

# A Hidden Markov Model with Abnormal States for Detecting Stock Price Manipulation

Yi Cao, Yuhua Li, Sonya Coleman, Ammar Belatreche, T.M. McGinnity  
Intelligent Systems Research Centre  
University of Ulster, UK

**Abstract**—Price manipulation refers to the act of using illegal trading behaviour to manually change an equity price with the aim of making profits. With increasing volumes of trading, price manipulation can be extremely damaging to the proper functioning and integrity of capital markets. Effective approaches for analysing and real-time detection of price manipulation are yet to be developed. This paper proposes a novel approach, called Hidden Markov Model with Abnormal States (HMMAS), which models and detects price manipulation activities. Together with the wavelet decomposition for features extraction and Gaussian Mixture Model for Probability Density Function (PDF) construction, the HMMAS model detects price manipulation and identifies the type of the detected manipulation. Evaluation experiments of the model were conducted on six stock tick data from NASDAQ and London Stock Exchange (LSE). The results showed that the proposed HMMAS model can effectively detect price manipulation patterns.

**Index Terms**—Hidden Markov Model, Anomaly Detection, Capital Market Price Manipulation

## I. INTRODUCTION

Since the financial crisis in 2008, new aspects of risk management have been identified by regulators around the world. Surveillance of financial exchange markets to prevent market abuse activities is one of the emerging needs in the financial sector. Financial market abuse can be split into three general categories. These include information based manipulation, where false information is released to affect equity prices; action based manipulation, where a person's actions change the real value of an asset; and trade based manipulation where a manipulator attempts to influence a stock by simply buying and selling [3]. Trade based market abuse is generally characterized by price manipulation, where the manipulation targets equity bid/ask prices only [3]. Most existing literature consists of empirical studies of the phenomenon and the market response to the manipulation [6], [10], [11], [13]. Very few address an analysis of the features of different price manipulation strategies [2] and corresponding detection approaches.

Intelligent techniques for monitoring trading activities are required in order to provide early warnings. This could include identifying unusual patterns within the equity bid/ask prices in real time, as a pre-cursor to market abuse activities. This paper proposes a new approach for price manipulations detection in stock exchange market. In this approach, features of stock bid/ask price are extracted by wavelet theory and then modelled by HMMAS based approach.

The remainder of the paper is organised as follows: Section II introduces price manipulation and presents a brief review

of price manipulation detection methods. The analysis of key patterns and features of typical price manipulations are discussed in Section III. Section IV presents the proposed HMMAS approach for detecting market abuse patterns, while its performance is evaluated in Section V. Finally, Section VI concludes the paper and discusses potential improvements and future work.

## II. PRICE MANIPULATION AND ITS DETECTION

### A. Price Manipulation

In capital markets, limit orders are the instructions to buy or sell equities. They indicate the trading intention of investors to buy or sell volumes of a specific equity at a specific price. The trade occurs on eligible orders with matched prices. The outstanding unmatched limit orders are recorded in the order books and the highest buy price of those orders determines the bid while the lowest price the seller is willing to accept decides the ask. The difference between the bid and ask price is defined as bid-ask spread.

Price manipulation can be carried out in many forms. The primary form is called “ramping” (or gouging [5] or momentum ignition [17]), which describes the activities of a broker who enters a purchase/sell order, usually called “spoofing order” at higher/lower prices to create the appearance of active interest in a security, followed by an order on the opposite side of the order book, and subsequently followed by a withdrawal of the first order. Once the first purchase/sell order is created, a price upward/downward movement is expected from the investor and the latter actions follow only when there is an enough potential profit for the investor. Ramping has roughly the same effect as the other two forms of price manipulation, colloquially known as “pump&dump” and “capping&pegging” schemes [5]. In both of these cases, exchange participants make a profit by generating a significant increase in price for a security and carry out a quick flip at the higher price to remove the already held positions in the former scheme, and to exercise the held derivative/option of the manipulated security of the latter. If profit is exploited in another market, the form of price manipulation is then termed cross order/market.

In September 2012, an analogous type of “Ramping” was reported and documented by the Financial Industry Regulatory Authority in the USA [12]. In this case, a liquidity-enhancing strategy, “quote stuffing”, was used to create the fictitious impression of buying interest on the bid side by quickly submitting and cancelling a mass of fake orders. As discussed

by NANEX [12], quote stuffing is the dominant manipulation strategy in stock exchanges of Copenhagen, Helsinki, London, Oslo and Stockholm and NASDAQ (See also in[17]).

All of above mentioned different forms of market abuse reduce to the same manipulations tactic: artificially pushing up/down the price of a security and taking advantage of the increased/decreased price. Profits are made in distinct ways as shown in Figure 1.

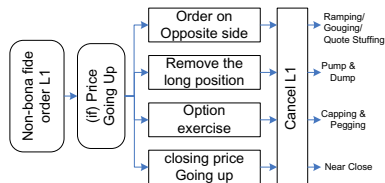
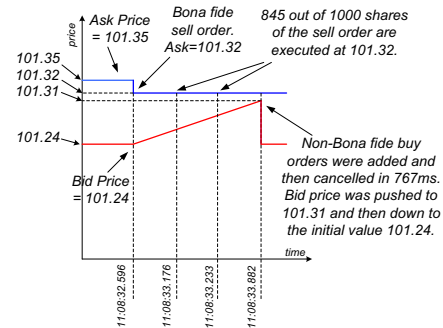


Fig. 1: Different types of price manipulation

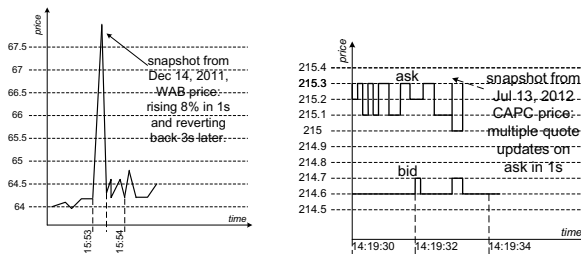
### B. Price Manipulation Detection

The automatic detection of price manipulation has been less studied in contrast to the volume of theoretical and empirical work on manipulation activities. Two computational approaches, logistic regression with artificial neural network and support vector machine, have been studied and compared for their use in detecting trade based manipulation within the emerging Istanbul Stock Market [13]. Similar work has been done by modelling the returns, liquidity and volatility of the equity price using linear and logistic regression [6]. Evaluations and comparisons of different techniques have been presented [6], [13], but the studies lack any reliable quantitative analysis of the relationship between the manipulation tactic and the relevant features like return, trading volume, liquidity and volatility. An association rule, has been proposed to detect closing price manipulation in the Thai Bond Market [11]. The approach is based on the assumption that trading time should be random regardless of the trader. It is stated as a supplementary to the existing surveillance system within the Thai Bond Market for identifying a specific type of trade based manipulation [11] rather than a general detection approach. A market close “ramping” detection algorithm developed by Smarts Group International (a surveillance technique provider), and applied by NASDAQ OMX is described in [2]. But this work mainly focuses on the relationship study of the market efficiency and the manipulations detected by the Smarts Group algorithm rather than the analysis of detection algorithms. The Smart Group algorithm triggers the alert of market close “ramping” if the percentage difference between the closing price and the price 15 minutes prior is greater than a certain threshold, which is set to a 99% histogram distribution cut-off of the historical price change during the benchmark period. The straight forward market proven detection approach from Smarts Group is the industry benchmark for our study.

Generally speaking, work to date mainly focuses on the detection of manipulation in pre-recorded datasets according to assumptions based on empirical studies. To the best of our knowledge, only limited efforts appear to have been made to



(a) Sawtooth Pattern



(b) Pulse Pattern

(c) Square Pattern

Fig. 2: Price Manipulation Patterns. (Reproduced based on data from [17])

carry out in-depth analysis of the strategic behaviour of price manipulation tactics [2], [10], and very little effort has been made to develop computational models of price manipulation strategies. The latter is the focus of this paper.

### C. Problem Formulation

According to the price manipulation scenarios, the typical patterns of bid/ask price fluctuation during a manipulation process have been reported by Credit Suisse and Nanex and are illustrated in Fig. 2.

In Fig. 2(a), the successively increasing “non-bona fide” buy orders pushed the bid price going up from 101.24 to 101.31 and the cancellation of them dropped down it back to the origin level thus making a sawtooth pattern. The increasing bid lured a trade to occur before the cancellation. The “sawtooth” is a typical pattern in bid/ask during “quote stuffing” [12] and is also discovered as manipulation by High Frequency Trading in [17], [9].

Unlike the “quote stuffing” strategy, where the bid price is increased along a slope, a “ramping” tactic usually enters one “spoofing order” at higher/lower prices to create the appearance of active interest on bid/ask sides of a security. Through this, the manipulated price can be moved up/down a percentage and revert to its previous level. The percentage of price movement is usually small compared to the long term price and is illustrated as a small pulse or square fluctuation as illustrated in Fig. 2(c) [18]. However, it can also be significant

as illustrated in Fig. 2(b). Although this unusual pulse rise is rare since most exchange markets such as Euronext NSC suspend continuous trading if prices change by more than 2% [7], it is also highlighted as one of the typical manipulation cases in [9].

### III. CHARACTERISATION OF PRICE MANIPULATION

The characterisation of irregular price manipulation patterns is key to their detection. The extraction of reliable features is crucial to the design of detection algorithms. The price manipulation detection system proposed in this paper is comprised of a feature extraction component and a detection model, which is trained based on normal data samples defined by the features extracted from the equity bid/ask price by the feature extraction component, which is discussed as the follows.

#### A. Common Features

Typical price manipulations are usually followed by either a remarkable bid/ask price change or special patterns (square or sawtooth) of short term small fluctuations. Correspondingly, two intuitive features of the manipulation activities can be extracted:

- 1) The amplitude of price (bid/ask) fluctuation:

$$\Delta P = \frac{P(t+1) - P(t)}{P(t+1)} \quad (1)$$

where  $P(t)$  is the ask/bid price at time  $t$

- 2) The rate of price (bid/ask) fluctuation (represented as first order derivative of  $P(t)$  with respect to  $t$ ):

$$\frac{dP(t)}{dt} = \lim_{\Delta t \rightarrow 0} \frac{P(t + \Delta t) - P(t)}{\Delta t} \quad (2)$$

where  $\Delta t$  is the time interval between the changes of ask/bid prices.

#### B. Additional Features

As illustrated in Fig. 2(a), the “sawtooth” patterns show mere changes of original equity price  $(101.31 - 101.24)/101.24 = 0.069\%$ . Similarly, “square” patterns are around 0.047% changes as illustrated in Fig. 2(c).

According to the capital market microstructure theory, the nature of the equity price movement follows the mean-reverting feature, the tendency to randomly oscillate away from, and over time back towards an “equilibrium” price level determined by the long-term mean of the equity.

The price fluctuations triggered by the manipulation strategies can be considered as short-termed oscillations with small amplitudes around the “equilibrium” level of the price (0.069% in Fig. 2(a) and 0.047% in Fig. 2(c)). Such tiny fluctuations are usually considered as the “contamination” of financial data [8]. The “noisy” part is where the manipulation patterns occur. Therefore, retrieving “noisy” short-term oscillation information from the price is crucial for detecting particular patterns. The wavelet analysis feature of separating the low and high frequency components of a signal while localising the high frequency components in time enables the application in the fields of economics and finance for de-noising financial time

series [8]. The power of the wavelet method in analysing frequency components of a signal and localising components in time could be utilised for feature extraction.

a) *The Wavelet Approach*: Signals  $y(t)$  can be decomposed into a combination of some scaling functions  $\phi_{M,n}(t)$  and wavelet functions  $\psi_{m,n}(t)$ , each factored by their corresponding approximation coefficients  $S_{M,n}$  and detail coefficients  $T_{m,n}$  [1].

$$y(t) = \sum_{n=-\infty}^{\infty} S_{M,n} \phi_{M,n}(t) + \sum_{m=-\infty}^M \sum_{n=-\infty}^{\infty} T_{m,n} \psi_{m,n}(t), \quad (3)$$

Once the signal is decomposed, the detail coefficients vectors can be filtered for processing the signals since it represents “high-frequency” components of the signal. In our approach, the traditional “hard thresholding” method used in signal denoising [8] is applied inversely, so removing the “equilibrium” level and keeping the “noisy” short term oscillation information is carried out with “hard” thresholding and the wavelet coefficients outside the thresholds are set to zeros:

$$T_{m,n} = \begin{cases} =T_{m,n} & \text{if } |T_{m,n}| \leq \lambda \\ =0 & \text{if } |T_{m,n}| > \lambda \end{cases} \quad (4)$$

where  $n$  is the shift parameter which controls the translation of the scaling and wavelet functions;  $\lambda$  is the fixed threshold and  $T_{m,n}$  are the detail coefficients at level  $m$ . Based on the above method, the wavelet procedure consists of the following steps:

- 1) **Decomposition**. For a given data series  $y(t)$ , calculate the discrete wavelet transform at a certain level  $m$  to obtain the wavelet coefficient vectors  $\{T_{m,n}, T_{m-1,n}, \dots, T_{1,n}\}$  and the scaling coefficient vector  $S_{m,n}$ .
- 2) **Thresholding**. Select a threshold  $\lambda$  and filter the wavelet coefficient vectors  $T_{m,n}, T_{m-1,n}, \dots, T_{1,n}$  through the “hard” thresholding. After the thresholding, a set of filtered wavelet coefficients  $\{\hat{T}_{m,n}, \hat{T}_{m-1,n}, \dots, \hat{T}_{1,n}\}$  is produced.
- 3) **Reconstruction**. Recombine the  $\hat{y}(t)$  using the filtered wavelet coefficients  $\{\hat{T}_{m,n}, \hat{T}_{m-1,n}, \dots, \hat{T}_{1,n}\}$ .

An experiment of the wavelet approach is shown in Fig. 3 (right parts of sub-figures). The equity price, illustrated in Fig.3, is filtered by the Symlets wavelet at level 5. The smoothed “equilibrium” values are removed from the original price and the short term oscillation is remained.

b) *Gradient Approach*: As discussed in Section III-A, the price change rate is significant feature during the price manipulation period. To capture this, the gradient of the original price and the filtered short term fluctuation are both calculated. Among the methods for calculating gradients, finite-difference is chosen and its central difference form is used for approximately estimating the first-order derivatives of the equity price time series.

An example of gradients of the filtered short term oscillation and the equity price are shown in Fig. 3(c) and 3(d) (right parts of sub-figures) respectively.

#### IV. DETECTION MODEL

In addition to the previously discussed feature extraction process, the algorithm for detecting the price manipulation activities comprises two parts: (1) learning the probability density functions of the extracted features; (2) modelling the transition probabilities among the states determined by the learned distributions.

##### A. Gaussian Mixture Model and Hidden Markov Model

A common approach for modelling the probability density of an observed variable is to approximate its unknown density with a Gaussian Mixture Model (GMM) [4]. A GMM is a weighted sum of  $M$  component Gaussian densities as given by the equation,

$$p(\mathbf{x}) = \sum_{i=1}^M w_i g(\mathbf{x} | \boldsymbol{\mu}_i, \Sigma_i), \quad (5)$$

where  $\mathbf{x}$  is a  $D$ -dimension continuous-valued data vector,  $w_i, i = 1, \dots, M$ , are the mixture weights, and  $g(\mathbf{x} | \boldsymbol{\mu}_i, \Sigma_i), i = 1, \dots, M$ , are the component Gaussian densities. Each component density is a  $D$ -variate Gaussian function of the form,

$$g(\mathbf{x} | \boldsymbol{\mu}_i, \Sigma_i) = \frac{\exp\left\{(-1/2)(\mathbf{x} - \boldsymbol{\mu}_i)^T \Sigma_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i)\right\}}{(2\pi)^{D/2} |\Sigma_i|^{1/2}}, \quad (6)$$

with mean vector  $\boldsymbol{\mu}_i$  and covariance matrix  $\Sigma_i$ . The mixture weights satisfy the constraint  $\sum_{i=1}^M w_i = 1$ . The extracted features are firstly studied by k-means clustering algorithm for determining the number of components of the probability density function (PDF), which is then modelled by GMM.

Since the availability of the market manipulation samples in capital market is much less than that of normal trading samples, the training dataset contains only normal bid/ask price information. The threshold setting strategy is borrowed from the industry benchmark algorithm from Smart Group [2]: the anomaly threshold is set to the 99% cumulative distribution of four extracted features of the bid/ask price. This is to identify the highest and lowest 0.5% frequent values for each feature.

To further capture the temporal dynamics between different features of the bid/ask time series, a Hidden Markov Model (HMM) [14] presents an ideal model for analysing time-series data by a state-based form and modelling the temporal dynamics among the “hidden” states of the Markov chain

Since the lack of manipulation data, as the approaches in [19], a HMM model is applied on the normal data only: the features of normal activities are modelled and the test cases “against” the model based on the selected threshold are reported as abnormal. However, two problems arise in this approach when applied to price manipulation detection. HMM is usually used to model 1-dimensional time series rather than the multiple features. Anomaly detection using HMM usually lacks the ability for recognizing the anomaly type and the probability density of being that type.

In order to address these challenges, an improved HMM with abnormal states is developed and presented in the following section.

##### B. Hidden Markov Model with Abnormal States (HMMAS)

For the bid/ask price of an equity with  $T$  points, four features can be extracted as discussed in Section III-B. Let  $F_t(t = 1, \dots, T)$  be the feature vector at the  $t^{\text{th}}$  point in time. Based on the previous discussion, the pre-processed four features at time  $t$  are  $F_t = [F_t^o, F_t^{og}, F_t^s, F_t^{sg}]$ , which are described as follows:

- 1) Movement of the original price:  $F_t^o$ ;
- 2) Gradient of the of original price:  $F_t^{og}$ ;
- 3) Movement of the short-term fluctuation (filtered by wavelet):  $F_t^s$ ;
- 4) Gradient of the short-term fluctuation:  $F_t^{sg}$ ;

As discussed in Section IV-A, the probability density functions (PDF) of the four features are learned separately by the Gaussian Mixture Model as described in Equation 5. Changes in the observation distribution can be detected by testing which mixture component in the PDF was most likely to have generated a given value. Each component of the PDF is designated as a sub-state of one feature. This gives a simplified “state view” of time-series data.

To designate “abnormal” states based on this “state view”, thresholds for “abnormal” regions are set on each learned PDF. As discussed in Section IV-A, “thresholds for abnormality” are set at the 99% cumulative distribution cut-off. This heuristic method is also applied in one-class Support Vector Machine [15], where the detection boundary is usually set to include most (for example 99%), but not all, training data to avoid overfitting. Hence, the thresholds for each feature can be set such that regions within the thresholds are considered “normal”. By doing this, the cases in the regions outside the thresholds are not simply assumed to be “abnormalities” but rather generate “dummy abnormal sub-states” for the Hidden Markov States (as illustrated in Fig. 4), which were extended to comprise “anomaly” (“manipulation”) states that can not be achieved in traditional HMMs.

Fig. 3 shows an example of the PDFs’ illustration of the four features of the time series data. In Fig. 3(a), the learned PDF of stock price is composed of three Gaussian distribution components and the thresholds are set to 0.5% and 99.5% cumulative probability. Hence, 4 sub-states, which consist of one “abnormal” sub-state (outside the thresholds) and three “normal” sub-states, are generated by the 3-component PDF with the thresholds. Similarly, 2 sub-states (one “abnormal” and one “normal”) are generated for the other three 1-component PDFs. The sub-states of four features are then quantified and combined as the hidden states of observed bid/ask price so that  $4 * 2 * 2 * 2 = 32$  hidden states are generated.

Only three among the 32 states are considered completely “normal” since each feature is within the “normal” region. The other 29 states are considered as “abnormal” since at least one feature is within the “abnormal” region. The manipulation types in real-life can be deduced using combinations of the different “abnormal” states.

According to the discussion in Section II, typical price manipulation is associated with combinations of different

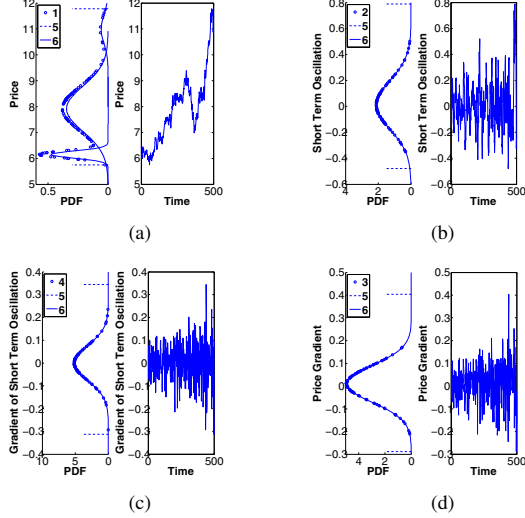


Fig. 3: Example of the features learned by GMM. (a) Equity Price (right) and its PDF (left); (b) Short Term Oscillation (right) and its PDF (left); (c) Gradient of Short Term Oscillation (right) and its PDF (left); (d) Gradient of Original Equity Price (right) and its PDF (left) (Legend: **1**: PDF of Price; **2**: PDF of price oscillation; **3**: PDF of gradient of price; **4**: PDF of price oscillation gradient; **5**: Threshold; **6**: Components of PDF; )

extracted features. The types, “Quote Stuffing”, “Momentum Ignition” and “Gouging” depend on instantaneous sweep of order books and are essentially associated with significant changes in gradients for short term oscillation and the original price rather than the large price fluctuation. Additionally, “Ramping”, “Pump&Dump”, “Capping&Pegging” and “Near Close” are associated with significant changes of the original price or short term fluctuations but not necessarily with the gradients. Consequently, the 29 “abnormal” states can be combined into three primary manipulation states, namely: “Quote Stuffing”, “Ramping” and “Other Abnormal”. Abnormal states where only the gradient features are abnormal fall into the manipulation state “Quote Stuffing” while those where only the price features are abnormal are “Ramping”. The manipulation state “Other Abnormal” is associated with cases where all features are abnormal. By this merging, the 32 states are further simplified into a six-state view of the observed bid/ask price with three normal states and three abnormal states. These states thus provide a complete spectrum of the bid/ask changes triggered by different trading behaviours as illustrated in Fig.4.

In summary, the proposed computational model for detecting price manipulation activities is comprised of three parts: (1) extracting the four features of the bid/ask price using wavelet and gradient approaches, (2) learning the PDF of the four features by GMM, and (3) feeding the quantised features sequence into the HMMAS models.

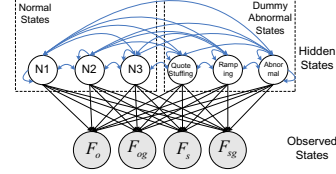


Fig. 4: Architecture of HMMAS

## V. EXPERIMENTS AND EVALUATION

### A. Experiment Data

The experiment data includes six representative stocks: Google, Microsoft and Apple from NASDAQ and ARM, BARCLAYS and Vodafone from LSE. These datasets were chosen because of their relatively higher trading volumes and more volatile price fluctuation, which are considered as important factors that might increase the likelihood of manipulation [10]. This dataset covers tick data over five trading days (from 11<sup>th</sup> – 15<sup>th</sup> June 2012) which consist of more than 2,000,000 time points in total for each stock. The trading data on 11 June 2012, which contains around 400,000 points, is used for training. The 5-fold cross-validation is used on these 400,000 data points in order to train the detection model. The trading data between 12<sup>th</sup> – 15<sup>th</sup> June 2012 is used for testing.

Due to the lack of the market abuse data, synthetically generating the abuse patterns is accepted for evaluating the model. Three typical manipulation patterns as shown in Fig.2, are synthetically created and injected into the testing datasets of different stock context making it a mixture of both “normal” and “abnormal” patterns. For each stock dataset, the model is tested on five different testcases  $n = 1, \dots, 5$ , where testcase  $n = 1$  contains 10 anomalies for each of the three types of patterns, thus 3 (types) \* 10 (anomalies per type) = 30 total numbers of injected anomalies. Testcase  $n = 2$  has  $3 * 20 = 60$  numbers and testcase  $n = 3, 4, 5$  has  $3 * 40 = 120$ ,  $3 * 80 = 240$ , and  $3 * 100 = 300$  total numbers of injected anomalies respectively. The trained models are then tested on the mixed test dataset to detect abnormal patterns. This is practical and acceptable for business people, especially when it is very costly to obtain real manipulation cases.

### B. Performance Metrics

The performance evaluation of the proposed model is based on the *precision* ( $Pre$ ), *sensitivity* ( $Sen$ ) and the *specificity* ( $Spe$ ) metrics which are defined as  $Pre = \frac{TP}{TP+FP}$ ,  $Sen = \frac{TP}{TP+FN}$  and  $Spe = \frac{TN}{FP+TN}$ , where TP is the true positive (normal cases detected as normal), TN is the true negative (manipulation cases detected as manipulation), FP is the false positive (manipulation cases detected as normal) and FN is the false negative (normal cases detected as manipulation).

Two other popular metrics for evaluating the performance of anomaly detection systems, namely the G-mean and F-measure, are used to evaluate the effectiveness of the proposed detection model. The G-mean ( $G$ ) is defined as  $G = \sqrt{Sen * Spe}$

and the F-measure (F) is defined as  $F = \frac{(1+\beta^2) * Sen * Pre}{\beta^2 * Sen + Pre}$ , where  $\beta$  is usually set to 1 [16].

TABLE I: Sensitivity values of six datasets

Dataset	Number of injected novelties				
	30	60	120	240	300
APPLE	0.7692	0.8247	0.8743	0.8660	0.8686
ARM	0.8081	0.8743	0.8815	0.8877	0.8978
BARCLAY	0.7619	0.8602	0.8743	0.8755	0.8791
GOOGLE	0.7692	0.8743	0.8815	0.8803	0.8869
MICROSOFT	0.7921	0.8649	0.8815	0.8889	0.9009
VODAFONE	0.7547	0.8511	0.8767	0.8743	0.8782

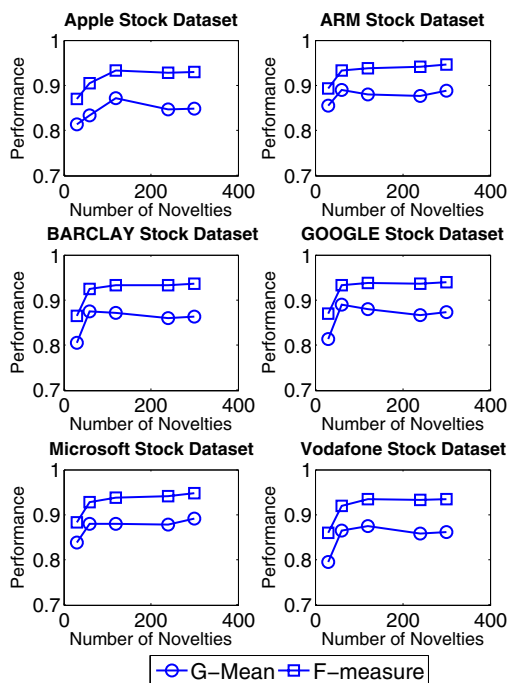


Fig. 5: G-mean and F-measure on six dataset

The experiment results are illustrated in Fig. 5. The best (i.e. the highest) F-measure and G-mean scores are obtained on the Microsoft stock price dataset with  $3 * 100 = 300$  injected anomalies (F=0.947 and G=0.89). The lowest F-measure and G-mean scores are obtained on Vodafone dataset with  $3 * 10 = 30$  injected anomalies (F=0.86 and G=0.796). Besides the extreme values, the F-measure is generally stable and within the range [0.9, 0.95]. But the G-mean is relatively volatile ranging from 0.8 to 0.89. It is notable that the *precision* values in all the experiments are as stable as one, which means no manipulation cases are detected as normal. However, the *sensitivity* values varied across different datasets as illustrated in Table I.

## VI. CONCLUSION

This paper presents a computational intelligence approach for price manipulation detection. Price changes caused by different types of market manipulation were characterised

using the gradient information and multi-level wavelet analysis of the price data. Based on the extracted features, a Hidden Markov Model with Abnormal States (HMMAS) was proposed for detecting the anomalies in the bid and ask prices. The proposed model was validated on real world financial datasets with synthetically injected anomalies. Performance evaluation results show that all synthetic price manipulation cases were successfully detected. However, the sensitivity measure varied across different datasets which indicates the varied rates of normal case detected as manipulation and further work is needed to optimise it.

## ACKNOWLEDGMENT

The authors gratefully acknowledge the support of Dr Patrick McParland and Bruce Bland from Fidessa Group plc. This project is supported by the companies and organisations involved in the Northern Ireland Capital Markets Engineering Research Initiative.

## REFERENCES

- [1] P.S. Addison. *The illustrated wavelet transform handbook: introductory theory and applications in science, engineering, medicine and finance*. Taylor & Francis, 2002.
- [2] M Aitken, F Harris, and S Ji. Trade-based manipulation and market efficiency: A cross-market comparison. In *22nd Australasian Finance and Banking Conference 2009*, 2009.
- [3] F Allen and D Gale. Stock Price Manipulation. *The Review of Financial Studies*, 5(3):503–529, 1992.
- [4] C Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.
- [5] D Cumming, S Johan, and D Li. Exchange trading rules and stock market liquidity. *Journal of Financial Economics*, 99(3):651–671, 2011.
- [6] D Diaz, B Theodoulidis, and P Sampaio. Analysis of stock market manipulations using knowledge discovery techniques applied to intraday trade prices. *Expert Systems with Applications*, 38(10):12757 – 12771, 2011.
- [7] N Hautsch and R Huang. The market impact of a limit order. *Journal of Economic Dynamics and Control*, 2011.
- [8] E Haven, X Liu, and L Shen. De-noising option prices with the wavelet method. *European Journal of Operational Research*, 222(1):104 – 112, 2012.
- [9] D Ian. Market abuse and surveillance. [www.bis.gov.uk/assets/foresight/docs/computer-trading/12-1076-eia17-market-abuse-and-surveillance](http://www.bis.gov.uk/assets/foresight/docs/computer-trading/12-1076-eia17-market-abuse-and-surveillance), 2012.
- [10] E Lee, K Eom, and K Park. Microstructure-based manipulation: Strategic behavior and performance of spoofing traders. *Journal of Financial Markets*, page in press, 2012.
- [11] J Mongkolkeha and S Tirapat. Marking the close analysis in thai bond market surveillance using association rules. *Expert Systems with Applications*, 36(4):8523 – 8527, 2009.
- [12] Nanex. Whac-a-mole is manipulation. [www.nanex.net/aqck2/3598.html](http://www.nanex.net/aqck2/3598.html), 2012.
- [13] H Ögüt, M. Mete Doğanay, and R Aktaş. Detecting stock-price manipulation in an emerging market: The case of turkey. *Expert Systems with Applications*, 36(9):11944–11949, November 2009.
- [14] L Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.
- [15] B Schölkopf, J Platt, J Shawe-Taylor, A Smola, and R Williamson. Estimating the support of a high-dimensional distribution. *Neural computation*, 13(7):1443–1471, 2001.
- [16] Y Tang, Y Zhang, N.V. Chawla, and S. Krasser. Svms modeling for highly imbalanced classification. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 39(1):281 –288, Feb. 2009.
- [17] J Tse, X Lin, and D Vincent. High frequency trading - measurement, detection and response. Technical report, Credit Suisse, 2012.
- [18] J Tse, X Lin, and D Vincent. High frequency trading - the good, the bad and the regulation. Technical report, Credit Suisse, 2012.
- [19] D. Yeung and Y. Ding. Host-based intrusion detection using dynamic and static behavioral models. *Pattern Recognition*, 36(1):229 – 243, 2003.