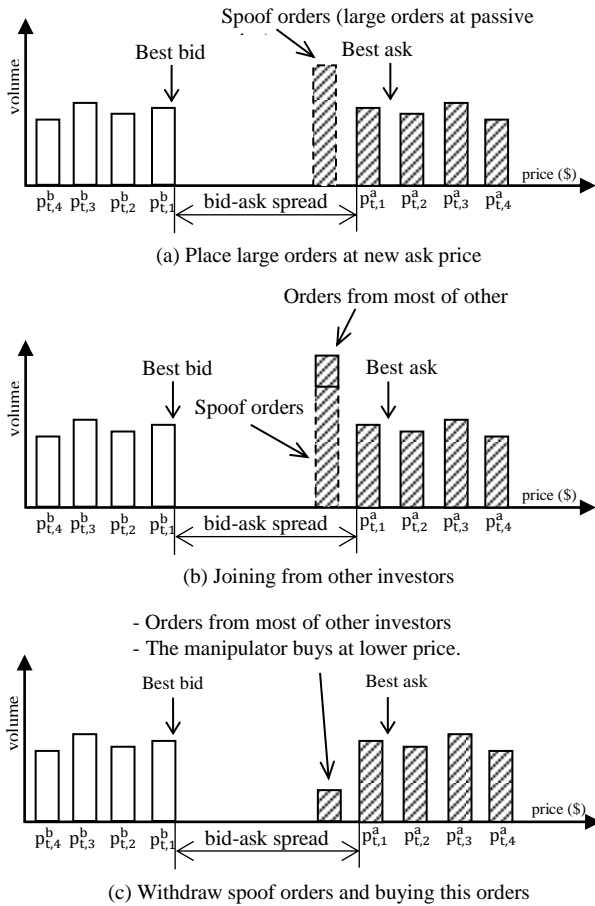(c) Withdraw spoof orders and buying this orders

Fig. 3. Spoof trading scheme. (a) Send large orders at new ask price (b) Joining from other investors (c) Withdraw spoof orders and buying this orders.

We defined three required conditions for spoof trading: order cancellation has its price close to the current bid or ask price, high cancellation volume, and high volume matched at the last buy or sell order.

$$spoof\_1 = \begin{cases} 1 \; ; \left| \frac{P_{sell}^{cancel}(t) - P_{buy}^{matched}(t)}{P_{sell}^{cancel}(t)} \right| < threshold4 \\ 0 \; ; otherwise \end{cases} \quad (5)$$

For the first condition defined in (5), the absolute value of the difference between the price of cancellation sell orders

$P_{sell}^{cancel}(t)$ and the current ask price $P_{buy}^{matched}(t)$ is lower than the threshold4 = 0.5%.

For the second condition, the amount of the cancellation sell orders $V_{sell}^{cancel}(t)$ is higher than threshold5, which is five times of the summation of matched orders since the starting point $\sum_{n=1}^{t-1} V^{matched}(n)$.

$$spoof\_2 = \begin{cases} 1 \; ; V_{sell}^{cancel}(t) > (\sum_{n=1}^{t-1} V^{matched}(n) \cdot threshold5) \\ 0 \; ; otherwise \end{cases} \quad (6)$$

$$spoof\_3 = \begin{cases} 1 \; ; V_{buy}^{matched}(t) > (\sum_{n=1}^{t-1} V^{matched}(n) \cdot threshold6) \\ 0 \; ; otherwise \end{cases} \quad (7)$$

For the third condition, the amount of matched buy orders $V_{buy}^{matched}(t)$ is higher than threshold6 = 50% of the summation of matched orders $\sum_{n=1}^{t-1} V^{matched}(n)$ since the starting point.

When the three conditions are satisfied, we treated them as a spoof trading event. Then, it will be used as a desired output in the training data for our neural network model.

$$spoof\ trading = spoof\_1 \wedge spoof\_2 \wedge spoof\_3 \quad (8)$$

IV. DATA AND MODEL SELECTION

Normally, neural networks are used as classifiers, which discriminate data into classes. We chose neural network as our model, because it can be trained to handle complex problems. In Fig. 4, we constructed a feedforward neural network model for detecting stock price manipulation using open, high, low, close price and volume (OHLCV) of five time steps as inputs. This model consisted of the input layer (25 nodes), the hidden layer (3 nodes), and the output layer (1 node). The model detected pump-and-dump and spoof trading events. The output of the neural network is a binary variable that indicates the probability of manipulation. We defined '1' and '0' as separated events or variable *r* as in (9), which were manipulated point and non-manipulated point respectively.

$$r = \begin{cases} 1 \; ; manipulated\ point \\ 0 \; ; non - manipulated\ point \end{cases} \quad (9)$$

The supervised training was used with back propagation algorithm. We expected to forecast stock price manipulation using level 1 data, which is easier to acquire. Then, the performance evaluation was evaluated by a statistical method.
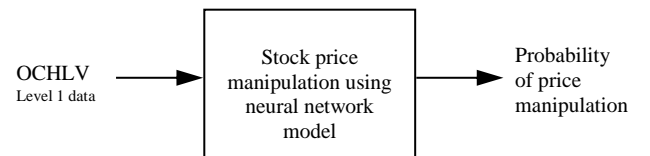

Fig. 4. Overall of neural network model for detecting stock price manipulation.

We chose trading data from NASDAQ (LOBSTER [13]). The data of three companies were gathered: Amazon, Intel, and Microsoft in intraday period on 21st June, 2012. These stocks display order cancellations in level 2 data. The data set was assumed to have price manipulations, because there were high trading volume and large price volatility.

For the pump-and-dump model, we defined a sliding window with 5 points $t = \{1, 2, …, 5\}$. The time frame from each period was 1 minute. The pump-and-dump position of Microsoft company from level 2 data is shown in the top panel of Fig. 5. The pump-and-dump position is detected in the interval from 49 to 53, which matches with 10:49 – 10:53 am. The cancellation volume at the dumping state is higher than matched volume as shown in the bottom panel of Fig. 5. We used a neural network model to train these pump-and-dump events. The input was level 1 data which has no information about order cancellations. The sample set was divided into two parts: the training set for training the neural network model and the test set for verifying the training results.
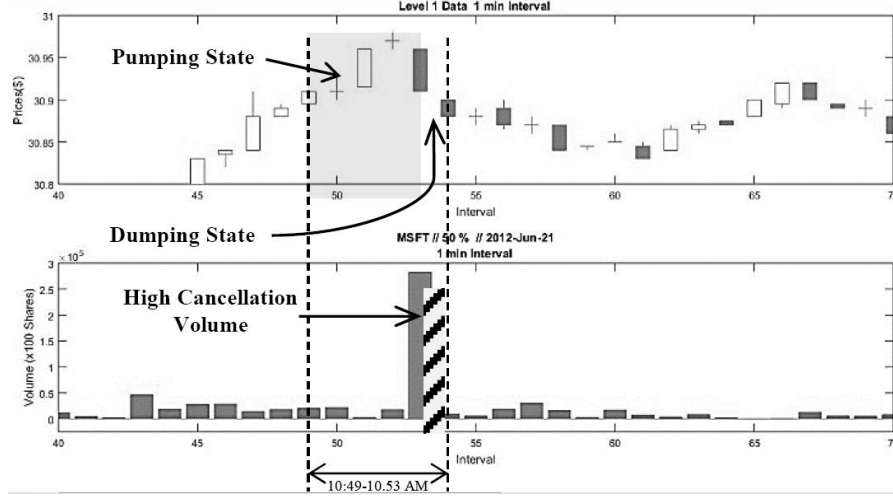

Fig. 5. Pump-and-dump position of microsoft company.

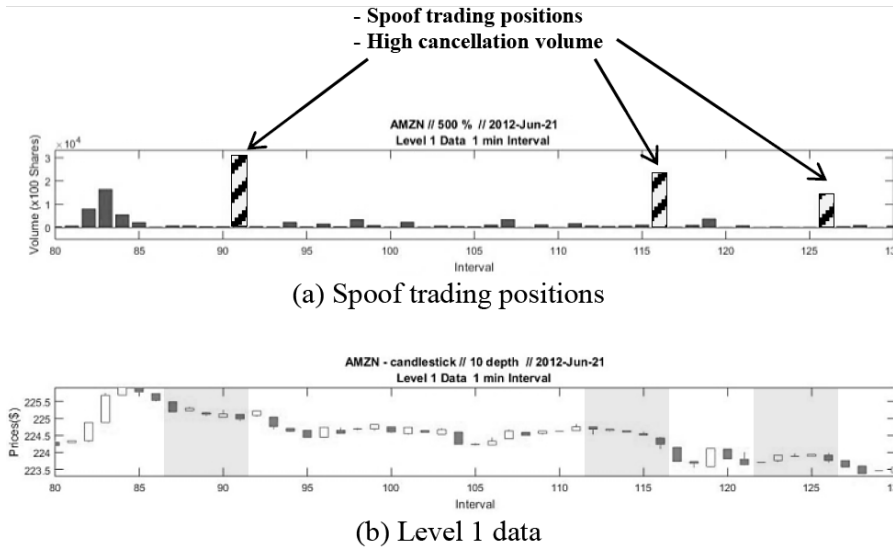(a) Spoof trading positions

(b) Level 1 data

Fig. 6. Spoof trading positions of Amazon Company: (a) Spoof trading positions (b) Level 1 data.

For spoof trading model, we defined the length of a sliding window as 5 points $t = \{1, 2, …, 5\}$ and time frame equaled 1 minute, which was similar to the pump-and-dump model. Fig. 6 shows the spoof trading positions of Amazon company with the price pattern from level 1 data. We display spoof trading positions in Fig. 6(a) at the wide upward diagonal bar. The spoof trading positions had high cancellation volume compared to the matched volume. Fig. 6(b) shows the price pattern from level 1 data. Although the spoof trading positions were found, we could not identify them using the price pattern from level 1 data.

V. EXPERIMENT AND RESULT

After we got pump-and-dump and spoof trading models, the performance of both models were tested and evaluated. Level 2 data was transformed to level 1 data (OHLCV) as

inputs for forecasting stock price manipulation. We selected 22 sets for sample data. These data were divided into two groups: non-manipulated stock 50%, and manipulated stock 50%. The model was trained 10 times per model, and we chose the one that had the smallest mean square error. We used the maximum number of training epochs and the target error weight as 500 and 0 respectively. After the training was completed, we tested the model using leave-one-out cross validation method [14]. We chose this method, because there were relatively fewer manipulated points compared to non-manipulated points.

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

The test results of pump-and-dump and spoof trading model are shown in Table I. The mean square error was computed by the square difference between the output of the model and the target value. The table also shows false positive and false negative error separately. False negative refers to misdetection: manipulated events detected as

non-manipulated events. False positive reflects false alarm rate: non-manipulated events detected as manipulated events. From the table, the neural network model for detecting spoof trading has significantly more average mean square error, false positive error, and false negative error than those of pump-and-dump.
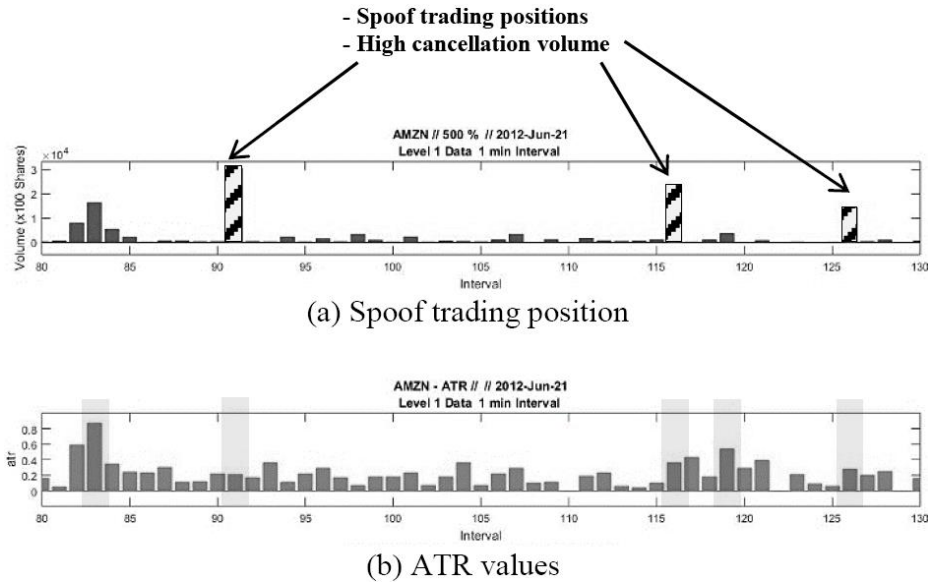

(a) Spoof trading position

(b) ATR values

Fig. 7. Price manipulation detection of Amazon Company: Spoof trading with ATR values. (a) Spoof trading position (b) ATR values.

We conclude that the price pattern from level 1 data cannot discriminate manipulated cases from normal cases in the spoof trading model. We further explored why the neural network training had an unsatisfied result. The average true range (ATR) was chosen and compared with the volumes of order cancellations, because it can measure the level of oscillation of the stock's price [15]. So, it is mostly applied for indicating a volatility of stock. The ATR can be calculated from a moving average of true ranges. The true ranges can be calculated from the maximum of three difference values: distance between current high and current low, distance between previous close and current high, and distance between previous close and current low. The period of moving average was 14. We expected that the spoof trading positions should have high ATR values. The ATR values of Amazon company are shown in Fig. 7(b). The graph shows that 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.

We assumed that the spoof trading occurred when there are high cancellation volume, high matched volume, and low delta price (the difference price between cancellation price and the best ask/bid price). In order to create an effective model for spoof trading detection, we used a technique similar to anomaly detection. One of the common models for anomaly detection is the two-dimensional Gaussian distribution. Normal orders were modeled as a 2D Gaussian distribution, where spoof orders were treated like anomaly. The two parameters that are the important features of spoof trading are the cancellation volume of the spoof orders and the matched volume from subsequent orders to take a profit. Trade data were collected in each time period with the length of a sliding window as t = {1, 2, 3, ..., n}. The time frame

from each period was 1 minute. To reduce the model into 2 dimensions, the samples that have high values of the delta price were discarded first, leaving with only samples that have low delta price. The remaining samples were plotted with cancellation volume as the horizontal axis and matched volume as the vertical axis.

Fig. 8 shows the decision boundary with the threshold of cancellation volume and the threshold of matched volume. The decision boundary was defined as a combination of two times standard deviation line from normal distribution, the threshold of cancellation volume and the threshold of matched volume. A sample is considered to be a spoof trading event when it is located outside the boundary line of two times standard deviation, and has both cancellation volume and matched volume exceed the thresholds.
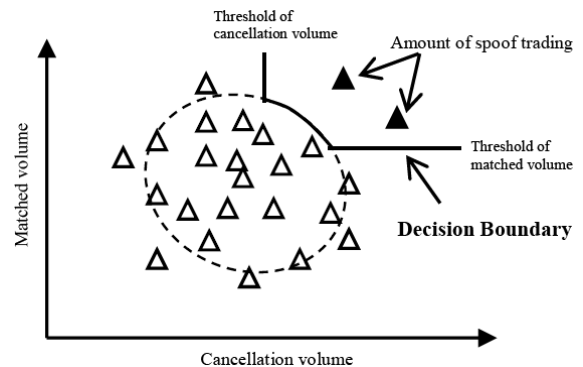

Fig. 8. Decision boundary of spoof trading model.

In the first step, we calculated the difference price between cancellation price and the best ask/bid price (called the delta price, ΔP). We defined $P^{cancel}(t)$ as the cancellation price of buy/sell orders. $P^{matched}(t)$ was the matched price of buy/sell orders. Current bid/ask price is represented by $P(t)$. $V^{cancel}(t)$ and $V^{matched}(t)$ were denoted as cancellation

volume of buy/sell orders, and matched volume of buy/sell orders respectively.

The delta price of each time frame can be computed by (12). Next, we discarded the samples with high values of the delta price, when the delta price is more than threshold = 0.1%.

$$P_{TF} = \sum_{t=1}^{n} \frac{((|P^{cancel}(t) - P(t)|)V^{cancel}(t)))}{V^{cancel}(t)} \quad (10)$$

$$E[P] = \sum_{t=1}^{n} \frac{P^{matched}(t)}{n} \quad (11)$$

$$\Delta P = \frac{P_{TF}}{E[P]} \quad (12)$$

Next, the average of cancellation volume and matched volume were calculated as follows.

$$E[V^{cancel}] = \sum_{t=1}^{n} \frac{V^{cancel}(t)}{n} \quad (13)$$

$$E[V^{matched}] = \sum_{t=1}^{n} \frac{V^{matched}(t)}{n} \quad (14)$$

The probability density function [16] of a 2-dimentional Gaussian distribution is given by (15). $x \sim N(\mu, \Sigma)$ is a vector-valued variable with mean μ and covariance matrix Σ of the cancellation volume and the matched volume.

$$p(x; \mu, \Sigma) = \frac{1}{(2\Pi)|\Sigma|^{\frac{1}{2}}} exp(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)) \quad (15)$$

$$x = \begin{bmatrix} x_{cancel} \\ x_{matched} \end{bmatrix} \quad (16)$$

$$\mu = \begin{bmatrix} \mu_{cancel} \\ \mu_{matched} \end{bmatrix} \quad (17)$$

$$\Sigma = \begin{bmatrix} \Sigma_{cancel,cancel} & \Sigma_{cancel,matched} \\ \Sigma_{matched,cancel} & \Sigma_{matched,matched} \end{bmatrix} \quad (18)$$

We used level 2 data as an input of the model. We calculated the mean, standard deviation, and covariance matrix of the matched volume and the cancellation volume of each company. Table II and III show mean and standard deviation values of buy orders and sell orders in three companies respectively.

Covariance matrices show the relationship of the changed of two variables. In this case, the matched volume and the cancellation volume of three companies were identified. A positive covariance means that the matched volume and the cancellation volume move together. A negative covariance means both of them move inversely. Table IV and V show covariance of buy orders and sell orders respectively. The three companies appear to have similar behaviors between the matched volume and the cancellation volume. Thus, these covariance values are positive.

TABLE II: RESULTS OF STATISTICAL METHOD: BUY ORDERS

Company	$\mu_{matched}$	$\sigma_{matched}$	μ_{cancel}	σ_{cancel}
	Buy orders			
AMZN	1129.43	1371.84	16838.59	10126.06
INTC	15697.75	18173.12	173927.84	132346.16
MSFT	13912.75	21394.40	185108.42	164045.34

TABLE III: RESULTS OF STATISTICAL METHOD: SELL ORDERS

Company	$\mu_{matched}$	$\sigma_{matched}$	μ_{cancel}	σ_{cancel}
	Sell orders			
AMZN	919.94	1005.41	15016.17	10112.42
INTC	12637.80	11929.04	154048.38	125823.51
MSFT	13912.75	21394.40	185108.42	164045.34

TABLE IV: COVARIANCE OF BUY ORDERS

Company	$\Sigma_{cancel,cancel}$	$\Sigma_{cancel,matched}$	$\Sigma_{matched,cancel}$	$\Sigma_{matched,matched}$
	Buy orders			
AMZN	102537046.20	5086943.032	5086943.03	1881942.619
INTC	17515506875.88	1182204877.41	1182204877.41	330262198.51
MSFT	26910874832.70	2071303627.52	2071303627.52	457720254.74

TABLE V: COVARIANCE OF SELL ORDERS

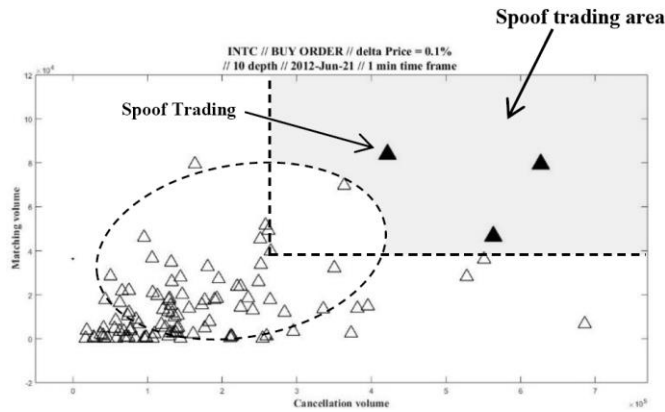| Company	$\Sigma_{cancel,cancel}$	$\Sigma_{cancel,matched}$	$\Sigma_{matched,cancel}$	$\Sigma_{matched,matched}$
	Sell orders			
AMZN	102261132.66	4199180.897	4199180.90	1010852.491
INTC	15831556050.62	736503752.84	736503752.84	142301967.45
MSFT	34968729699.40	1773363548.86	1773363548.86	385601990.27

Fig. 9. Spoof trading points of Intel Company using 2-dimentional Gaussian model.

The spoof trading position was verified by 2-dimentional Gaussian method. Fig. 9 shows three spoof trading points of the Intel company. The thresholds of the matched volume and the cancellation volume were used to define a decision boundary. They were defined based on the standard deviation of 95% confidence interval. Cumulative distribution function (CDF) describes a probability that a variable is less than or equal to the independent variable of the function. A point that has CDF value more than 95% confidence interval is defined as a manipulated point.

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 activities 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. Order cancellations are important indicators to identify price manipulation. A set of less-detailed in 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. Therefore, we had to create a model with the 2-dimentional Gaussian distribution and use level 2 data for detecting spoof trading. We assumed that the spoof trading occurred when three conditions were met: low value of the difference price between cancellation price and current bid/ask price, high cancellation volume, and high matched volume. We show that a spoof trading position in the sample data can be detected when CDF is more than 95% confidence interval, and the matched volume and the cancellation volume are more than the thresholds.

REFERENCES

[1] T. Leangarun, P. Tangamchit, and S. Thajchayapong, "Stock price manipulation detection using a computational neural network model," in *Proc. The 8th International Conference on Advanced Computational Intelligence*, Chiangmai, Thailand, February 14-16, 2016, pp. 337-341.
[2] F. Allen and D. Gale, "Stock-price manipulation," *The Review of Financial Studies*, vol. 5, no. 3, pp. 503-529, 1992.
[3] R. A. Jarrow, "Market manipulation, bubbles, corners, and short squeezes," *Journal of Financial and Quantitative Analysis*, vol. 27, pp. 311-336, March 1992.
[4] A. K. Rajesh and G. J. Wu, "Stock market manipulation-theory and evidence," University of Michigan Business School Working Paper, 2004.
[5] H. Öğüt, 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, pp. 11944-11949, 2009.
[6] J. Mongkolnavin and S. Tirapat, "Marking the close analysis in Thai bond market surveillance using association rules," *Expert Systems with Applications*, vol. 36, no. 4, pp. 8523 – 8527, 2009.
[7] F. Rahnamay, H. Kordlouie *et al.*, "Forecasting stock price manipulation in capital market," *World Academy of Science, Engineering and Technology,* vol. 56, pp. 151-161, 2011.
[8] F. Yang, H. Yang, and M. Yang, "Discrimination of China's stock price manipulation based on primary component analysis," *Behavior, Economic and Social Computing*, pp. 122-127, 2014.
[9] 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, pp. 310-330, 2015.
[10] K. Menyah and K. Paudyal, "The components of bid-ask spreads on the London stock exchange," *J. Banking, Finance*, vol. 24, no. 11, pp. 1767–1785, 2000.
[11] About.com, 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
[12] D. Cumming, S. Johan, and D. Li, "Exchange trading rules and stock market liquidity," *J. Financial Economics*, vol. 99, no. 3, pp. 651–671, 2011.
[13] LOBSTER, Atlanta, GA, USA. (2012). Limit Order Book System [Online]. Available: http://www.lobster.wiwi.hu-berlin.de
[14] S. Arlot, "A survey of cross-validation procedures for model selection," *Statistics Surveys,* vol. 4, pp. 40-79, 2010.
[15] S. Yamanaka, "Average true range," *Stocks & Commodities Magazine*, vol. 20, no. 3, pp. 76-79, 2012.
[16] M. Słoński, "Gaussian mixture model for time series-based structural damage detection," *Computer Assisted Methods in Engineering and Science*, vol. 19, pp. 331-338, 2012.

Teema Leangarun received her B.Eng. degree in control systems and instrumentation engineering from King Mongkut's University of Technology Thonburi, Bangkok, Thailand in 2014, where she is currently a master's student. Her research topic involves computational finance and anomaly detection with the King Mongkut's University of Technology Thonburi, Bangkok, Thailand.

Poj Tangamchit received his Ph.D. degree in electrical and computer eng. in 2003 from Carnegie Mellon University, USA. He is currently an associate professor at the department of Control Systems and Instrumentation Engineering at King Mongkut's University of Technology Thonburi, Bangkok, Thailand. His research involves robotics, artificial intelligence, and computational finance.

Suttipong Thajchayapong received his Ph.D. degree in electrical and electronic engineering from Imperial College London, U.K. He is a researcher with the National Electronic and Computer Technology Center, National Science and Technology Development Agency (NSTDA). His research interests include anomaly detection, wireless networks, and intelligent transportation systems.