Scaling and universality in the position profiles of order cancellations in an emerging stock market

Gao-Feng Gu[1,2], **Fei Ren**[1,2,3] **and Wei-Xing Zhou**[1,2,3]

[1] School of Business, East China University of Science and Technology, Shanghai 200237, China

[2] Research Center for Econophysics, East China University of Science and Technology, Shanghai 200237, China

[3] School of Science, East China University of Science and Technology, Shanghai 200237, China

E-mail: `wxzhou@ecust.edu.cn`

Abstract. We have studied the empirical distribution of cancellation positions through rebuilding the limit-order book using the order flow data of 23 liquid stocks traded on the Shenzhen Stock Exchange in the year 2003. We find that the probability density function (PDF) of relative price levels where cancellations allocate obeys the log-normal distribution. We then analyze the PDF of normalized relative price levels by removing the factor of order numbers stored at the price level, and find that the PDF has a power-law behavior in the tails for both buy and sell orders. When we focus on the probability distribution of cancellation positions at a certain price level, we find that the PDF increases rapidly in the front of the queue, and then fluctuates around a constat value until the end of the queue. In addtion, the PDF of cancellation positions can be fitted by the exponent function for both buy and sell orders.

Keywords: Econophysics, Cancellation, Probability distribution, Limit-order book

PACS numbers: 89.65.Gh, 02.50.-r, 89.90.+n

1. Introduction

In order-driven markets, order cancellation plays an important role in the price formation and makes significant contribution to the order book activity. When limit orders at the best price canceled completely, mid-price (the mean of best bid and best ask) and spread will change as well. If cancellation occurs inside the limit-order book (LOB), it also affects the shape of LOB and has potential effects on price formation.

The motivation of order cancellation is to avoid risk which includes non-execution (NE) risk and free option (FO) risk. The former risk arises when the current price moves away from the submitting price, and limit orders may not to be executed immediately, which may cause traders to suffer the opportunity cost [1]. In order to avoid NE risk, traders may submit more aggressive limit orders or market orders to increase the execution probability. The latter risk arises when good or bad news arrives, which drives the intrinsic value of asset underestimated or overestimated based on the current market price. Traders cancel their stale orders in time to prevent them from being traded at unfavorable prices. Thus, if traders concern the NE risk, they will revise their orders to increase the price priority, while if traders focus on the FO risk, they will revise orders to decrease the price priority. In the Australian security market NE risk is the major reason for order cancellation, especially for large orders [2].

With the reason that order cancellation concerns the dynamics of LOB and there are not enough cancellation data in the past, only a few papers refer to the study of order cancellation. However, in the recent years, with the development of information technology and rapid development of electronic trading venues, traders can easily make an order cancellation at their private computers, so it causes a striking increase of order cancellation in the financial markets. For example, more than 20% of submitted orders are cancellation orders, and 40% of limit orders are canceled on the New York Stock Exchange [3]. The percentage of cancellation is even higher in the futures market [4], in which the cancellation rate is as high as 68.3%.

Yeo [3] studied the database of 148 stocks traded on the New York Stock Exchange in the year 2001 and found that about 95% order cancellations occur within 10 minutes after submission. He also reported that traders tend to resubmit limit orders with more aggressive prices after a cancelation to avoid NE risk. Fong and Liu [2] studied the order flow data of 40 stocks listed on the Australian Securities Exchange (ASX) in August 2000. They found that half cancellations occur within the first two price levels in the LOB. They also indicated that cancellation probability increases with order size, but decreases with order aggressiveness. Liu[5] presented a simple model for order revision and cancellation. The paper shows that cancellation is positively related to order submission risks, and that large capitalization stocks tend to have more cancellations. Hasbrouck and Saar [6] focused on the cancellations on the Island electronic communications network (ECN). They observed that a large number of limit orders are canceled very shortly after their submissions and two sharp jumps appear in the probability density function (PDF) of order cancellation. Biais *et al* [7] analyzed the database of 40 stocks in 1991 in the Paris Bourse. They found that cancellations at the bid or ask side are relatively frequent after large sales or purchases. The paper also shows that cancellations follow each other quickly on the same side of the LOB. Ellul *et al.* [8, 9] studied the data sample of 148 stocks trading on the NYSE. They find that when a buy or sell order is canceled, the most likely subsequent event is the arrival of the same kind of limit orders.

The relationship between cancellation and bid-ask spread has been studied wildly, while the results are controversial. Some scholars concluded that cancellation is positively related to spread [2, 3]. However, others found that cancellation is more likely to happen when spread is tight [5, 8, 9]. Order cancellation is more frequent at the opening and closing time on a trading day, displaying a U-shape intra-day pattern [2, 4, 5, 7, 9, 8].

In the paper, we will study the empirical distribution properties of cancellation positions by rebuilding the LOBs using 23 liquid stocks traded on the Shenzhen Stock Exchange (SZSE) in the Chinese stock market. The rest of paper is organized as follows. In Section 2, we will describe briefly the database we adopt. Section 3 presents the probability distributions of price levels where cancellations allocated in the LOB. We further focus on the cancellation position at a certain price level and study its probability distribution in Section 4. Section 5 summarizes the results.

2. Data sets

In the year 2003, there were two kinds of auctions held on the Shenzhen Stock Exchange, called opening call auction and continuous auction. Opening call auction is held to generate the opening price at the beginning of a trading day. It corresponds to the process of one-time centralized matching of buy and sell orders accepted during the period from 9:15 am and 9:25 am. Continuous auction operates from 9:30 am to 11:30 am and 13:00 pm to 15:00 pm. It refers to the process of continuous matching of buy and sell orders on a one-by-one basis. The interval between opening call auction and continuous auction (9:25 am - 9:30 am) is the cool period, when the Exchange is open to orders routing from members, but does not process orders or process cancelations. Closing call auction (14:57 pm - 15:00 pm) was not held until July 1 2006 to generate the closing price.

In an order-driven market, LOB is a queue of limit orders waiting to be executed according to the price-time priority, and it is the base of continuous double auction mechanism. Price levels in the LOB are discrete. The difference between two adjacent price levels is the tick size or its multiple. In the Chinese stock market, tick size is 0.01 CNY for all the A shares of stocks. Figure 1 presents the structure of LOB. In the figure, we denote x as the price level in the LOB. In the buy LOB, orders with higher limit prices have executing priority and are stored in the front of the LOB, so x decreases with buy price. In contrast, in the sell LOB, priority is given to the orders with lower limit prices, and x increases with the sell price. The highest buy price allocated at the first price level in the buy LOB ($x = 1$) is called the best bid, and the lowest sell price allocated at the first price level in the sell LOB ($x = 1$) is called the best ask. Spread is defined as the difference between the best bid and best ask. When limit orders are submitted with the same price, they are also stored in a queue at a certain price level according to the arriving time. Denote y as the sequence of orders' arriving time, and the smaller value of y indicates the earlier arriving time among the orders with the same submitting price.

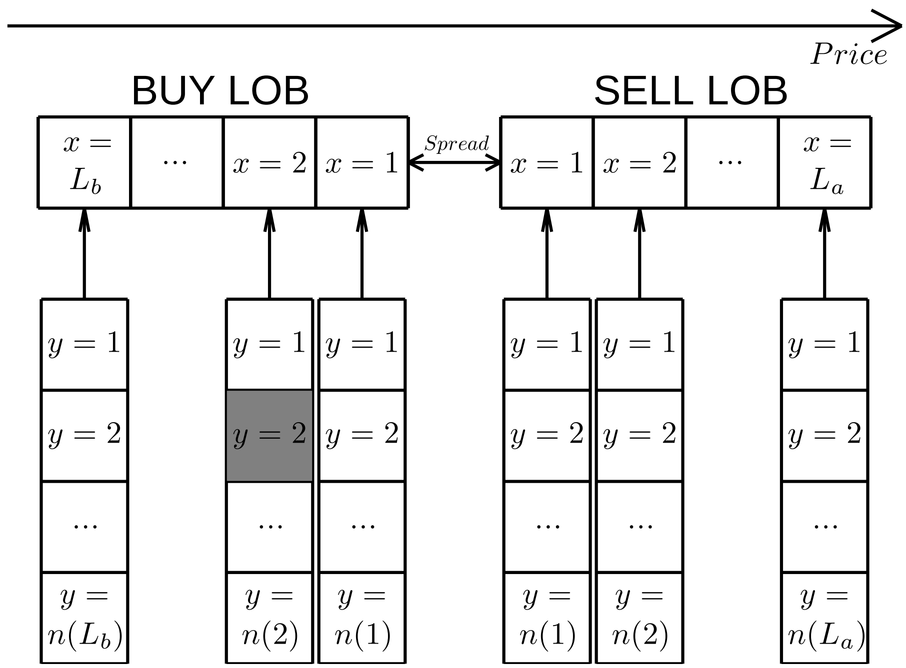

Figure 1. The structure of limit order book.

Our study is based on the order flow data of 23 liquid stocks extracted from the A-share market on the SZSE in 2003. The database contains details of order placement and order cancellation with the time stamp accurate to 0.01 second. In the paper we focus on the empirical distribution of cancellation positions in the LOB in the continuous auction. Table 1 depicts the number of cancellations $C_{b,s}$ for both buy orders and sell orders of 23 stocks. We also calculate the ratio $r_{b,s}$ of the number of cancellations to the number of order placement for the same kind order and find that the ratio fluctuates within a wide range, with [4.2%, 24.7%] for buy orders and [4.5%, 24.3%] for sell orders. There is an interesting feature that the ratio of buy orders r_b is close to sell orders r_s for each stock, which means that a large proportion of buy orders canceled corresponds to a large number of cancellations taking place for sell orders, and vice versa. For all the stocks, we have the mean values $\overline{r}_b = 16.1\%$ for buy orders and $\overline{r}_s = 17.1\%$ for sell orders.

The cancellation probability is related to the order aggressiveness of submission. Table 2 presents the proportions of cancellation of different aggressiveness for both buy and sell orders of 23 stocks. We study four kinds of cancellation orders according to

Table 1. The number of cancellations for buy orders and sell orders in the continuous auction for 23 stocks. The columns from left to right show the stock code, the number of cancellations for buy orders (C_b), the ratio of the number of cancellations to the number of order placement for buy orders (r_b), the number of cancellations for sell orders (C_s), the ratio of the number of cancellations to the number of order placement for sell orders (r_s).

Stock	Buy orders		Sell orders		Stock	Buy orders		Sell orders	
	C_b	$r_b(\%)$	C_s	$r_s(\%)$		C_b	$r_b(\%)$	C_s	$r_s(\%)$
000001	317,015	19.2	274,929	18.3	000406	94,100	20.3	91,776	20.1
000002	34,577	4.2	43,801	4.9	000429	36,999	20.1	36,174	17.4
000009	183,804	21.9	187,421	20.7	000488	32,585	18.2	33,854	18.4
000012	114,662	24.7	106,361	24.3	000539	26,950	15.0	27,087	15.9
000016	60,219	21.1	59,189	19.2	000541	19,715	18.8	19,847	16.6
000021	157,174	23.8	152,853	22.0	000550	122,865	23.6	129,606	22.9
000024	42,593	22.3	45,594	19.6	000581	27,236	18.6	29,490	16.2
000027	33,058	7.1	38,647	5.8	000625	123,361	22.2	131,719	22.6
000063	25,681	7.2	34,815	7.5	000709	65,704	18.7	64,017	17.1
000066	110,289	24.3	106,695	21.8	000720	16,558	11.4	14,209	10.7
000088	6,861	5.6	6,917	4.5	000778	43,576	18.4	47,360	16.4
000089	20,909	7.0	22,984	7.6					

their submission aggressiveness, that is, unfilled orders, orders placed in the bid-ask spread, orders placed at the same best price, and orders placed inside the LOB which are marked as r_1, r_2, r_3 and r_4 in the table. It is shown that the cancellation probability decreases with the submission aggressiveness. Most of the cancellations take place inside the LOB, and only a few orders canceled when they are not fully executed or placed in the spread. For the 23 stocks analyzed, we have $\overline{r}_{b,1} = 1.1\%$, $\overline{r}_{b,2} = 3.3\%$, $\overline{r}_{b,3} = 24.4\%$ and $\overline{r}_{b,4} = 71.2\%$ for buy orders, and $\overline{r}_{s,1} = 1.0\%$, $\overline{r}_{s,2} = 2.9\%$, $\overline{r}_{s,3} = 21.5\%$ and $\overline{r}_{s,4} = 74.6\%$ for sell orders.

3. The PDF of cancellation positions in the limit-order book

In this section, we will study the probability distribution of cancellation position in the LOB. Denote $x(t)$ as the price level where a cancellation allocated in the LOB at time t. Here we use the event time t, i.e. when a cancelation happens, it increases by one ($t = t + 1$). Suppose an order which has been marked with gray color in Figure 1 is canceled at time t. The canceled order is stored at the second price level in the buy LOB, so we have $x(t) = 2$. Generally speaking, $x(t) = n$ indicates a cancellation happening at the n-th price level at time t in the buy or sell LOBs.

The length of LOB ($L_b(t)$ or $L_s(t)$), defined as the number of price levels existing in the buy or sell LOB, plays an important role in the behavior of order cancellation

Table 2. The proportion of order cancellation according to the submission aggressiveness for 23 stocks. The columns show the stock code, the proportions of unfilled buy orders ($r_{b,1}$) and sell orders ($r_{s,1}$) canceled, the proportions of buy orders ($r_{b,2}$) and sell orders $r_{s,2}$) canceled in the bid-ask spread, the proportions of buy orders ($r_{b,3}$) and sell orders ($r_{s,3}$) canceled at the same best price, and the proportions of buy orders ($r_{b,4}$) and sell orders ($r_{s,4}$) canceled inside the LOB.

Stock	Buy orders (%)				Sell orders(%)			
	$r_{b,1}$	$r_{b,2}$	$r_{b,3}$	$r_{b,4}$	$r_{s,1}$	$r_{s,2}$	$r_{s,3}$	$r_{s,4}$
000001	0.9	1.5	18.4	79.2	0.7	0.7	18.0	80.6
000002	0.9	1.1	25.9	72.2	0.6	0.7	22.4	76.3
000009	0.8	1.1	24.9	73.2	0.5	0.5	21.6	77.4
000012	0.7	2.2	22.6	74.5	0.8	2.4	21.6	75.2
000016	0.7	2.8	24.8	71.7	0.6	2.4	23.7	73.4
000021	0.8	2.8	20.2	76.2	0.6	2.2	19.0	78.2
000024	0.6	3.7	25.7	70.0	0.6	3.7	22.0	73.7
000027	1.6	1.6	23.6	73.2	1.1	1.2	18.3	79.4
000063	1.8	4.4	21.2	72.5	2.6	3.7	17.5	76.2
000066	0.9	3.1	22.1	73.9	0.9	2.2	20.5	76.4
000088	2.2	10.3	29.4	58.1	1.2	9.2	21.0	68.6
000089	1.4	2.8	24.1	71.7	2.0	2.6	22.0	73.3
000406	0.9	1.2	24.5	73.4	1.3	1.3	22.7	74.7
000429	0.8	1.8	28.9	68.5	0.8	1.6	25.4	72.2
000488	1.4	5.7	19.2	73.7	0.9	3.9	20.0	75.2
000539	1.0	3.9	21.9	73.2	1.1	4.1	20.3	74.5
000541	1.1	4.7	23.2	70.9	1.2	5.0	24.2	69.5
000550	1.1	3.5	21.3	74.2	1.3	2.3	18.7	77.7
000581	1.1	5.0	24.5	69.3	1.4	4.7	19.1	74.8
000625	1.3	3.5	20.3	74.8	1.1	3.0	17.2	78.7
000709	0.8	0.8	29.2	69.2	0.6	0.7	26.1	72.6
000720	1.0	6.7	40.9	51.4	1.1	4.9	30.3	63.7
000778	0.6	2.5	24.2	72.7	0.8	2.6	22.2	74.5

process. For example, it is a completely different situation that a trader cancels an order at the second price level ($x(t) = 2$) when $L(t) = 5$ from the situation that he cancels an order at the second price level when $L(t) = 50$. It is trivial to study the probability distribution of absolute values of $x(t)$, and we focus on the relative price levels $X(t)$ instead, which is defined as follows

$$X(t) = \frac{x(t)}{L_{b,s}(t)} , \tag{1}$$

where $X(t)$ varies in the range $(0, 1]$. According to the definition, we know that a small value of X means the canceled order allocated in the front of LOB, while a large value of X suggests cancellation happening at the end of LOB.

We analyze the probability density function (PDF) $f(X)$ of relative price levels where cancellations allocate for both sell and buy LOBs. In figure 2, we illustrate the probability distributions $f(X)$ for four stocks randomly chosen from 23 stocks we analyzed. Since the order waiting in the LOB is executed only if all the orders ahead have been removed, that is, the orders at the end of the LOB are less to be executed and the cancellation probabilities increase as well, we expect that $f(X)$ should increase monotonically with respect to the price level X. However, the probability density functions $f(X)$ first increase sharply, and then fall down rapidly for both buy orders and sell orders, which deviates from our expectation. The special shape of PDF $f(X)$ may be cause by the reason that in the Chinese stock market many impatient traders place aggressive orders close to the same best price to increase the transaction probability, which is confirmed by the fact that the limit-order book shape has a maximum away from the same best price ($X = 1$) but extremely close to the same best price [10]. Once the impatient traders can not make an immediate transaction, they will cancel orders to catch the price in order to avoid NE risk, which results in a large cancellation probability in the front of the LOB. In addition, the PDFs $f(X)$ are not identical between buy orders and sell orders. For sell orders, they have higher cancelation probability in the front and at the end of the LOB, however buy orders have higher cancelation probability in the middle of the LOB. The sharply increasing value of $f(X)$ located at $X = 1$ may be caused by the 10% price limit trading rules in the Chinese stock market.

In figure 2, we find that the PDF $f(X)$ of relative price levels follows a log-normal distribution, that is,

$$f(X) = \frac{1}{\sqrt{2\pi}\sigma X} \exp\left[-\frac{(\ln X - \mu)^2}{2\sigma^2}\right] , \tag{2}$$

where μ is the location parameter and σ is the scaling parameter. We also analyze the remaining stocks and find that their PDFs $f(X)$ are all log-normally distributed. Using the least squares fitting method, we calculate the parameters μ and σ for buy and sell LOBs of 23 stocks, which are listed in table 3. In order to compare the performance of log-normal distribution fit, we introduce a new statistical variable χ, defined as the root mean square (r.m.s) of the difference between the best fit and the empirical data, which is depicted in table 3 as well. For the 23 stocks we calculate the mean values of μ and σ, and have $\overline{\mu} = -2.06 \pm 0.16$ and $\overline{\sigma} = 1.10 \pm 0.07$ for buy orders and $\overline{\mu} = -2.32 \pm 0.12$ and $\overline{\sigma} = 1.29 \pm 0.09$ for sell orders. The values of μ obtained form the PDFs of sell orders are all smaller than buy orders, while the values of σ of sell orders are larger than buy orders. It means that there exists a higher peak in the PDF $f(X)$ of sell orders on the one hand, and the peak is more closer to the same best price in sell LOB than buy LOB on the other hand. It implies sellers are more eager to make a transaction immediately, which is consistent with the situation that it was a bear stock market during the year 2003 and market participants were more willing to sell their orders.

Since all probability density functions of relative price levels of 23 stocks are log-normally distributed, we aggregate the 23-stock data together and treat all the stocks

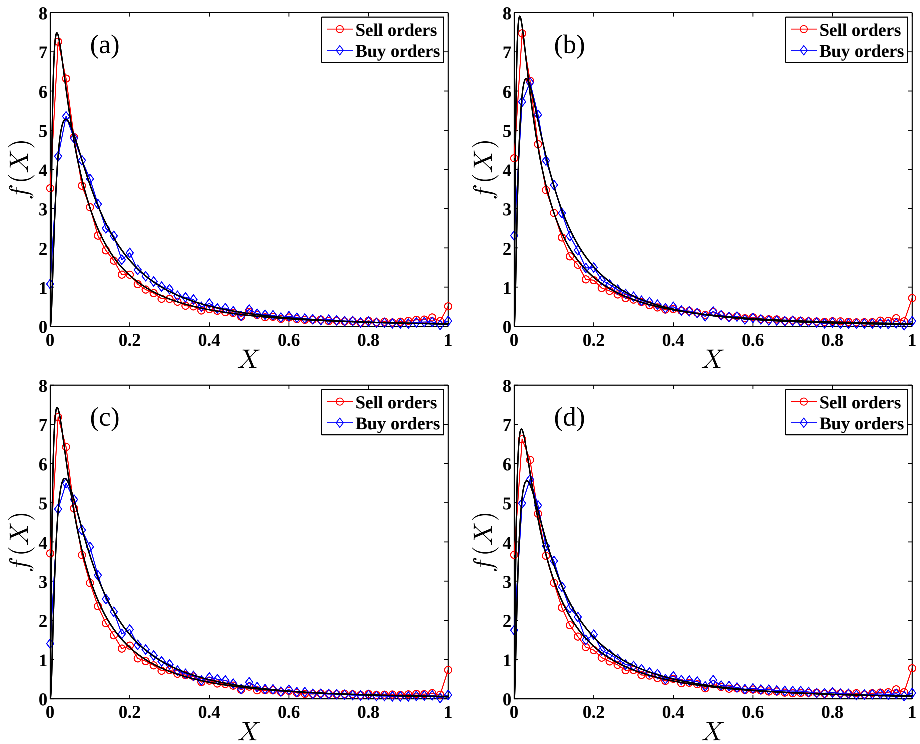

Figure 2. (Color online) Probability distributions $f(X)$ of relative price levels for both buy orders and sell orders of four stocks 000012 (a), 000021 (b), 000066 (c) and 0000625 (d). The solid lines are log-normally distributed and determined by the least squares fitting method.

as an ensemble in order to obtain a better statistic. As expected, the PDF $f(X)$ of ensemble relative price levels obey the log-normal distribution for both buy orders and sell orders. Using the least squares fitting method, we obtain the parameters $\mu = -2.17$ and $\sigma = 1.09$ with the r.m.s $\chi = 0.07$ for buy orders, and $\mu = -2.35$ and $\sigma = 1.31$ with the r.m.s $\chi = 0.09$ for sell orders.

Except for the factor of the total price levels in the LOB, the number of orders stored in the LOB also affects traders' cancelation strategy. especially the number of orders stored at the price level at which the cancellation takes place. Denote $n(x,t)$ as the number of orders stored at price level x at time t. The cancelation probability of an order at price level $x(t)$ is related to $n(x,t)$ as well. We expect that the cancelation probability increases with $n(x,t)$. In order to remove this impact, we define the normalized relative price levels $\widehat{X}(t)$ by

$$\widehat{X}(t) = X(t) \Bigg/ \frac{n(x,t)}{\sum_{x=1}^{L_{b,s}(t)} n(x,t)} \quad , \tag{3}$$

where $L_{s,b}(t)$ is the length of buy or sell LOB, and $\sum_{x=1}^{L_{b,s}(t)} n(x,t)$ indicates the total number of orders stored in the buy or sell LOB at time t.

In figure 3, we present the probability distribution $f(\widehat{X})$ of normalized relative price levels $\widehat{X}$ for both buy and sell orders of four stocks. We find that the PDFs $f(\widehat{X})$ first

Table 3. Characteristic parameters of probability distributions of relative price levels X for both buy and sell orders of 23 stocks. μ and σ are the parameters of log-normal distribution for relative price levels X using the least squares fitting method, and χ is the r.m.s of the difference between the best fit and the empirical data applying the log-normal distribution.

Stock	Buy orders			Sell orders		
	μ	σ	χ	μ	σ	χ
000001	-2.38	1.11	0.11	-2.54	1.45	0.10
000002	-2.20	1.08	0.34	-2.40	1.20	0.09
000009	-2.25	1.00	0.21	-2.26	1.22	0.15
000012	-2.08	1.08	0.07	-2.34	1.32	0.08
000016	-2.03	1.05	0.16	-2.31	1.23	0.09
000021	-2.25	1.12	0.08	-2.36	1.38	0.08
000024	-1.98	1.10	0.11	-2.30	1.26	0.08
000027	-2.16	1.02	0.10	-2.30	1.21	0.10
000063	-2.08	1.19	0.07	-2.34	1.42	0.10
000066	-2.14	1.08	0.07	-2.34	1.30	0.09
000088	-1.70	1.14	0.26	-2.17	1.35	0.10
000089	-2.05	1.11	0.18	-2.35	1.21	0.10
000406	-2.25	1.02	0.09	-2.43	1.24	0.13
000429	-2.01	0.97	0.19	-2.36	1.16	0.08
000488	-1.97	1.21	0.13	-2.19	1.29	0.09
000539	-1.87	1.10	0.22	-2.17	1.31	0.07
000541	-1.90	1.18	0.19	-2.45	1.34	0.13
000550	-2.05	1.10	0.07	-2.27	1.29	0.09
000581	-1.82	1.09	0.17	-2.30	1.29	0.08
000625	-2.11	1.15	0.08	-2.25	1.32	0.10
000709	-2.15	0.99	0.12	-2.54	1.10	0.15
000720	-1.89	1.24	0.15	-1.99	1.41	0.15
000778	-2.06	1.09	0.13	-2.38	1.33	0.09

increase with the normalized relative price level $\widehat{X}$, and then decrease rapidly for buy and sell orders, with the maximum value away from the same best price, which is in line with the results presented in figure 2. In addition, the PDF $f(\widehat{X})$ follows a power-law decay when $\widehat{X} > \widehat{X}_{\rm min}$ for both buy and sell orders, that is,

$$f(\widehat{X} > \widehat{X}_{\rm min}) \sim \widehat{X}^{-\alpha} \tag{4}$$

where α is the power-law tail exponent and $\widehat{X}_{\rm min}$ is the lower threshold of the power-law scaling range.

There are many methods to determine the threshold $\widehat{X}_{\rm min}$ and the power-law tail exponent α. Newman *et al* [11]. recently proposed an efficient quantitative method to

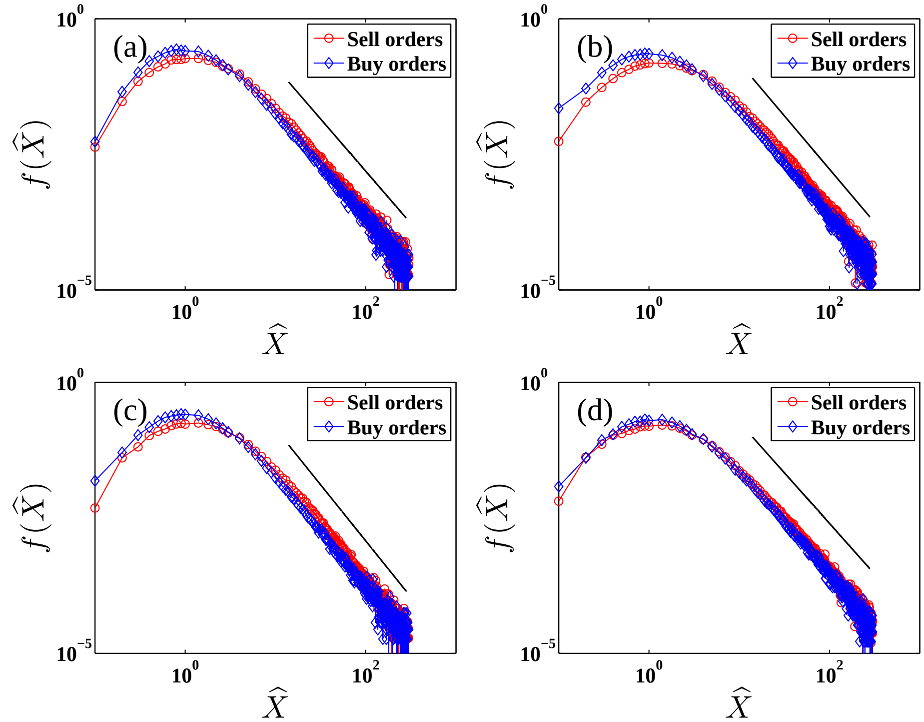

Figure 3. (Color online) Probability distributions $f(\widehat{X})$ of normalized relative price levels $\widehat{X}$ for both buy and sell LOBs of four stocks 000012 (a), 000021 (b), 000066 (c) and 0000625 (d). The solid lines are the fitted functions using the maximum likelihood estimation method.

estimate $\widehat{X}_{\min}$ and α based on the Kolmogorov-Smirnov test. The Kolmogorov-Smirnov statistic (KS) is defined as

$$KS = \max_{\widehat{X} > \widehat{X}_{\min}} \left(|P - F_{PL}| \right) , \tag{5}$$

where P is the cumulative distribution of normalize relative price levels $\widehat{X}$ and F_{PL} is the cumulative distribution of the best power-law fit. We first determine the threshold $\widehat{X}_{\min}$ by minimizing the KS statistic, and then estimate the power-law tail exponent α with the data in the range $\widehat{X} > \widehat{X}_{\min}$ using the maximum likelihood estimation (MLE) method, that is,

$$\alpha = 1 + m \left[\sum_{t=1}^{m} \ln \frac{\widehat{X}(t)}{\widehat{X}_{\min}} \right]^{-1} , \tag{6}$$

where m is the length of the data points in the range $\widehat{X} > \widehat{X}_{\min}$. The standard error $\hat{\sigma}$ on α, which is derived from the width of the likelihood maximum, can be expressed by

$$\hat{\sigma} = \frac{\alpha - 1}{\sqrt{m}} . \tag{7}$$

We study the probability distribution of normalized price levels $\widehat{X}$ for the rest stocks and find that they all obey the power-law distribution in the tail. Applying the method mentioned above, we calculate the characteristic parameters, that is, the tail exponent

α, the threshold $\widehat{X}_{\min}$ and the standard error $\hat{\sigma}$, which are illustrated in table 4. We find that the power-law exponent α varies in the range [1.80, 2.51] for the buy orders and [1.74, 2.42] for the sell orders. The exponents α are all close to 2, with the mean value $\overline{\alpha} = 2.06 \pm 0.16$ for buy orders and $\overline{\alpha} = 2.12 \pm 0.17$ for sell orders. Moreover, for the 23 stocks analyzed, only 8 stocks have the tail exponents of buy orders greater than sell orders.

Table 4. Characteristic parameters of probability distribution of normalized relative price levels $\widehat{X}$ for both buy and sell orders of 23 stocks. α and $\widehat{X}_{\min}$ are the parameters of power-law tail distribution based on the KS-test for the normalized relative price levels $\widehat{X}$ and $\hat{\sigma}$ is the standard error on α using the likelihood maximum estimation method.

Stock	Buy orders			Sell orders		
	α	$\hat{\sigma}$	$\widehat{X}_{\min}$	α	$\hat{\sigma}$	$\widehat{X}_{\min}$
000001	1.82	0.001	4.85	2.09	0.002	10.61
000002	1.96	0.005	4.07	2.16	0.006	8.54
000009	2.01	0.002	4.06	2.35	0.003	10.41
000012	1.96	0.003	5.63	2.07	0.003	9.18
000016	2.07	0.004	5.04	2.19	0.005	11.10
000021	2.05	0.003	11.23	2.19	0.003	9.75
000024	2.10	0.005	7.75	1.90	0.004	3.94
000027	2.08	0.006	7.73	2.06	0.005	10.88
000063	2.18	0.007	18.49	2.26	0.007	19.34
000066	2.10	0.003	9.79	2.22	0.004	18.98
000088	2.51	0.026	19.39	2.42	0.028	22.00
000089	2.11	0.008	6.36	1.93	0.006	3.68
000406	2.01	0.003	4.89	2.13	0.004	19.40
000429	2.10	0.006	3.30	2.19	0.006	15.17
000488	2.45	0.012	15.53	1.98	0.005	17.29
000539	2.04	0.006	5.67	2.23	0.008	27.91
000541	1.95	0.007	3.79	1.79	0.006	2.76
000550	1.99	0.003	6.24	2.19	0.003	13.87
000581	2.12	0.007	7.86	2.01	0.006	11.04
000625	2.00	0.003	9.31	2.18	0.003	12.15
000709	2.02	0.004	3.26	2.23	0.005	12.01
000720	1.80	0.006	3.35	1.74	0.006	3.83
000778	2.07	0.005	7.55	2.19	0.005	18.74

With the reason that the PDFs of normalized relative price levels $\widehat{X}$ for 23 stocks are similar to one another, we trade the 23 stocks as an ensemble and aggregate the data together and calculate the probability distribution aggregating 23 stocks for both buy orders and sell orders. As expected, the ensemble PDF is similar to the individual stock, with the power-law distribution in the tail. Using the method based on maximum likelihood estimation and KS-test, we obtain $\alpha = 2.02$ with $\widehat{X}_{\min} = 7.34$ for buy orders and $\alpha = 2.15$ with $\widehat{X}_{\min} = 10.59$ for sell orders, which are close to the mean values $\overline{\alpha} = 2.06$ and $\overline{\alpha} = 2.12$ mentioned above, respectively.

4. The PDF of cancellation positions at the price level

Limit-order book is constructed by the price-time priority principle which consists of price priority principle and time priority principle. The formal principle has been mentioned in section 3, and in this section we focus on the latter principle, studing the probability distribution of cancellation positions at a certain price level. Time priority principle concerns the orders having the same submitted prices and indicates that the priority is given to the order which is earlier received by the exchange trading system. At a certain price level in the LOB, orders are also stored in a waiting queue according to their arriving time, and front orders in the queue are earlier submitted than the back ones.

Denote $y(x,t)$ as the position where a canceled order allocated in the queue at the x-th price level at time t. For example, the canceled order marked with gray color in figure 1 is located at the second position at the second price level ($x = 2$) in the buy LOB at time t, so we have $y(2,t) = 2$. In order to consider the number effect of orders stored at the x-th price level, we introduce the relative position variable $Y(x,t)$ instead of $y(x,t)$, which reads,

$$Y(x,t) = \frac{y(x,t)}{n_{b,s}(x,t)} \ , \tag{8}$$

where $Y(x,t)$ varies in the range $(0,1]$, and $n_{b,s}(x,t)$ is the number of orders stored at the x-th price level at time t for buy LOB or sell LOB.

Now we focus on the probability distribution of relative positions $Y(x,t)$ which can quantitatively describe the properties of time priority principle. We calculate the PDFs of $Y(x,t)$ at the x-th price level for both buy and sell orders, and the PDFs $f(Y)$ of the first four price levels ($x = 1,\ 2,\ 3,\ 4$) of stock 000001 are illustrated in figure 4. We find that the probability distributions are almost the same. The function $f(Y)$ goes to zero when the relative position Y approaches to zero. As Y increases, $f(Y)$ first increases rapidly in the range $Y \leq 0.1$, then it fluctuates around a constat level until the end of the queue.

In order to approximate the probability distribution $f(Y)$, we empirically apply an exponential function, that is,

$$f(Y) = \frac{1}{z}(1 - e^{\beta Y}) \ , \tag{9}$$

where z is a normalized constant and we have $z = \int_0^1 (1 - e^{\beta x}) \mathrm{d}x$. Using the least squares fitting method, we estimate the parameter β for both buy orders and sell orders of the four stocks, and obtain the parameters $\beta = -30.34$ with $\chi = 0.22$ (χ is the r.m.s of the difference between the best fit and the empirical data, which has mentioned in table 3)

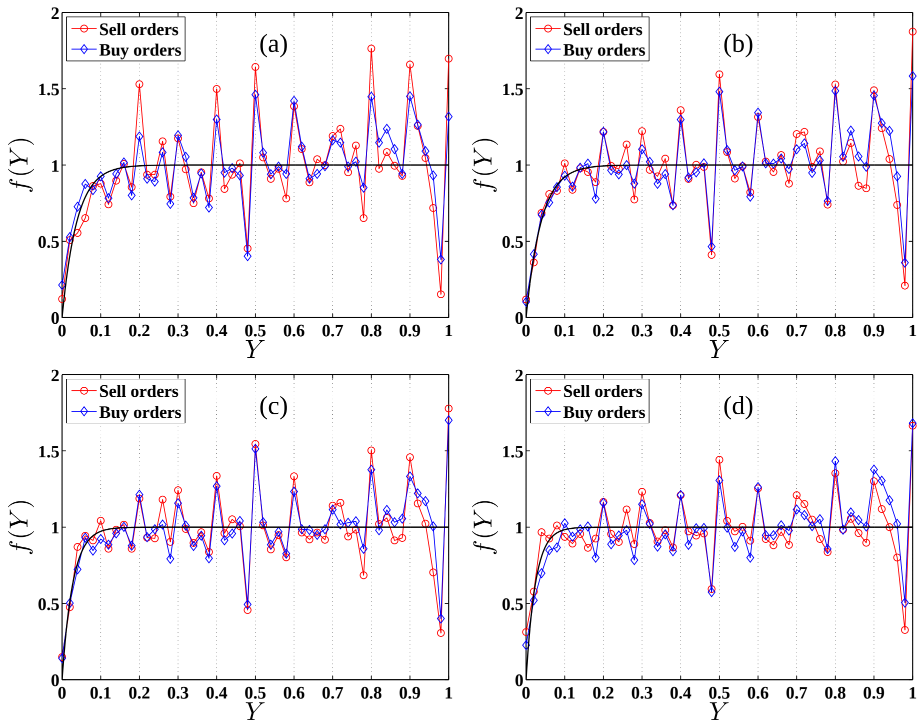

Figure 4. (Color online) Probability density functions $f(Y)$ of relative positions $Y(x,t)$ at the first price level ($x = 1$) (a), second price level ($x = 2$) (b), third price level ($x = 3$) (c) the fourth price level ($x = 4$) (d) for both buy and sell orders of stock 000001. The solid curves are the best fitting function illustrated in Eq. (9).

for the buy orders and $\beta = -21.51$ with $\chi = 0.30$ for the sell orders at the first price level. We have $\beta = -24.89$ with $\chi = 0.22$ for the buy orders and $\beta = -25.78$ with $\chi = 0.26$ for the sell orders at the second price level. We get $\beta = -22.25$ with $\chi = 0.21$ for the buy orders and $\beta = -29.54$ with $\chi = 0.24$ for the sell orders at the third price level. We have $\beta = -25.82$ with $\chi = 0.20$ for the buy orders and $\beta = -20.35$ with $\chi = 0.20$ for the sell orders at the fourth price level.

We also study the PDF $f(Y)$ of relative positions for a randomly chosen price levels and find that it has a similar distribution. So we treat the PDFs $f(Y)$ at all price levels are similar and aggregate the data $Y(x,t)$ at all price levels together. Figure 5 presents the PDFs $f(Y)$ of the relative positions at all price levels for four stocks randomly chosen from the 23 stocks. We find that the total PDF is similar to the PDF of the relative positions at a certain price level mentioned in figure 4.

With the same method we estimate the parameter β for both buy orders and sell orders of the 23 stocks, and results are illustrated in table 5. The mean value of β for buy orders are $\overline{\beta} = -21.49$, and for sell orders it is $\overline{\beta} = -20.92$.

Moreover, we observe an interesting feature that the PDF $f(Y)$ has periodic peaks at $Y = 0.1m$ when $m = 1,\ 2, \cdots,\ 10$ for both buy and sell orders. The underlying mechanism of this periodic behavior is unknown, and it may be caused by the trading strategy of impatient traders or people's irrational number preference, such as 5, 10

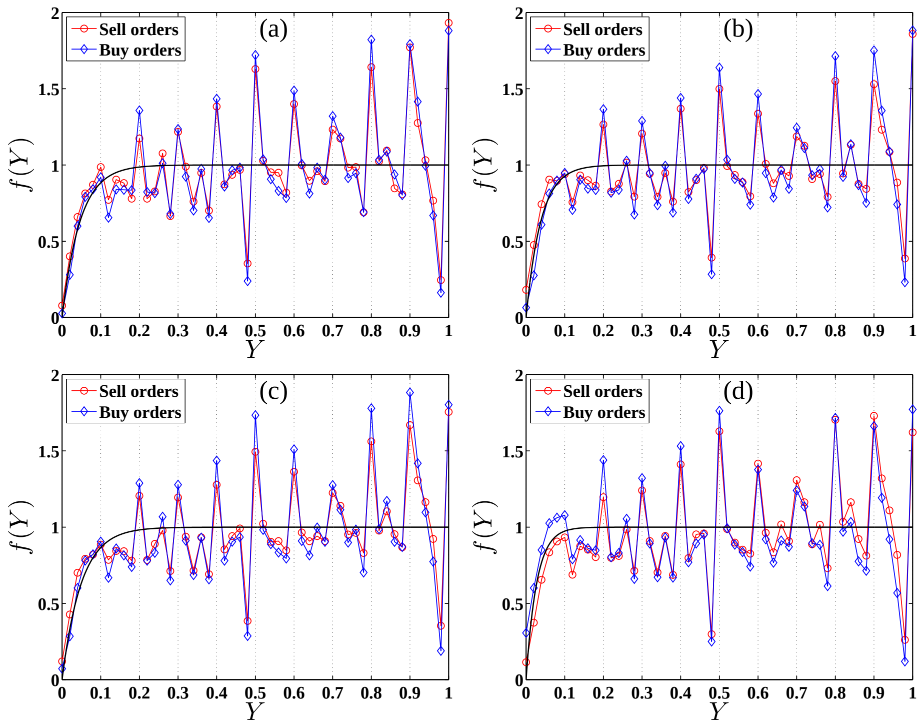

Figure 5. (Color online) Probability density functions $f(Y)$ of aggregating relative positions for both buy and sell orders of four stocks, 000012 (a), 000021 (b), 000066 (c) and 000625 (d). The solid curves are the best fitting function illustrated in Eq. (9).

Table 5. Characteristic parameters of probability distributions of aggregating relative positions Y at all price levels for both buy and sell orders of 23 stocks. β is the parameter of the fit function expressed in Eq. (9) and χ is the r.m.s of the difference between the best fit and the empirical data of relative positions.

Stock	Buy orders		Sell orders		Stock	Buy orders		Sell orders	
	β	χ	β	χ		β	χ	β	χ
000001	-33.78	0.17	-36.57	0.16	000406	-15.12	0.29	-18.94	0.24
000002	-9.42	0.29	-14.35	0.22	000429	-10.15	0.36	-12.42	0.28
000009	-14.07	0.22	-18.90	0.19	000488	-23.40	0.41	-12.32	0.37
000012	-18.81	0.31	-24.12	0.27	000539	-29.23	0.70	-28.82	0.31
000016	-12.72	0.36	-18.21	0.30	000541	-9.55	0.43	-16.41	0.39
000021	-21.26	0.29	-30.91	0.23	000550	-35.81	0.30	-25.01	0.25
000024	-17.67	0.41	-17.23	0.32	000581	-30.32	0.39	-14.85	0.33
000027	-13.00	0.32	-13.04	0.23	000625	-40.17	0.32	-22.77	0.28
000063	-44.70	0.36	-33.26	0.24	000709	-13.25	0.26	-12.91	0.24
000066	-17.81	0.32	-22.36	0.24	000720	-28.33	1.10	-34.40	0.86
000088	-32.43	0.51	-21.70	0.46	000778	-12.38	0.35	-17.15	0.30
000089	-10.89	0.41	-14.46	0.31					

or their multiples [12]. This periodic strip pattern is also observed in the snapshot of stocks traded on the London Stock Exchange [13], in the LOB shape in the Chinese stock market [10], and in the PDF of order sizes in the opening call auction in the Chinese stock market [14]. The interesting feature call for further investigation, which is however beyond the scope of this work.

5. Conclusion

Cancellation plays an important role in the dynamics of price formation in the financial markets, and helps traders avoid NE risk or FO risk. In the paper, we analyze the order flow data of 23 liquid stocks traded on the Shenzhen Stock Exchange in the whole year 2003, and study the empirical distributions of cancellation positions by rebuilding the buy and sell Limit-order books. We first analyze the probability distribution of relative price levels where cancellation allocates and find that the probability density function (PDF) obeys the log-normal distribution. Moreover, Comparing the PDFs of buy order and sell order, we present that the PDF of sell order has a higher peak near to the same best price than the buy orders. In order to remove the number effect, we then study the PDF of cancellation positions for the normalized relative price levels , and find that the PDF follows the power-law behavior in the tail. Using the KS test and MLE method, we calculate the power-law exponent for buy and sell orders, and find that they are both close to 2. Next, we study the probability distribution of cancellation positions at a certain price level, and find that the PDF increases rapidly in the front of the queue at a certain price level and then fluctuates around a constat value until the end of the queue. In addition, the PDF can be fitted by the exponent function for both buy and sell orders.

Acknowledgments:

This work was partly supported by the National Natural Science Foundation of China (Grants No. 71101052, 71131007 and 11075054), the “Chen Guang” project (Grant No. 2010CG32) sponsored by Shanghai Municipal Education Commission and Shanghai Education Development Foundation, and the Fundamental Research Funds for the Central Universities.

References

[1] M. D. Griffiths, B. F. Smith, D. A. S. Turnbull, and R. W. White. The costs and determinants of order aggressiveness. *J. Financ. Econ.*, 56:65–88, 2000.

[2] K. Y.L. Fong and W. M. Liu. Limit order revisions. *J. Behav. Financ.*, 34:1873–1885, 2010.

[3] W. Y. Yeo. Cancellations of limit orders. Working paper, National University of Singapore, 2005.

[4] M. Coppejans and I. Domowitz. Screen information, trader activity, and bid-ask spreads in a limit order market. Working Paper, Duke University, 1999.

[5] W. M. Liu. Monitoring and limit order submission risks. *J. Financ. Markets*, 12:107–141, 2009.

[6] J. Hasbrouck and G. Saar. Limit orders and volatility in a hybrid market: the Island ECN. Working paper, New York University, 2002.

[7] B. Biais, P. Hillion, and C. Spatt. An empirical analysis of the limit order book and the order flow in the Paris Bourse. *J. Financ.*, 50:1655–1689, 1995.

[8] A. Ellul, C. W. Holden, P. Jain, and R. Jennings. Determinants of order choice on the New York Stock Exchange. Working paper, Indiana University, 2003.

[9] A. Ellul, C. W. Holden, P. Jain, and R. Jennings. Order dynamics: recent evidence from the NYSE. *Journal of Empirical Finance*, 14:636–661, 2007.

[10] G.-F. Gu, W. Chen, and W.-X. Zhou. Empirical shape function of limit-order books in the Chinese stock market. *Physica A*, 387:5182–5188, 2008.

[11] A. Clauset, C. R. Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Rev.*, 51:661–703, 2009.

[12] S. N. Dorogovtsev, J. F. F. Mendes, and J. G. Oliveira. Frequency of occurrence of numbers in the World Wide Web. *Physica A*, 360:548–556, 2006.

[13] Z. Eisler, J. Kertész, and F Lillo. The limit order book on different time scales. In J. Kertész, S. Bornholdt, and R. N. Mantegna, editors, *Proc. SPIE*, volume 6601, page 66010G, Florence, Italy, 2007. AIP.

[14] G.-F. Gu, F. Ren, X.-H. Ni, W. Chen, and W.-X. Zhou. Empirical regularities of opening call auction in Chinese stock market. *Physica A*, 389:278–286, 2010.