Journal of Universal Computer Science, vol. 14, no. 14 (2008), 2288-2308
submitted: 30/9/07, accepted: 30/4/08, appeared: 28/7/08 © J.UCS

Market Microstructure Patterns Powering Trading and Surveillance Agents

Longbing Cao
(University of Technology, Sydney, Australia
lbcao@it.uts.edu.au)

Yuming Ou
(University of Technology, Sydney, Australia
yuming@it.uts.edu.au)

Abstract: Market Surveillance plays important mechanism roles in constructing market models. From data analysis perspective, we view it valuable for smart trading in designing legal and profitable trading strategies and smart regulation in maintaining market integrity, transparency and fairness. The existing trading pattern analysis only focuses on interday data which discloses explicit and high-level market dynamics. In the mean time, the existing market surveillance systems available from large exchanges are facing crucial challenges of diversified, dynamic, distributed and cyber-based misuse, mis-disclosure and misdealing of information, announcement and orders in one market or crossing multiple markets. Therefore, there is a crucial need to develop innovative and workable methods for smart trading and surveillance. To deal with such issues, we propose the innovative concept *microstructure pattern analysis* and corresponding approaches in this paper. Microstructure pattern analysis studies trading behaviour patterns of traders in market microstructure data by utilizing market microstructure knowledge. The identified market microstructure patterns are then used for powering market trading and surveillance agents for automatically detecting/designing profitable and legal trading strategies or monitoring abnormal market dynamics and trader's behaviour. Such trading/surveillance agent-driven market trading/surveillance systems can greatly enhance the analytical, discovery and decision-support capability of market trading/surveillance than the current predefined rule/alert-based systems.

Keywords: market microstructure pattern, data mining, agents, market surveillance
Categories: I.2.6, H.1.1, M.0, M.1

1 Introduction

In many types of markets such as capital and electricity markets, *market surveillance* plays important mechanism roles in designing market models and business rules, as well as regulation roles in maintaining the market integrity, transparency and fairness [Dehdashti, 05; UN, 05; O'Hara, 01]. Many current program trading systems are based on predefined trading strategies. The existing research on trading pattern analysis mainly focus on interday data. One the other hand, the existing market surveillance systems usually rely on surveillance rules for alerting of suspect findings in the market. Such rules are predefined and based on business rules. Additional surveillance rules come from statistics and reporting results, which can capture more sophisticated abnormal trading behaviour and market movement. These rules play

important roles in filtering obvious offences against market business rules, regulation rules, and explicitly exceptional market dynamics.

However, the existing trading pattern analysis loses the in-depth information hidden in the market microstructure. With regard to market surveillance rules, the existing systems available from large exchanges are facing crucial challenges of identifying diversified, dynamic, distributed and cyber-based misuse, mis-disclosure and misdealing of information, announcement and orders in one market or crossing multiple markets. Such challenges cannot be handled by the existing systems and techniques usually used in exchanges.

The current trading pattern analysis and price movement analysis mainly focus on interday data, in particular, closing prices. The resulting analytical results are not workable for real-time market surveillance because they cannot catch and filter the microstructure behaviour every second of every day. In fact, currently, there is no analytical work reported on analyzing tick-by-tick data for scrutinizing either profitable trading strategies or abnormal trading behaviour. There is a crucial need to develop breakthrough methodologies and techniques to discover hidden knowledge on the market microstructure data under the increasing financial and trading globalization.

In this paper, to deal with the above issues, we study *market microstructure behaviour* surrounded by *market microstructure data*. We propose the innovative concepts and corresponding approaches to identifying *market microstructure patterns*. *Market microstructure behaviour* consists of investor's actions and interactions among investors in one market or crossing multiple markets, as well as their embodiment in market dynamics. *Microstructure pattern analysis* studies trader's behaviour patterns in market microstructure data by following and involving market microstructure theories.

General market microstructure patterns consist of *positive* and *negative market microstructure patterns* in *time series* and *activity sequences*. In addition, hybrid patterns and more advanced activity microstructure patterns may be identified in market microstructure time-series and sequences.

Microstructure patterns are very useful for smart trading or surveillance. If the patterns legally make sense, they are helpful for smart trading. Some other exceptional microstructure patterns may reflect abnormal trading behaviour in the market. The identified microstructure patterns can then be used for powering market trading/surveillance agents that automatically detect/monitor the market dynamics and trader's behaviour patterns. For instance, market surveillance agents can be developed for market surveillance officers and management teams to present them alerts and indicators of abnormal market movements. Such market microstructure pattern-driven market trading/surveillance systems can greatly enhance the analytical, discovery and decision-support capability of market trading/surveillance than the current predefined strategy/alert-based systems.

In fact, microstructure behaviour can be seen in many financial applications, for instance, derivative market, foreign currency exchange market, and index exchange market. The innovative methodology of microstructure pattern analysis presents new and powerful approaches to enhancing the existing market surveillance systems and trading performance.

2 Market Microstructure and Data

2.1 Market Microstructure

Market microstructure [Madhavan, 00; Harris, 03] is a branch of finance concerned with the details of how exchange occurs in markets. The major thrust of market microstructure research examines the ways in which the working processes of a market affects determinants of transaction costs, prices, quotes, volume, and trading behaviour[1]. Microstructure theory focuses on how specific trading mechanisms affect the price formation process[2]. It is devoted to theoretical, empirical, and experimental research on the economics of securities markets, including the role of information in the price discovery process, the definition, measurement, control, and determinants of liquidity and transactions costs, and their implications for the efficiency, welfare, and regulation of alternative trading mechanisms and market structures."[3] The theory of market microstructure applies to the exchange of real or financial assets in financial markets. Market microstructure deals with issues of market structure and design, price formation and discovery, transaction and timing cost, information and disclosure, and market maker and investor behaviour.

2.2 Market Microstructure Data

Transactional data recording the investor behaviour in markets obeying market microstructure theory present a unique structure. We call such data *market microstructure data*. On the one hand, market microstructure data presents syntactic components and representation nothing special from transactions normally accumulated in many applications such as e-commerce data, retail data and telecom data. On the other hand, market microstructure data does present differences. In summary, market microstructure data presents some major characteristics that are not usually seen in many other applications.

- The data indicates *rich semantics*. The semantics is led by market microstructure theories and somehow embodied through the syntactic representation. For instance, any order is associated with either ask or bid, and orders are likely partnered from ask and bid sides. An ask-order transaction distinguishes from a bid-order one.
- The data presents *time frame* and *gradient*. Market microstructure data is associated with timeframe. Further, it also presents *time gradient*. For instance, opening and closing prices are *daily* data, while a trading price, belonging to intraday data, is usually associated with a particular time point in a hundredth of one second.
- The data presents *granularity dynamics*. There are both interday and intraday data in markets. Both interday and intraday data presents different granularities. For example, stock market index consists of both intraday and interday categories, its intraday elements also include various catalogues. Some are broad-based while others are sector-based.

[1] http://en.wikipedia.org/wiki/Market_microstructure
[2] O'Hara, Maureen, Market Microstructure Theory, Blackwell, Oxford, 1995
[3] http://www.nber.org/workinggroups/groups_desc.html

- The data is *heterogeneous*. Market data consists of numerical, categorical, sequential/serial, and textual elements. If more than one market must be involved, then the data structures are likely different. Processing and mining such mixed data needs effective methods.

The above characteristics of market microstructure data distinguish it from the data usually seen in the mainstream data mining research. Such semantic information (1) brings challenges to the existing pattern mining approaches, and also (2) provides guideline or hints for itemset construction and pattern mining.

Typical market microstructure data can be described by the data model in Figure 1. It shows a three-dimensional model: time (t_i), multiple levels of ask/bid prices (P_{Asi}/P_{Bii}) and volumes (V_{Asi}/V_{Bii}) associated with a particular point of time t_i.

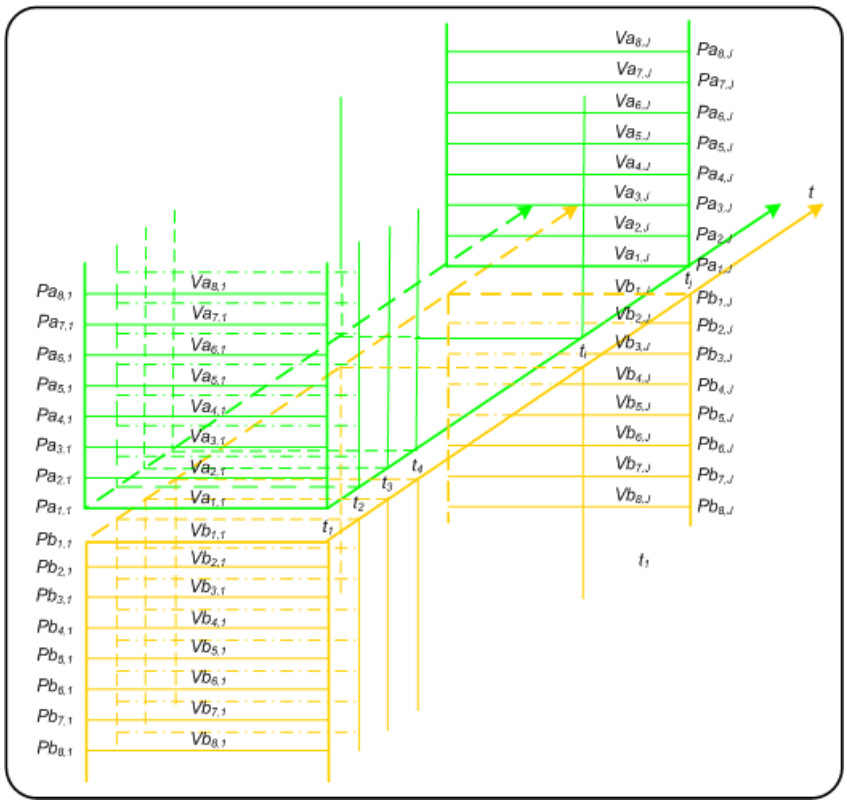

Figure 1: Market microstructure data

In financial markets, market microstructure data consists of basic data categories including *orders* (*O*), *trades* (*T*), *indices* (*I*) and market data (*M*). It involves several dimensions: time (*t*), value (*v*) and direction (*d*). Basic trading types can be classified into buy (*B*), sell (*S*) and hold (*K*). To represent a general element in market microstructure data, we use the following model:

$$a.x^{j}_{y,d} \quad (1)$$

where, $a \in \{S, I, \ldots\}$ is the asset traded: security (S), index (I), etc.; $x \in \{p, v, q\}$ refers to the target attribute: price (p), volume (v) and value (q) of a at time t, $y \in \{O, T\}$; d refers to the order direction, if y = 'T', $d \in \{B, S, K\}$ refers to buy (B), sell (S) and hold (K), while $d = \{B_i, A_s\}$ if y = 'Q' indicating the data comes from either bid (B_i) or ask (A_s) side; $j = \{1, \ldots, J\}$ refers to the number of x corresponding to time t. In addition, investor's actions $u \in \{n, l, m, w\}$ on an order can be: add (n), delete (l), amend (m) or withdraw (w). For instance, $CBA.p^{54}_{O,B_i}$ refers to the price of no. 54 bid quote on the security CBA.

2.3 Processing Market Microstructure Data

Before the mining can be undertaken on market microstructure data, the data characteristics described in the above sections have to be properly catered through data preparation. The semantics, granularity, timeframe and heterogeneity surrounding market microstructure data determine the preparation tasks and guideline for itemset construction. We discuss a few general techniques here.

The semantic information is helpful to guide us to build more meaningful series and sequences. For instance, with order types and directions, we can construct sequences by extracting investor's actions in markets. As an example, in the following, we show how to represent and construct microstructure order sequences guided by domain knowledge.

2.3.1 Vector-Based Microstructure Order Representation

Considering the fact that every order follows market microstructure theory and indicates information about order holder's intention, the representation of such order sequences should reflect them accordingly. In stock markets, although the values of a particular order attribute vary from order to order, they actually reflect a trader's intentions, and cater for the particular stage of their lifecycles. For instance, for a single time point, an order may present in one of the following states (s) in its lifecycle: $s \in \{$*new, traded partly, traded entirely, deleted, outstanding*$\}$. Further, even for the same values of a particular order attribute, they may indicate divided circumstances that reflect investor's varying motivation and behaviour [Shleifer, 00].

Therefore, the proper representation of an order should reflect order holder's intention, actions and the order's lifecycle. For this purpose, we propose *vector-based order representation*, namely a multi-dimensional vector O represents an order in terms of attributes that describe the above aspects. For instance, in stock market, a five-dimension vector $O(d, \delta, \rho, \varphi, \varepsilon)$ is defined to represent an order. Dimension d reflects the trade direction of an order, δ stands for the probability that an order is traded, ρ measures the size of an order, φ represents how many trades the order leads to, and ε reflects the balance of an order at the time of market close.

These five dimensions are defined as follows:

$$d \in \{B, S, K, B_i, A_s\} \tag{2}$$

$$\delta = \begin{cases} \delta_H, & if\ \Delta p < 0 \\ \delta_M, & if\ 0 \le \Delta p \le 1 \\ \delta_L, & if\ \Delta p > 1 \end{cases} \tag{3}$$

$$\Delta p = \begin{cases} p_2 - p_1, & if\ d \in \{B, B_i\} \\ p_1 - p_2, & if\ d \in \{S, A_s\} \end{cases} \tag{4}$$

$$\rho = \begin{cases} \rho_S, & if\ \rho < 3000 \\ \rho_M, & if\ 3000 \le \rho < 10000 \\ \rho_L, & if\ \rho \ge 10000 \end{cases} \tag{5}$$

$$\varphi = \begin{cases} \varphi_0, & if\ an\ order\ has\ never\ been\ traded \\ \varphi_1, & if\ an\ order\ has\ been\ traded\ only\ once \\ \varphi_N, & if\ an\ order\ has\ been\ traded\ over\ once \end{cases} \tag{6}$$

$$\varepsilon = \begin{cases} \varepsilon_0, & if\ an\ order\ has\ been\ completely\ traded \\ \varepsilon_1, & if\ an\ order\ is\ outstanding \\ \varepsilon_{-1}, & if\ an\ order\ has\ been\ deleted \end{cases} \tag{7}$$

where p_1 is the order price and p_2 is the last trade price in the market when the order is placed.

The order vector $O(d, \delta, \rho, \varphi, \varepsilon)$ encloses plenty of semantics: (1) indicating the direction, probability and size of an order to be traded, (2) reflecting an order's dynamics during its lifecycle. The vector actually provides a mechanism to transform microstructure-based orderbook into vector-based order sequences. For example, the following Table 1 and Table 2 show the orderbook data and trade data related to order O100.

Order number	Order date	Order time	Account ID	Security code	Trade direction	Order price	Order volume
O100	28/06/2005	09:54:07	A123	S123	B	10.00	1000

Table 1: Orderbook data related to order O100

Order number	Trade date	Trade time	Account ID	Security code	Trade direction	Trade price	Trade volume
O078	28/06/2005	09:59:52	A348	S123	S	10.10	500
O102	28/06/2005	09:59:52	A980	S123	B	10.10	500
O067	28/06/2005	09:59:56	A690	S123	S	10.00	200
O100	28/06/2005	09:59:56	A123	S123	B	10.00	200
O089	28/06/2005	10:07:49	A531	S123	S	10.00	300
O100	28/06/2005	10:07:49	A123	S123	B	10.00	300

Table 2: Trade data related to order O100

Following the $O(d, \delta, \rho, \varphi, \varepsilon)$ specification, the above order lifecycle can be transformed into the following order sequence: $(B, \delta_M, \rho_S, \varphi_N, \varepsilon_1)$.

2.4 Constructing Vector-Based Microstructure Order Sequences

Through data processing, market microstructure transactional data can be expressed in terms of *microstructure series* and *microstructure sequences*.

Microstructure series refer to both interday and intraday time-series data of numerical attributes, for instance, closing price series and trading price series. Time series consist of both original and derived series. Derived time series data is of those attributes aggregated on top of original time series, for instance, *volatility* series and *sharpe ratio* series. For example, the following shows a buy-order price series.

$$\{ p_B^1, ..., p_B^m \}$$

Microstructure sequences refer to ordinal and categorical data that is related to investor's behaviour. In markets, investor actions may be placing, amending, deleting and withdrawing orders. Corresponding sequences can be constructed for each of them. In addition, trades consist of sequences as well. For instance, we can construct all sell-side put-order sequences as per investor.

$$\{ o_{S,A_s}^1, ..., o_{S,A_s}^j \}$$

Based on the above vector-based order representation $O(d, \delta, \rho, \varphi, \varepsilon)$, microstructure orderbook can be transformed into order vectors. For intraday microstructure data, an order at most lasts for one day since its generation. This indicates the length of an order lifecycle is maximally one trading day. In addition, orders placed by different investors indicate different intentions, beliefs and desires. Thus it is domain-friendly to construct order sequences in terms of trading day and order investors.

A microstructure order sequence Ω consists of sequences of orders in vectors for a trader within a trading day,

$$\Omega = \{O_1(d^1, \delta^1, \rho^1, \varphi^1, \varepsilon^1), O_2(d^2, \delta^2, \rho^2, \varphi^2, \varepsilon^2), ..., O_i(d^j, \delta^j, \rho^j, \varphi^j, \varepsilon^j), ...\} \quad (8)$$

in which O_j is an order vector. Therefore, an order sequence Ω systematically reflects an investor's intention, his/her order lifecycles and trading activities in a market. For example, if a trader entered three orders $(B, \delta_L, \rho_L, \varphi_0, \varepsilon_{-1})$, $(B, \delta_L, \rho_L, \varphi_0, \varepsilon_{-1})$, and $(S, \delta_H, \rho_M, \varphi_1, \varepsilon_0)$ on July 16, 2004, then the corresponding order sequence can be expressed as:

$$\Omega = \{(B, \delta_L, \rho_L, \varphi_0, \varepsilon_{-1}), (B, \delta_L, \rho_L, \varphi_0, \varepsilon_{-1}), (S, \delta_H, \rho_M, \varphi_1, \varepsilon_0)\}.$$

3 Market Microstructure Patterns

Market microstructure patterns are identified in market microstructure time-series and sequences. In microstructure time-series, general microstructure patterns may consist of the following types:

- Microstructure time-series patterns

Microstructure sequential patterns identified in microstructure sequences may consist of

- Positive/negative microstructure patterns:
- Sequential microstructure patterns
- Exceptional microstructure patterns
- Microstructure activity patterns

Further, in data combining time series and sequences, combined patterns may be discovered:

- Combined microstructure patterns

We interpret them respectively in the following subsections.

3.1 Microstructure Time-series Pattern Analysis

Financial market data is typical time series data. On microstructure time-series data, the following microstructure time-series patterns may be identified.

(1) Microstructure single time-series patterns

In financial markets, basic microstructure time-series consist of *price* series, *volume* series, *value* series, *index* series. Based on these series, derived series can be generated such as *profit* series, *abnormal return* series, and *sharpe ratio* series. Pattern analysis on such series includes *outliers* and *trends*. Outliers indicate exceptional trading patterns, while *trends* present the dynamics of the normal movement.

(2) Microstructure cross-time-series patterns

With the involvement of domain knowledge, cross-time-series analysis has potential to disclose much richer information about market movement and suspect exceptional activities. The aim is to find microstructure cross-time-series patterns.

We don't expand these patterns' discovery in this paper. Rather, our focus is on microstructure activity pattern analysis, which to the best of our knowledge, has never been investigated before, and has potential to in-depth understanding of market dynamics from the perspective of market microstructure.

3.2 Microstructure Activity Pattern Analysis

By contrast to microstructure time-series pattern analysis, microstructure activity pattern analysis is a totally new topic. We here focus only on microstructure trading behaviour, namely an investor's sequences of actions in one or cross markets. Sequences of investor's trading actions can be constructed in terms of various strategies by introducing domain experts' guide, for instance,

- S_1: The lifecycle of an order,
- S_2: The lifecycle or selected investment life interval of an investor,
- S_3: The lifecycle of an investor on selected assets in one market,
- S_4: The lifecycle of an investor on selected assets cross markets.

Correspondingly, *microstructure activity pattern analysis* intends to discover behaviour patterns in the activity sequences. There are two ways to identify such patterns. One is to ignore the order in the sequences, and focus on *positive*, *negative* or *hybrid activity associations*. The other is to identify sequential activity patterns. Many innovative types of sequential activity patterns may be identified.

Investor's activity patterns may be *exceptional* if they are against average people's expectation. If this is the case, it likely indicates some abnormal behaviour in the market. Algorithms and exception monitoring indicators can be carried over by

market surveillance agents. They can then detect and alert exceptional trading in real-time and automatic manner. On the other hand, those *constructive and legal* trading behaviour patterns (including exceptionally well-performed patterns satisfying legal and regulatory rules) are useful for powering trading agents to better the investment performance. Such trading agents can autonomously detect the market movement and make smart trading decisions on behalf of their investors.

Let D be the microstructure data, $f_j \in \{f_1, \ldots, f_m\}$ be the original attributes in D, after the vector-based transformation, we get the new attributes $f_j^v \in \{d, \delta, \rho, \varphi, \varepsilon\}$. $A = \{a_1, a_2, ..., a_m\}$ are the activity set extracted from the orderbook data by following the vector-based order representation specification.

(1) Positive/negative/hybrid microstructure activity associations

We then define *positive* microstructure activity associations as $\{a_i, a_j\}$ (where ',' indicates a simple combination without ordering) indicating the occurrence of some activities in markets, while *negative* microstructure activity patterns $\{\neg a_i, \neg a_j\}$ indicate non-occurrence of the itemsets. Further, combined associations may present in form of $\{\neg a_i, a_{i+k}\}$.

(2) Sequential microstructure activity patterns

Sequential microstructure activity patterns consist of trading activities in some order, the ordering is represented by ';'. Similar to association pattern analysis, *positive* $\{a_i; a_{i+k}\}$, *negative* $\{\neg a_i; \neg a_{i+k}\}$ and *hybrid sequential* $\{\neg a_i; a_{i+k}\}$ microstructure activity patterns may be found in ordinal activity set.

(3) Exceptional microstructure activity patterns

Both microstructure activity associations and sequences can be *exceptional* if they are out of normal cycle. Exceptional activity patterns may present as outliers. Let Ψ be the exceptions, exceptional activity patterns can be generalized in terms of activity algebra: .

(4) Impact-targeted microstructure activity patterns

If the impacts of activities (denoted by the set Φ) can be defined and categorized by domain experts, then impact-targeted microstructure activity patterns [Cao, 081; Cao, 082] can be identified. For instance, the following forms of impact-targeted microstructure activity patterns may be examined.

- Positive and negative impact-targeted activity pattern mining

	Target occurred (Φ)	Target disappeared ($\overline{\Phi}$)
Pattern appearing (A)	$A \to \Phi$	$A \to \overline{\Phi}$
Pattern disappearing ($\overline{A}$)	$\overline{A} \to \Phi$	$\overline{A} \to \overline{\Phi}$

Table 3: Positive and negative impact-targeted activity pattern mining

- Sequential impact-contrasted activity patterns

Let A be a sequence of activities, sequential impact-contrasted activity patterns consist of a pair of patterns that are associated with A as follows (D_1 and D_2 refer to two datasets):

$$\begin{cases} A \to \Phi, & D_1 \\ A \to \overline{\Phi}, & D_2 \end{cases} \quad (9)$$

- Sequential impact-reversed activity patterns

Sequential impact-reversed activity patterns consist of a pair of patterns, in which one of it is an underlying impact-targeted pattern associated with an impact, the other is a derivative impact-targeted activity pattern that is associated with another impact, the impact is opposite to that of the underlying pattern. Sequential impact-reversed activity patterns can be described as follows, in which D_i and D_j could be the same or different datasets.

$$\begin{cases} A \rightarrow \Phi, & D_i \\ \{A;Q\} \rightarrow \overline{\Phi}, & D_j \end{cases} \quad (10)$$

In fact, certain outcome-targeted exceptional activity patterns can be viewed as one type of impact-driven activity patterns, in which the impacts refer to the exceptions, which may be predefined or ad hoc.

4 Approaches for Identifying Microstructure Patterns

In this section, we briefly illustrate two approaches and corresponding case studies in identifying market microstructure patterns. One is to detect announcement pre-disclosure-associated volatility deviations on time series data by segmentation, the other is to identify exceptional trading activity patterns related to stock price manipulation.

4.1 Segmenting Volatility Deviation Associated with Announcement Pre-disclosure

Finance literatures have disclosed that the announcement arrival and the resolution of its informational impact are directly related to the dynamics of the market volatility [Andersen, 97; Mitchell, 94; Rahman, 02]. On top of such domain findings, we develop methods [Ou, 07; Yu, 07] to segment microstructure time-series for finding volatility deviations. Such deviations, if identified, may indicate announcement pre-disclosure. Rules can then be defined and loaded to market surveillance agents for automated detection of potentially exceptional company announcements on market movements. Thus this section discusses the detection of movement changes (denoted by turning points) on microstructure time-series.

4.1.1 Volatility and segmentation

The market return volatility is defined as follows.

$$\tau = \frac{std(r_t)}{\sqrt{T}} \quad (11)$$

$$r_t = \ln \frac{P_t}{P_{t-1}} \quad (12)$$

$$P_t = \frac{\sum V \bullet P}{\sum V} \tag{13}$$

where τ is the return volatility for a time range of T, r_t is the logarithmic return, P_t is the Volume Weighted Average Price (VWAP) at the time [t-T, t], V is the volume, and P is the price. Using VWAP instead of trading price can deduct some noises in high frequency intraday data. A sliding window with size T is used to calculate the return volatility, which is a time series.

Segmentation [Keogh, 01] refers to the process of partitioning a time series into multiple regions that essentially share similar distribution and indicate similar trading behaviour. Segmentation is helpful for understanding the dynamics including the turning points of return volatility. In segmenting the price/volatility time series, *piecewise linear fitting* has been demonstrated to fit in high frequency data that contains much noise. We use the following *piecewise segmented model* to segment the volatility time series.

$$Y = \begin{cases} g_1(t,\alpha_1) + e_1(t), & 1 < t < \theta_1 \\ g_2(t,\alpha_2) + e_2(t), & \theta_1 < t < \theta_2 \\ \vdots & \\ g_k(t,\alpha_k) + e_k(t), & \theta_{k-1} < t < \theta_k \end{cases} \tag{14}$$

where the $f_i(t,w_i)$ is the linear function that fits in segment i, and $e_i(t)$ is its error. This model splits the time series into segments that have minimal total error:

$$\mu = \sum_{i=1}^{k} e_i(t) \tag{15}$$

4.1.2 Experimental results

Experiments have been carried out on five stock intraday datasets of AMP Limited Australia in Australian Stock Exchange to compare the detection performance of volatility-based turning points with that of trading prices-based. Table 4 shows a sample announcement dataset which include *trading date*, *number of announcements* released during normal trading hours (from 10am to 4pm) on that day and *release time of announcements*.

No.	Trading date	Number of announcements	Release time	Release time
1	20/11/2001	2	11:12:32	13:25:36
2	27/02/2003	2	12:31:44	13:53:55
3	13/03/2003	2	12:33:11	14:41:55
4	31/03/2003	1	11:16:23	
5	07/05/2003	1	12:01:45	

Table 4: Sample announcement datasets

Both price and volatility data is segmented based on the piecewise segmented model (formulas 11-15) to spot the turning points and see their performance difference. The interval T for calculating VWAP is one minute. The size of the sliding window to calculate volatility is 30 minutes. The *turning points* are defined as the points that separate two adjacent segments with divided trends and have the shortest distance from the release time of announcements.

	Dataset 1		
Release time	11:12:32	13:25:36	Number of segments
Real index	73	206	
Turning point of price	71	235	14
Turning point of volatility	79	225	7
	Dataset 2		
Release time	12:31:44	13:53:55	Number of segments
Real index	152	234	
Turning point of price	147	241	18
Turning point of volatility	159	243	15
	Dataset 3		
Release time	12:33:11	14:41:55	Number of segments
Real index	154	282	
Turning point of price	155	279	13
Turning point of volatility	159	277	4
	Dataset 4		
Release time	11:16:23		Number of segments
Real index	77		
Turning point of price	71		6
Turning point of volatility	77		3
	Dataset 5		
Release time	12:01:45		Number of segments
Real index	122		
Turning point of price	119		10
Turning point of volatility	125		4

Table 5: Some experimental results of detecting the turning points

Some of the segmentation results are shown in Table 5. The “Real index” rows indicate the points associated with the release time of announcements. It shows that

with the similar segmentation effect, the number of segments needed for detecting the turning points of return volatility is less than that for trading price. This means that return volatility can capture the announcement arrival better than trading price. The turning points of return volatility are distinguished from the noise around them, and the trends before and after them are quite different. Therefore, they are easy to be spotted by less segments. Whereas the turning points of trading price tend to be overwhelmed by the noise surrounding them so that they need more segments.

4.2 Detecting Exceptional Microstructure Behaviour Associated with Price Manipulation

In the previous sections, we present an approach to construct the vector-based order sequences based on domain knowledge. In this section, we introduce an approach to detect exceptional microstructure trading activities associated with the price manipulation on the order sequences.

4.2.1 Target Data and Benchmark Data

The data for extracting and constructing order sequences is intraday orderbook data. We reconstruct the microstructure orderbook into microstructure order sequences by following our proposed vector-based sequence construction.

The *target data* and *benchmark data* are selected in terms of a sliding time window (assuming size *m*). As shown in the Figure 2, all data on the current target day comes into the *target data set*, the data on the *m* previous days before the target day are called *benchmark day 1, 2, …, m* respectively. All data drawn from the *benchmark day j* fits into *benchmark data set j*. Through sliding the time window in the available data set, we can generate target and benchmark data sets for a target trading date. Intuitively we can understand that there is some sort of correlation between a target data set and its benchmark data. The closer the *benchmark day* is to the target trading day, the bigger the degree of correlation is. Based on such intuition, each *benchmark day* is assigned a weight (ω_j) by the following formula:

$$\omega_j = (1 + \gamma)^{j-1} \tag{16}$$

where $\gamma \geq 0$ is the volatility of the market. The more volatile the market is, the bigger γ is.

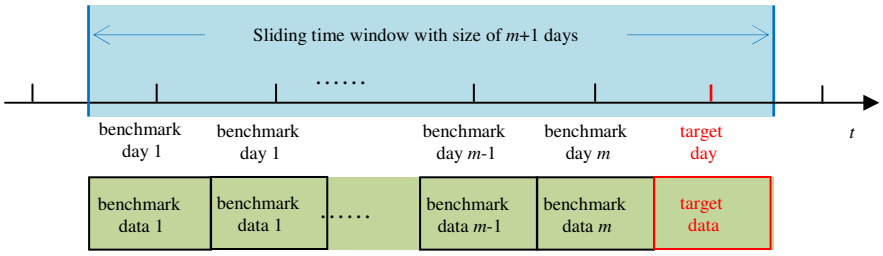

Figure 2: Constructing target and benchmark data sets

4.2.2 Algorithms for Identifying Exceptional Activity Patterns

To identify exceptional microstructure patterns in target data set, we define two interestingness measures: *Intentional Interestingness* (I_i) and *Exceptional Interestingness* (I_e).

Definition 1. *Intentional Interestingness* (I_i): I_i quantifies the intentional interestingness of a pattern as defined in the following formula:

$$I_i = Sup_t \times \frac{|\Omega|}{AvgL_t} \quad (17)$$

where

- Sup_t is the support of sequence Ω in the *target data set* of day t,
- $|\Omega|$ is the length of sequence Ω,
- $AvgL_t$ is the weighted average length of sequences in the *target data*,

I_i is positively related to the support of the *target data* and the length of its pattern. This metric reflects that investors tend to use a series of orders to deploy their intentions. Therefore an order sequence with a higher support and a longer length may happen on the *target day* at a higher probability.

Definition 2: *Exceptional Interestingness* (I_e): I_e quantifies the exceptional interestingness of a pattern as defined in the following formula:

$$I_e = \frac{\frac{Sup_t}{AvgL_t} \times \sum_{j=1}^{m} \omega_j}{\sum_j^m \left(\frac{SupB_j}{AvgLB_j} \times \omega_j \right)} \quad (18)$$

where

- $SupB_j$ is the support of sequence Ω in the *benchmark data j*,
- $AvgLB_j$ is the weighted average length of sequences in the *benchmark data j*,
- ω_j is the weight for the *benchmark data j*, and
- m is the number of *benchmark days*.

I_e is negatively related to the supports in the *benchmark data set*. The lower support the pattern has, the more likely the pattern is exceptional. I_e reflects how exceptional a pattern presents on *target day* than on *benchmark days*.

Consequently, a sequence is an *exceptional microstructure pattern*, if it satisfies the conditions:

- $I_i \geq I_{i0}$, and
- $I_e \geq I_{e0}$,

where I_{i0} and I_{e0} are the thresholds given by users or domain experts for the *intentional interestingness* and *exceptional interestingness* respectively.

The algorithm for mining *exceptional microstructure patterns* is described in Figure 3.

ALGORITHM: *Mining Exceptional Activity Patterns*
INPUT: trading dataset *TD*, order dataset *OD*, *m*, γ, I_{i0}, I_{e0}
OUTPUT: exceptional activity patterns *EAP*

EAP = Ø; */* exceptional patterns*/*
BS = Ø; */*benchmark sequences*/*
/ initialise the benchmark sequences */*
FOR each trading day *j* from trading day 1 to day *m*
    S = GenSeq(TD_j, OD_j); */*generate sequences*/*
    BS = *BS* + *S*; */*add the sequences to benchmark sequences*/*
ENDFOR
/ mining exceptional activity patterns*/*
FOR each trading day *j* from day *m* + *1* to the last trading day
    S = GenSeq(TD_j, OD_j); */*generate sequences from targeted data*/*
    P = MinePatterns(*S*) */*mine patterns from the sequences*/*
    FOR each pattern P_j in *P*
        I_{ij} = GetII(P_j); */* quantify the intentional interestingness*/*
        I_{ej} = GetEI(P_j, *BS*, γ); */*quantify the exceptional interestingness*/*
        */*add the pattern into EAP, if it meets the conditions*/*
        IF $I_i \geq I_{i0}$ and $I_e \geq I_{e0}$
            $EAP = EAP + P_j$;
        ENDIF
    ENDFOR
    Replace the sequences generated from trading day *j* – *m* in *BS* with *S*;
ENDFOR
OUTPUT exceptional activity patterns *EAP*;

Figure 3: Algorithm for Mining Exceptional Microstructure Patterns

4.2.3 Experimental results

The above approach has been tested on a real stock data set. The data consists of 240 trading days from 2005 to 2006 for a security. There were 213,898 orders entered by traders during this period. These orders led to 228,186 trades.

Table 4 shows samples of the discovered exceptional microstructure patterns. These patterns reflect the traders' exceptional intentions on the corresponding day. For example, on May 24, 2005, the I_i and I_e for the pattern $\{(S, \delta_m, \rho_s, \varphi_1, \varepsilon_0), (S, \delta_m, \rho_s, \varphi_1, \varepsilon_0)\}$ are *0.054* and *11.2* respectively. This indicates a strong intention and exception of trading activities conducted on that day.

Date	Exceptional Microstructure Patterns	I_i	I_e	Return %	AR %
05/01/2005	$\{(B, \delta_H, \rho_S, \varphi_0, \varepsilon_{-1}),(B, \delta_H, \rho_S, \varphi_0, \varepsilon_{-1}),(B, \delta_H, \rho_S, \varphi_0, \varepsilon_{-1}),(B, \delta_H, \rho_S, \varphi_0, \varepsilon_1)\}$	0.026	+∞	1.68	0.77
14/01/2005	$\{(B, \delta_H, \rho_S, \varphi_0, \varepsilon_{-1}),(B, \delta_H, \rho_S, \varphi_0, \varepsilon_{-1}),(B, \delta_H, \rho_S, \varphi_1, \varepsilon_0)\}$	0.025	7.6	-2.68	-1.38
18/01/2005	$\{(S, \delta_M, \rho_S, \varphi_0, \varepsilon_{-1}),(S, \delta_M, \rho_S, \varphi_0, \varepsilon_{-1}),(S, \delta_M, \rho_S, \varphi_0, \varepsilon_{-1}),(S, \delta_M, \rho_S, \varphi_1, \varepsilon_0)\}$	0.026	10.8	2.25	1.85
20/01/2005	$\{(S, \delta_M, \rho_S, \varphi_0, \varepsilon_{-1}),(S, \delta_H, \rho_S, \varphi_1, \varepsilon_0)\}$	0.038	5.4	-2.98	-1.56
28/01/2005	$\{(B, \delta_H, \rho_S, \varphi_1, \varepsilon_0),(B, \delta_M, \rho_S, \varphi_1, \varepsilon_0)\}$	0.030	8.2	-2.63	-1.97
01/02/2005	$\{(B, \delta_H, \rho_S, \varphi_0, \varepsilon_{-1}),(B, \delta_H, \rho_S, \varphi_0, \varepsilon_{-1}),(B, \delta_H, \rho_S, \varphi_0, \varepsilon_{-1})\}$	0.030	5.2	0.79	0.93
13/05/2005	$\{(S, \delta_M, \rho_S, \varphi_0, \varepsilon_{-1}),(S, \delta_M, \rho_S, \varphi_0, \varepsilon_{-1}),(S, \delta_M, \rho_S, \varphi_0, \varepsilon_{-1})\}$	0.026	5.1	-0.95	-2.24
24/05/2005	$\{(S, \delta_M, \rho_S, \varphi_1, \varepsilon_0),(S, \delta_M, \rho_S, \varphi_1, \varepsilon_0)\}$	0.054	11.2	6.82	6.38
16/06/2005	$\{(B, \delta_M, \rho_S, \varphi_1, \varepsilon_0),(S, \delta_M, \rho_S, \varphi_0, \varepsilon_{-1})\}$	0.025	6.1	-2.64	-2.47
17/06/2005	$\{(S, \delta_H , \rho_S, \varphi_0, \varepsilon_{-1})\}$	0.033	19.0	-1.88	1.00
01/07/2005	$\{(B, \delta_H, \rho_S, \varphi_0, \varepsilon_{-1}),(B, \delta_H, \rho_S, \varphi_0, \varepsilon_{-1}),(B, \delta_H, \rho_S, \varphi_1, \varepsilon_0)\}$	0.028	54.0	-2.21	-1.28
13/07/2005	$\{(B, \delta_M, \rho_S, \varphi_0, \varepsilon_{-1})\}$	0.028	5.6	9.55	9.12
15/09/2005	$\{(B, \delta_H , \rho_S, \varphi_0, \varepsilon_{-1}),(B, \delta_H , \rho_S, \varphi_0, \varepsilon_{-1})\}$	0.035	8.8	1.43	3.49
05/12/2005	$\{(S, \delta_M, \rho_S, \varphi_1, \varepsilon_0),(S, \delta_M, \rho_S, \varphi_1, \varepsilon_0)\}$	0.035	8.6	1.85	3.41

Table 6: Exceptional microstructure pattern samples (m = 10, γ = 0.01, I_{i0} = 0.025, and I_{e0} = 5); AR stands for the security's abnormal return.

Table 7 further shows the trading days with local exceptional patterns for different values of m and γ when I_{i0} = 0.025 and I_{e0} = 5. The symbol √ indicates that exceptional microstructure patterns are identified on that day.

Date	Return %	AR %	Exceptional Microstructure Pattern	m=30 γ=0.05	m=20 γ=0.05	m=15 γ=0.05	m=10 γ=0.05	m=10 γ=0.03	m=10 γ=0.02
05/01/2005	1.68	0.77	$\{(B, \delta_H, \rho_S, \varphi_0, \varepsilon_{-1}),(B, \delta_H, \rho_S, \varphi_0, \varepsilon_{-1}),(B, \delta_H, \rho_S, \varphi_0, \varepsilon_{-1}),(B, \delta_H, \rho_S, \varphi_0, \varepsilon_1)\}$		√	√	√	√	√
14/01/2005	-2.68	-1.38	$\{(B, \delta_H, \rho_S, \varphi_0, \varepsilon_{-1}),(B, \delta_H, \rho_S, \varphi_0, \varepsilon_{-1}),(B, \delta_H, \rho_S, \varphi_1, \varepsilon_0)\}$	√	√	√	√	√	√
18/01/2005	2.25	1.85	$\{(S, \delta_M, \rho_S, \varphi_0, \varepsilon_{-1}),(S, \delta_M, \rho_S, \varphi_0, \varepsilon_{-1}),(S, \delta_M, \rho_S, \varphi_0, \varepsilon_{-1}),(S, \delta_M, \rho_S, \varphi_1, \varepsilon_0)\}$	√	√	√	√	√	√
20/01/2005	-2.98	-1.56	$\{(S, \delta_M, \rho_S, \varphi_0, \varepsilon_{-1}),(S, \delta_H, \rho_S, \varphi_1, \varepsilon_0)\}$		√	√	√	√	√
28/01/2005	-2.63	-1.9	$\{(B, \delta_H, \rho_S, \varphi_1, \varepsilon_0),(B, \delta_M, \rho_S, \varphi_1, \varepsilon_0)\}$	√	√	√	√	√	√

	3	7		√	√	√	√	√	√
01/02/2005	0.79	0.93	{(B, δ_H, ρ_S, φ_0, ε_{-1}),(B, δ_H, ρ_S, φ_0, ε_{-1}), (B, δ_H, ρ_S, φ_0, ε_{-1})}				√	√	√
15/03/2005	-9.09	-6.13	{(B, δ_M, ρ_S, φ_1, ε_0),(B, δ_H, ρ_S, φ_1, ε_0)}	√					
13/05/2005	-0.95	-2.24	{(S, δ_M, ρ_S, φ_0, ε_{-1}),(S, δ_M, ρ_S, φ_0, ε_{-1}), (S, δ_M, ρ_S, φ_0, ε_{-1})}		√	√		√	√
24/05/2005	6.82	6.38	{(S, δ_M, ρ_S, φ_1, ε_0),(S, δ_M, ρ_S, φ_1, ε_0)}	√	√	√	√	√	√
16/06/2005	-2.64	-2.47	{(B, δ_M, ρ_S, φ_1, ε_0),(S, δ_M, ρ_S, φ_0, ε_{-1})}	√	√	√	√	√	√
17/06/2005	-1.88	1.00	{(S, δ_H , ρ_S, φ_0, ε_{-1})}	√	√	√	√	√	√
01/07/2005	-2.21	-1.28	{(B, δ_H, ρ_S, φ_0, ε_{-1}),(B, δ_H, ρ_S, φ_0, ε_{-1}), (B, δ_H, ρ_S, φ_1, ε_0)}	√	√	√	√	√	√
13/07/2005	1.78	3.59	{(B, δ_M, ρ_S, φ_0, ε_{-1})}	√	√	√	√	√	√
15/09/2005	1.43	3.49	{(S, δ_M, ρ_S, φ_1, ε_0),(S, δ_M, ρ_S, φ_1, ε_0)}	√	√	√	√	√	√
05/12/2005	1.85	3.41	{(S, δ_M, ρ_S, φ_1, ε_0),(S, δ_M, ρ_S, φ_1, ε_0)}	√	√	√	√	√	√

Table 7: Exceptional microstructure patterns identified by different values of m and γ (I_{i0} = 0.025 and I_{e0} = 5). AR stands for the security's abnormal return.

Figures 4 and 5 further illustrate the performance of our approach under different thresholds of I_{i0} and I_{e0}.

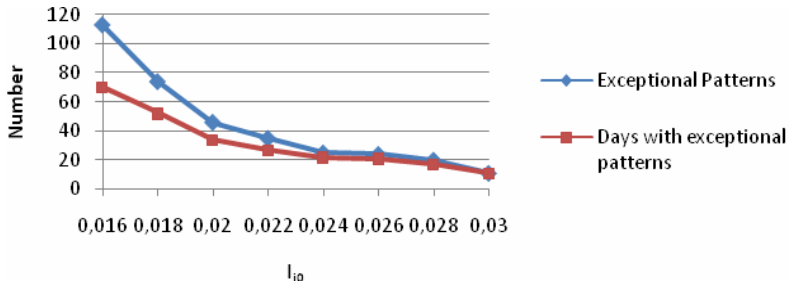

Figure 4: The number of exceptional patterns and the days with patterns for different I_{i0} when $m = 10$, $\gamma = 0.01$ *and* $I_{e0} = 5$.

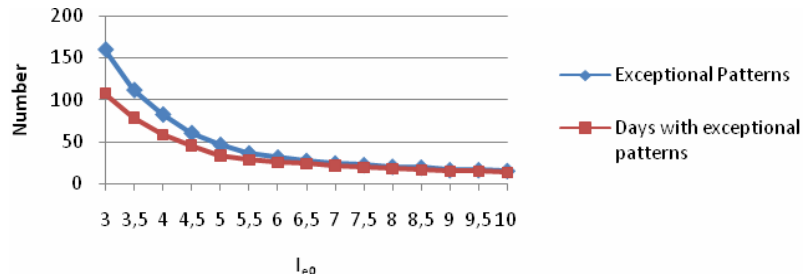

Figure 5: The number of exceptional patterns and the days with patterns for different I_{e0} when $m = 10$, $r = 0.01$ *and* $I_{i0} = 0.02$.

To evaluate the actionability of our findings in the business world, we calculate the *return* and *abnormal return* of the security. In stock markets, *return* reflects the gain or loss of a single security over a specific period while *abnormal return* indicates the difference between the *return* of a security and the *market return*. As shown in Table 1, the *return* and *abnormal return* on *24/05/2005* are as high as *6.82%*, *6.38%* respectively, which are aligned with the values of I_i and I_e for the pattern $\{(S, \delta_m, \rho_s, \varphi_1, \varepsilon_0), (S, \delta_m, \rho_s, \varphi_1, \varepsilon_0)\}$. These results from both technical and business sides present business people strong indicators showing that there likely was price manipulation on that day.

5 Microstructure Pattern-Powered Trading and Surveillance Agents

With the identified market microstructure patterns, market trading and/or surveillance agents can be armed with the rules for identifying market microstructure patterns. Such market trading agents can be (1) either integrated into automated trading systems for directly detecting signals for trading recommendation, (2) or trusted as

trading program in market trading. Similarly, market surveillance agents with such rules can monitor the market dynamics in real time, and generate alerts once abnormal things happen. Collaborated with other agents, more work such as pattern comparison can be conducted by surveillance agents.

Figure 6 illustrates the architecture of market microstructure pattern-powered trading and surveillance agents. Such agents fulfil roles of detecting microstructure patterns and generate alerts if the patterns are triggered. The figure also illustrates the scenarios and working process of microstructure pattern-powered trading and surveillance. It shows the protocols the agents following and constraints obeyed. From the programming perspective, it illustrates input/output parameters, method selection and execution control. The trading/surveillance agents interact with Data Preparation Agents to get data represented and series/sequence constructed. Coordinator Agents maintain the execution control, coordination and matchmaking of multiagents in the trading and surveillance operation. If a pattern is triggered, a trading signal or a surveillance alert is then automatically generated. The signal/alert is responded by MessageProcess Agents.

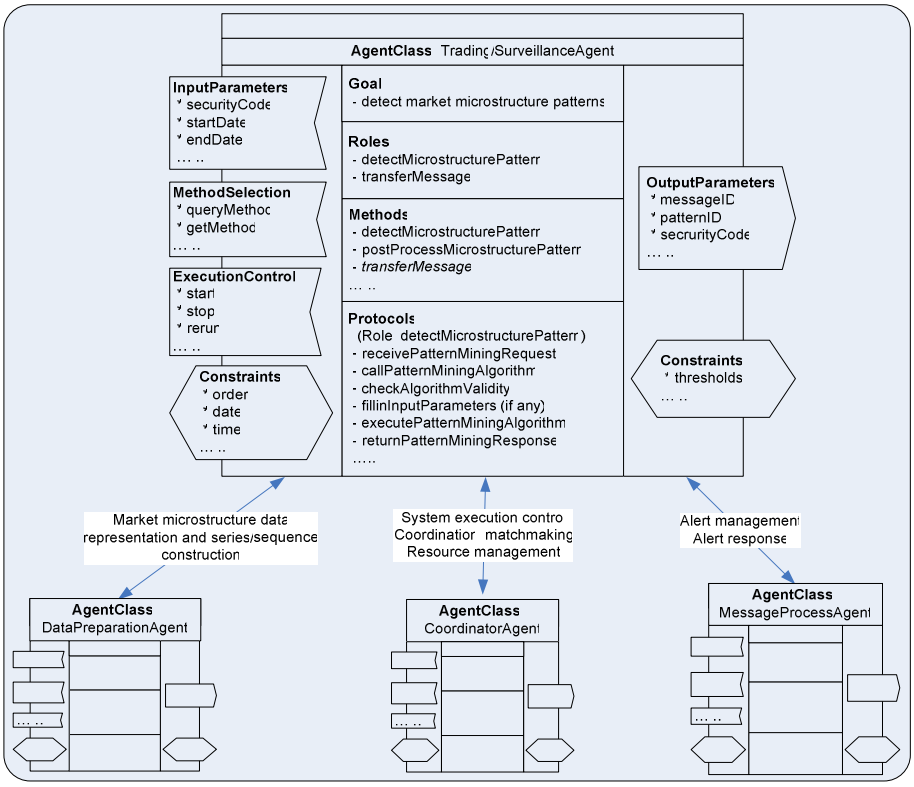

Figure 6: Architecture of market microstructure pattern-powered trading and surveillance agents

6 Conclusions

With the financial and economic globalization, market trading and surveillance are facing unprecedented chances and challenges. Technical means in supporting smart market trading and surveillance need to be greatly enhanced to adapt to such new circumstances and requirements. In this paper, we have proposed the breakthrough concept of *market microstructure patterns*. Market microstructure patterns are identified in market microstructure time-series and sequences, which are constructed in market microstructure data. We have discussed some general patterns and illustrated two specific approaches in detecting announcement pre-disclosure-associated price/volatility deviations through segmentation, and identifying market microstructure activity patterns associated with price manipulation.

Substantial experiments on real-life orderbook data have shown that market microstructure pattern analysis opens a new and effective means for crucially understanding and analyzing market dynamics. The resulting findings can greatly enhance the learning, detection, adaption and decision-making capability of trading and surveillance agents by considering semantics, granularity, timeframe, and heterogeneity.

Under the umbrella of market microstructure pattern analysis, many research issues wait for investigation, for instance, representation of microstructure data, pattern analysis crossing multiple time series/sequences.

Acknowledgements

The project is partially supported by Australian Research Council Discovery grants DP0773412, LP0775041 and DP0667060.

References

[Andersen, 97] Andersen, T., Bollerslev, T.: Heterogeneous information arrivals and return volatility dynamics: uncovering the long-run in high frequency returns, The Journal of Finance, vol. 52, pp. 975-1005, 1997.

[Cao, 081] Cao, L., Zhao, Y., Zhang, C.: Mining Impact-Targeted Activity Patterns in Imbalanced Data, IEEE Trans. on Knowledge and Data Engineering, 2008 (to appear)

[Cao, 082] Cao, L., Zhao, Y., Zhang, C., Zhang, H.: Activity Mining: from Activities to Actions, International Journal of Information Technology & Decision Making, 7(2), pp. 259 - 273, 2008.

[Dehdashti, 05] Dehdashti, E.: Monitoring and surveillance of wholesale electricity markets: roles, responsibilities and challenges, Power Engineering Society General Meeting, IEEE, vol.3, pp.3035-3041, 2005.

[Harris, 03] Harris, L.: Trading and Exchanges: Market Microstructure for Practitioners, Oxford University Press, 2003.

[Keogh, 01] Keogh, E., Chu, S., Hart, D., Pazzani, M.: An online algorithm for segmentation time series, Proceedings of IEEE International Conference on Data Mining, pp. 289-296, 2001.

[Madhavan, 00] Madhavan, A.: Market microstructure: A survey, Journal of Financial Markets, vol. 3, pp. 205- 258, 2000.

[Mitchell, 94] Mitchell, M.L. & Mulherin, J.H: The impact of Public Information on the stock market, The Journal of Finance, vol.49 no.3 pp.923-925, 1994

[O'Hara, 01] O'Hara, M.: Designing markets for developing countries, International Review of Finance, vol. 2, pp.205-215, 2001

[Ou, 07] Ou, Y., Cao, L., Yu, T., Zhang, C.: Detecting Turning Points of Trading Price and Return Volatility for Market Surveillance Agents, Proceedings of the 2007 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology - Workshops, IEEE Computer Society, pp. 491-494, 2007.

[Rahman, 02] Rahman, S., Lee, C., Ang, K.: Intraday return volatility process: evidence from NASDAQ stocks, Review of Quantitative Finance and Accounting, vol. 19, pp. 155-180, 2002.

[Yu, 07] Yu, T., Cao, L., Ou, Y., Zhang, C.: Segmentation algorithms in high frequency time-series data, The 6th International Conference on Computational Intelligence in Economics and Finance (CIEF 2007), pp. 445-451, 2007.

[Shleifer, 00] Shleifer, A.: Inefficient Markets: An Introduction to Behavioral Finance, Oxford University Press, USA, 2000.

[UN, 05] United Nations.: Market Surveillance In The Unece Region, 2005.