Procedia Computer Science
Volume 80, 2016, Pages 752–761
ICCS 2016. The International Conference on Computational Science

Detecting Informative Patterns in Financial Market Trends based on Visual Analysis.

Javier Sandoval[2], Jaime Nino[3], German Hernandez[4], and Andrea Cruz[1]

[1] Universidad Nacional, Bogota, Colombia
amcruzm@unal.edu.co
[2] Universidad Nacional,Universidad Externado Bogota, Colombia
jhsandovala@unal.edu.co
[3] Universidad Nacional, Bogota, Colombia
jhninop@unal.edu.co
[4] Universidad Nacional Bogota, Colombia
gjhernandezp@unal.edu.co

Abstract

This paper presents a graphical representation that fully depicts the price-time-volume dynamics in a Limit Order Book (LOB). Based on this pattern representation, a clustering technique is applied to predict market trends. The clustering technique is tested on information from the USD/COP market. Competitive trend prediction results were found, and a benchmark for future extensions was settled.

Keywords: Market Trend, Limit Order Book, Pattern Extraction.

1 Introduction

Every exchange market is organized as a dealership or a limit order market. Current markets share features from both categories. However, it is possible to identify markets with a dominant side. In a dealership market, a customer's order is filled at a single price quoted by a specialist, an institutional investor that is in charge of providing liquidity. The price that the specialist provides for a certain quantity of the corresponding asset does not affect the price that it would quote for different quantities. In dealership markets, dealers are expected to quote prices that clear the current buy and sell pressure. Examples of markets organized as dealership markets are NASDAQ and LSE.

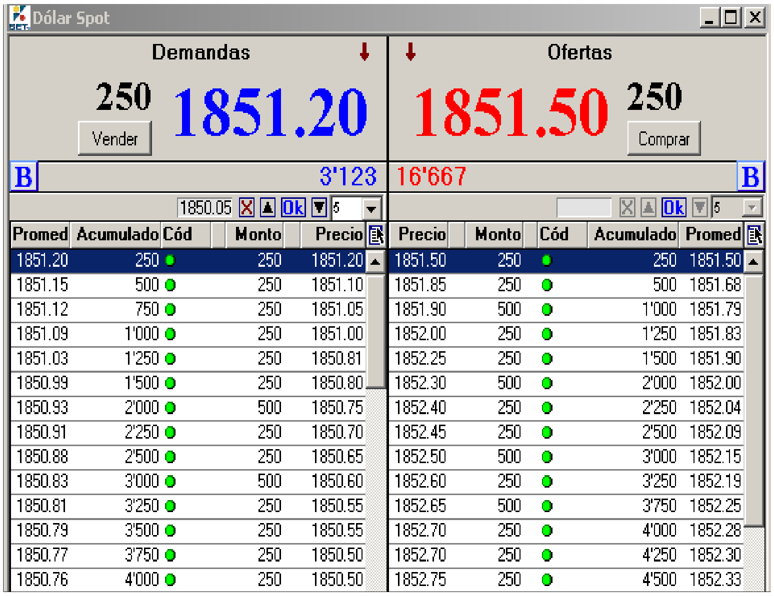
In contrast, several exchange markets, including the EURONEXT and the Colombian stock and FX markets, are organized into what is called, order-driven markets, markets that do not have a specialized dealer. In limit order markets, participants carry on their trades submitting

Selection and peer-review under responsibility of the Scientific Programme Committee of ICCS 2016

doi:10.1016/j.procs.2016.05.365

either limit orders or market order. This differentiation is important because it is directly related to the cost assumed by the agent when executing a buy/sell transaction.

Informally, a limit order is defined as an order that has to be fulfilled at a specific volume and price. If market conditions are not adequate for executing a limit order, the latter will be queued in what is called the limit order book until execution or cancellation, whichever occurs first. Orders in the limit order book are organized by price, with the best prices on top. It means the buy order with the highest price on the buy side of the book and the sell order with the lowest price on the sell side of the book. If multiple orders have the same price, they are organized by time of arrival following a FIFO (First In First Out) mechanism.


Dólar Spot

Demandas: 250 1851.20 Vender — 3'123

Ofertas: 1851.50 250 Comprar — 16'667

Promed	Acumulado	Cód	Monto	Precio	Precio	Monto	Cód	Acumulado	Promed
1851.20	250		250	1851.20	1851.50	250		250	1851.50
1851.15	500		250	1851.10	1851.85	250		500	1851.68
1851.12	750		250	1851.05	1851.90	500		1'000	1851.79
1851.09	1'000		250	1851.00	1852.00	250		1'250	1851.83
1851.03	1'250		250	1850.81	1852.25	250		1'500	1851.90
1850.99	1'500		250	1850.80	1852.30	500		2'000	1852.00
1850.93	2'000		500	1850.75	1852.40	250		2'250	1852.04
1850.91	2'250		250	1850.70	1852.45	250		2'500	1852.09
1850.88	2'500		250	1850.65	1852.50	500		3'000	1852.15
1850.83	3'000		500	1850.60	1852.60	250		3'250	1852.19
1850.81	3'250		250	1850.55	1852.65	500		3'750	1852.25
1850.79	3'500		250	1850.55	1852.70	250		4'000	1852.28
1850.77	3'750		250	1850.50	1852.70	250		4'250	1852.30
1850.76	4'000		250	1850.50	1852.75	250		4'500	1852.33

Figure 1: Snapshot of the limit order book on February 2nd 2011 at 10:20:09 A.M extracted from the SETFX, the interbank exchange platform. The best fourteen prices and volumes are presented. first column shows average prices, second column shows cumulative volume, third column shows order volume and forth column shows order prices. Demandas, ofertas, monto, precio stand for bids, offers, quantity and price respectively.

Figure 1 shows a snapshot of the limit order book for the Colombian exchange market on February 2nd, 2011 at 10:20:09 A.M extracted from the SETFX, the interbank exchange platform. Suppose a market agent wants to sell 750 thousand dollars at 1852.32. She has to use a limit order. The later will be placed on the sell (right) side of the limit order book, after the order to sell 250 thousand dollars at 1852.3 and before the order to sell 250 thousand dollars at 1852.33. Before the new limit order could be fulfilled, orders above of it have to be executed or canceled.

In contrast to limit orders, market orders define a specific volume and no price at which they have to be executed. Therefore, market orders take their prices from the limit orders needed to be fulfilled. Now, assume that an agent delivers a market order to sell 750 thousand dollars. Because the agent's order is a market order, it will be matched with best limit orders on the corresponding limit order book side. If the market order is a buy (sell) order, it will be executed using the limit order book sell (buy) side. Therefore, the previous market order will be fulfilled using the first two limit orders of the book buy side. The selected limit and market orders leave the market as executed orders, and the limit order book changed.

Depending on which type of orders participants use, transaction costs will be different. Observing figure 1, if an agent wants to buy 250 thousand dollars, he has to choose between a market or a limit order. If he decided to use a market order, the execution price would be 1851.5. If he decided to use a limit order, the best he could do is to place an order on the top of the buy side book, for example, selecting 1851.21 as the limit price. Assuming that the limit order is executed (remember that the latter is not guaranteed), it will be a difference of 29 cents between both transactions.

The LOB evolves discretely in time when market buyers and sellers insert, modify or delete limit orders. However, LOB also changes when market participants demand liquidity using market orders [3]. Formally, an LOB is considered a variable length list consisted of limit orders where transactions occur at non-uniform time intervals[4].

Recently, there has been a strong interest in modeling and analyzing LOB dynamics to extract richer information than closing prices to predict market trends [5, 7–12]. Closing prices are publicly available today to any investors, while LOB information is usually only available to brokers or professional investors willing to pay for this extra information. However, there is not overwhelming evidence to support the extra predictability ability of the LOB, or the problem is too technique or market dependent. For example,[9] and [5] state that the whole LOB helps to determine price direction, whereas [1] presents results where the further the movement from the best quotes on the LOB, the less information they convey. On the other hand, [8] uses changes on the LOB to forecast directly transaction price dynamics without obtaining statistically significant results. As evidence is mixed, the problem of predicting financial prices using LOB information seems to be too technique or market dependent.

Although current electronic financial markets generate a massive amount of information, the standard way of approaching LOB modeling has been parametric models that describe the LOB dynamics in a compact way or flexible agent models based on simulation. See for example [6], where a Kalman filter is used to estimate current LOB's states from a linear dynamic system and [2], where simulation is used to determine agents behavior when exposed to LOB information to determine trading decisions.

Although parametric and simulation-based models are helpful to understand the dynamics of LOBs, new efficient methods and hash functions to handle and operate LOB data are necessary [13], specially if researchers and practitioners want to apply non-parametric techniques to process LOB data. Under this circumstance, this paper starts providing a methodological approach to managing LOB data in order to extract useful information to construct profitable trading strategies.

This work is organized as follows: Section 2 presents the experimental setup. Section 3 shows results and discussion. Finally, Section 4 presents conclusions and extensions for future work.

2 Experimental Setup

This section presents a method for handling large amounts of data in financial markets. Experiments were conducted using real trading data from the USD/COP market between March 1 and August 31, 2012. Complete LOB of USD/COP is rebuilt from every market event observed.

Available information includes limit order timestamps, prices, and volumes; LOB data is summarized every minute and every 20 COP cents in the time/price domain respectively. LOB volume was quantized in multiples of USD 250,000, the minimum trading volume for the USD/COP market. The maximum volume observed for a particular order during the period analyzed was 43.5 million US dollars. The maximum price found was 1,862.6 COP, and the minimum was 1,742.2 COP.

First, this work depicts an LOB Graphical representation to facilitate human traders localize visual patterns. The suggested visualization was built as follows:

- Book Event Aggregation: Due to USD/COP Market Depth, we chose an aggregating interval of one minute. We suggest that more liquid markets use aggregation intervals of 30 seconds or even 1 second.
- Price quantization: LOB order prices were mapped to a discretized price grid with a granularity of 20 COP cents. This quantization helps to deal with the LOB sparsity.
- Volume quantization: Inside the discretized price ranges for every time interval, We averaged tick-by-tick volumes; the higher the averaged volume in certain price level, the lowest the level of lightness used to represent it.

A similar representation named BookMap X-ray[1] is provided by VeloxPro. This representation is also volume inspired. However, it is built in grayscale and it does not take advantage of chromatic differences or relationships presented in data.

Definition 1. *Pattern: A matrix representing aggregated volume in the LOB for a specific window in the price/time domain.*

Definition 2. *Trend: A trend in financial markets represents the sign of the price movements in a particular time interval. One last price, higher(lower) than an initial price shows a positive(negative) trend. According to the distance between prices, a trend is classified as strong or weak.*

This paper holds the assumption that frequent volume structures in the price/time domain are informative. If repeating volume patterns were found, it would be reasonable to count every appearance of each pattern and store the number of times that the pattern is associated with a specific trend (in this case bullish or bearish), to calculate the probability of the pattern being related to a given trend. This process allows labeling frequent patterns in bullish or bearish formations with the purpose of building a classifier. More elaborate techniques to connect patterns and price predictions could be explored, however, naive classifiers should always be examined first to settle a benchmark for future extensions.

[1] http:\\www.bookmap.com.

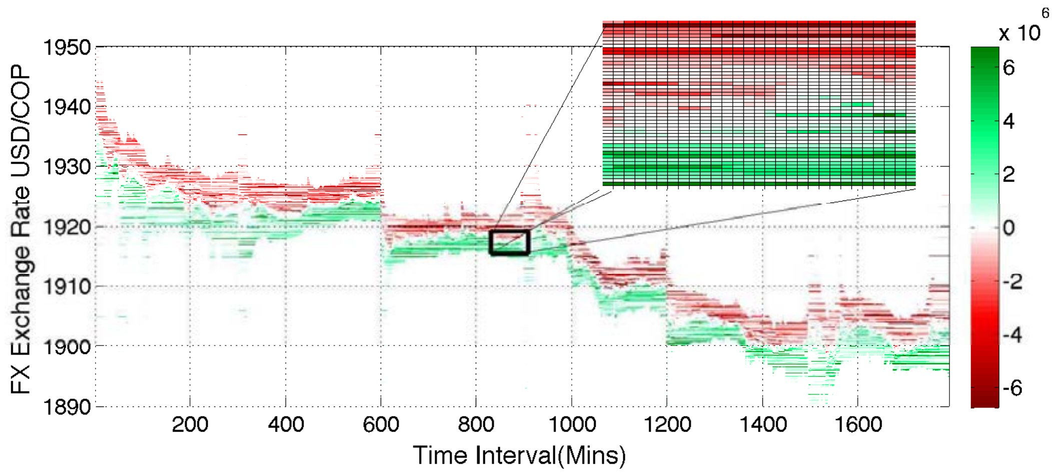


Figure 2: Example of the Visual Representation used over the USD/COP LOB data. Y-axis represents USD/COP exchange rates and x-axis captures time intervals using an aggregating one-min interval. Red means sell volume and green represents buy volume.

2.1 Pattern Exploration and Cluster Analysis.

The original 6-month dataset of LOB information was divided into two subsets. The first 70% of the data was used to train a K-means algorithm and to learn naive market trend abilities for each cluster. The last 30% of available data was used to test the generalization ability of the naive classifier based on LOB image clusters.

As shown in Figure 3, the training subset LOB was used to build non-overlapping images of the LOB centered around the bid-ask spread valid during each image. In detail, every image patch represents 30 units of the aggregating interval chosen (in this case 1 Min) and 30 units of the discretized price grid centered around the mid-price, calculated using the best quotes for every image patch.

Figure 4 shows a sample of images obtained following previous procedure. No pre-processing technique was run over the image training sample. Therefore, we have built raw LOB images centered around the bid-ask spread representing current volume dynamics in the most interesting part of the LOB. Images on Figure 4 use green for buy volume and red for sell volume. The highest the color intensity the highest the volume at that particular price value. White spaces correspond to an absence of volume in the bid-ask spread.

The resulting images of the training subset correspond to 875 visual LOB patches. Over these patches, We run a K-means algorithm using Euclidean distance and $K = 13$. The number of clusters was experimentally chosen in order to balance the dimensionality reduction process and the ability to predict future market trends affected by clustering overfitting. If the cluster number was too low, the dimensionality reduction effect is high, but the trend prediction ability is low and vice-versa.

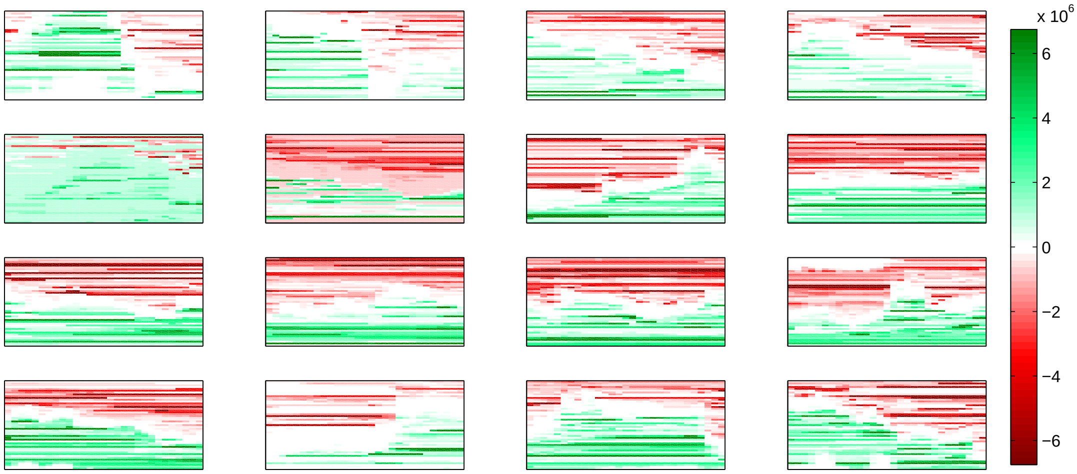


Figure 3: 30×30 volume windows representing 30 min and 3 COP up and down from every mid-price calculated using lowest best buy price and highest best sell price during the selected time window. Red means sell volume and green represents buy volume.

Figure 4 presents graphically the 13 cluster centroids in the lowest-most row and a sample of LOB image patches that belong to each image class. The cluster model learned that there were certain LOB patches with a strong volume on one side and certain price patterns. For example, Cluster 1 shows uptrend prices with higher sell than buy volume. Cluster 2 shows the opposite case.

Cluster classes showed a similar frequency of occurrence except for two cases, class 10 and 5. Class 10 corresponds to an initial uptrend followed by ranging prices in a context of light volume and class 5 corresponds to an initial downtrend followed by stable prices also in a context of low volume. See Figure 5.

After obtaining clusters centroids, We replaced every sample by its corresponding class centroid and determined the connection between each cluster class and a price trend during the next time window. This information was used to construct a naive classifier based on current visual LOB information observed.

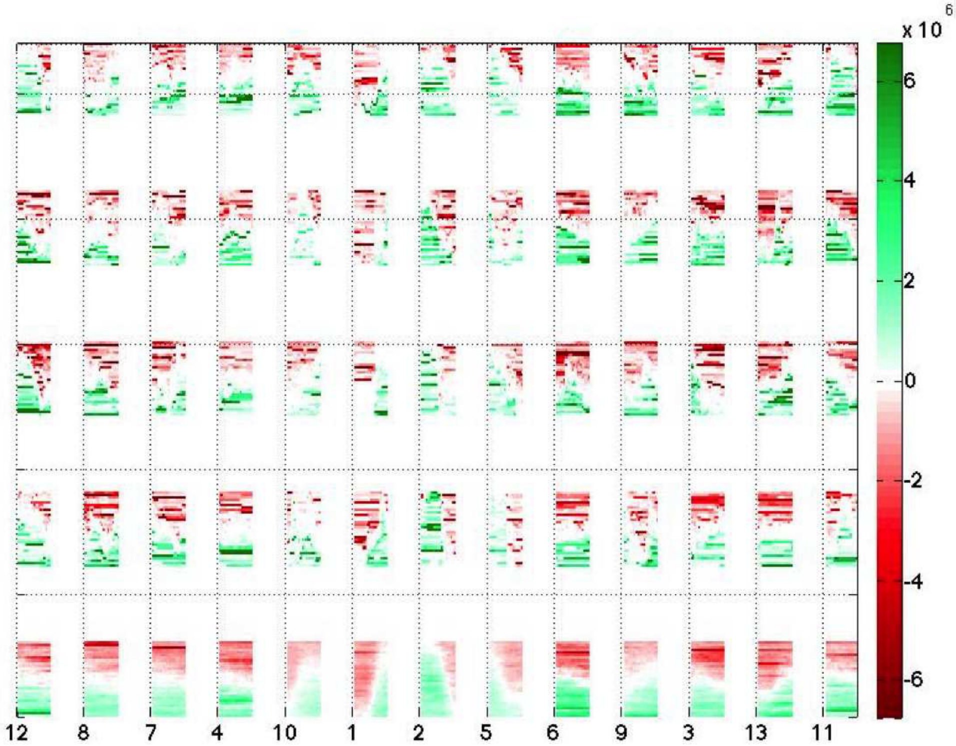


Figure 4: Clustering of visual patches using K-means and 13 clusters. Bottom-most row presents a visual representation of every cluster's centroid. Clusters in the x-axis are organized according to their probability of being associated to a particular market trend. Clusters $12, 8, 7, 4, 10$ are related to uptrend movements, clusters $5, 6, 9, 3, 13, 11$ are related to downtrend movements and clusters $1, 2$ showed a weak tendency to downtrend regimes.

Figure 6 presents conditional probabilities for all clusters organized in descending order, according to the conditional probability of observing a positive trend in the subsequent LOB visual patch i.e. during the next 30 mins. The training data shows that cluster 12 had the highest prediction power for uptrend movements and cluster 11 had the highest prediction power for downtrend movements. There were also found some cluster classes that do not give clear evidence of being related to a particular market trend.

2.2 Testing Performance of Cluster Analysis

After obtaining 13 cluster classes and their corresponding subsequent market trend sign, the last 30% LOB observations, 375 samples were used to test the generalization ability of the selected cluster model. We classified market trend in the subsequent image of every visual patch and compared forecasting results with realized price directions. Table 1 presents the corresponding Confusion Matrix calculated over the testing dataset. The True positive rates correspond to 56.84% and 58.39% for up-trend and down-trend respectively. Moreover, as presented on Table

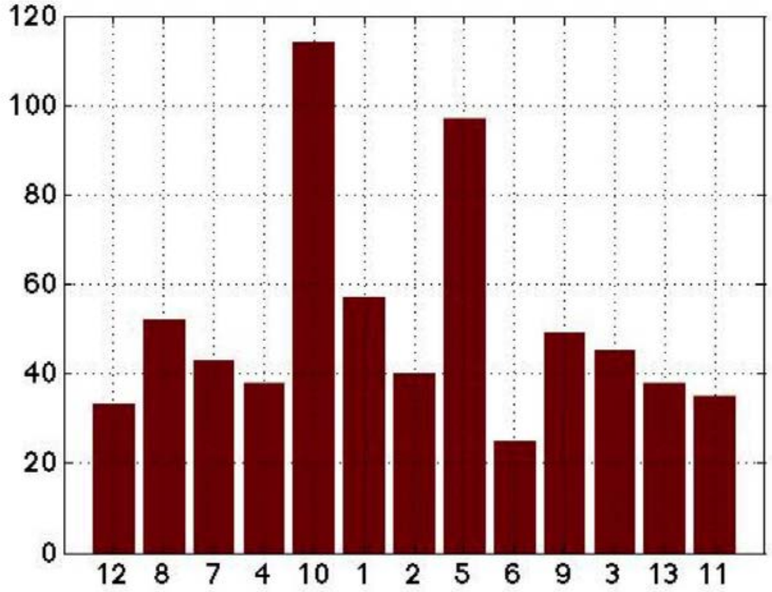


Figure 5: Histogram showing how frequent patterns from each clusters are during the training sample. X-axis presents each cluster and y-axis is observed frequency during training sample.

2, original interpretation of market trend for classes found in the training set, also persisted during the testing set. The naive classifier showed a market trend prediction ability superior to a random classification. These results will serve as benchmarks for future extensions focused on taking advantage of image details to improve model's prediction ability.

	UpTrend Outcome	DownTrend Outcome
UpTrend Prediction	56.84%	43.16%
DownTrend Prediction	41.61%	58.39%

Table 1: Prediction ability of the naive classifier calculated over the testing data.

	Training Set	Testing Set
UpTrend	$12, 8, 7, 4, 10$	$4, 8, 7, 12, 10$
DownTrend	$11, 13, 3, 9, 6, 5, 2, 1$	$9, 6, 5, 2, 13, 11, 3, 1$

Table 2: Market Trend and Cluster association calculated over the training and testing set. Classes are presented descendently in terms of prediction strength

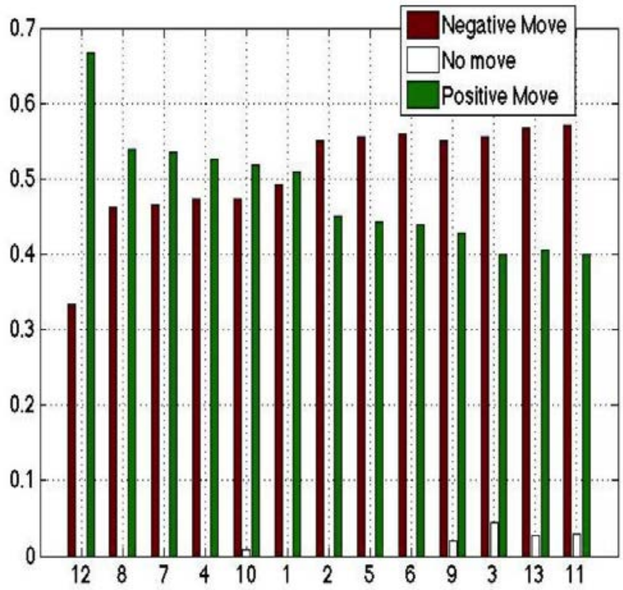


Figure 6: Conditional probability distribution of future price movement in next 30-min time window conditional to the current observation of elements from each cluster. X-axis presents patterns from each cluster and y-axis shows observed probability during training sample.

3 Conclusions

This work introduced a novelty visual representation of the LOB used to trained a cluster model aimed at forecasting short-term financial market trends. The model was tested on LOB information from the USD/COP using 13 clusters, which were aligned with their corresponding future market trend. The model's generalization ability was tested on the last 30% observations of the dataset using this naive classifier. Every class in the testing set kept its trend association found during the training phrase. Moreover, the confusion matrix of the proposed classifier showed superior classification results than random guessing. We expect these results to be a benchmark for future extensions that will focus on learning from details of image patches to improve model's forecasting ability.

4 Acknowledgment

The authors thank Algocodex for providing the dataset used in this work.

References

[1] Hee Joon Ahn, Jun Cai, and Yan Leung Cheung. Price clustering on the limit-order book: Evidence from the Stock Exchange of Hong Kong. *Journal of Financial Markets*, 8(4):421–451, 2005.

[2] Wei Cheng, Shan Cun Liu, He Ying Jiao, and Wan Hua Qiu. How does limit order book information affect trading strategy and market quality: Simulations of an agent-based stock market. *Proceedings - International Conference on Management and Service Science, MASS 2009*, 2009.

[3] J. Doyne Farmer, Paolo Patelli, and Ilija I. Zovko. The Predictive Power of Zero Intelligence in Financial Markets. 2003.

[4] Martin D Gould, Mason a Porter, Stacy Williams, Mark McDonald, Daniel J Fenn, and Sam D Howison. Limit Order Books. *Quantitative Finance*, 13(11):42, 2010.

[5] He Huang and Alec N Kercheval. A generalized birth-death stochastic model for high-frequency order book dynamics. *Quantitative Finance*, 12(April 2012):547–557, 2012.

[6] Jian Jiang and Wing Lon Ng. Revealing intraday market efficiency - Estimating diurnal price densities in limit order books. *Proceedings - 2009 International Conference on Information and Financial Engineering, ICIFE 2009*, (1985):8–12, 2009.

[7] Mario Onorato and Bunhill Row. An integrated pricing model for defaultable loans and bonds. *European Financial Management*, pages 1–20.

[8] D Palguna and I Pollak. Non-parametric Prediction of the Mid-price Dynamics in a Limit Order. *IEEE Statistical Signal Processing Workshop*, pages 896–899, 2012.

[9] Roberto Pascual and David Veredas. What pieces of limit order book information matter in explaining order choice by patient and impatient traders? *Quantitative Finance*, 9(5):527–545, 2009.

[10] Angelo Ranaldo. Order aggressiveness in limit order book markets. *Journal of Financial Markets*, 7(1):53–74, 2004.

[11] N Vvedenskaya, Y Suhov, and V Belitsky. A non-linear model of limit order book dynamics. *2011 IEEE International Symposium on Information Theory Proceedings*, pages 1260–1262, 2011.

[12] Ming-Chang Wang, Lon-Ping Zu, and Chau-Jung Kuo. The state of the electronic limit order book, order aggressiveness and price formation. *Asia-Pacific Journal of Financial Studies*, 37(2):245–296, 2008.

[13] Kilian Weinberger, Anirban Dasgupta, Josh Attenberg, John Langford, and Alex Smola. Feature Hashing for Large Scale Multitask Learning. *Proceedings of the 26th Annual International Conference on Machine Learning*, (Icml):(pp. 1113–1120)., 2009.