The Information Content of Limit Order Book and Short-term Stock Returns based on Information Technology

Yaqing Liu[1] and Hongbing Ouyang[2]

[1]*School of Economics and Management, North China University of Technology, China*
E-mail: liu_yaqing@yeah.net
School of Economics, Huazhong University of Science and Technology, China
E-mail: ouyang_hb@yeah.net

Abstract

This paper proposes a set of measurement to analyze the information content of limit order book based on cointegration theory and error correction model, and conduct empirical studies using the tick-by-tick transaction data of Shanghai Stock Exchange 180 Index components. Empirical results indicate that limit-order book is informative, even the orders other than the best bid and ask can also contribute approximately 41% to the price discovery. Further, this paper also pays a lot of attention to examine the relationships between the limit-order book and the short-run future return of the stocks. We find that the information of a limit-order book can help investors to forecast the future short-term return and thus improve the overall welfare of them.

Keywords: limit order book, information content, price discovery, best bid and ask, short-term stock returns

1. Introduction

It is generally considered that limit order book offered by electronic automatic trading system can improve the information transparency of stock market greatly, which enables the investors to observe the best quotes and desired trading volume and can be executed instaneously. However, no consensus has been reached about whether limit order book is informative.

The information content of limit order book mainly refers to the information regarding stock price volatility, price discovery and stock returns caused by buy and sell order imbalance. Cohen, *et al.,* propose a gravitational pull model to explain how limit order book are affected by the order placement strategy of the traders and find that it is limit orders rather than market orders that determine the spreads [1]. Classifying the traders into pre-committed and value-motivated ones, Harris discusses a so-called "quote-matching strategy" and also draws a conclusion that limit order book is informative [2]. Handa proposes a static model and find that the probability that traders carried out transactions with informed traders determines whether they submit limit orders [3]. Similarly, Bloomfield, O'Hara and Saar find in electronic market that informed traders would submit more limit orders than market orders [4].

Harris and Panchapagesan, by studying the trading strategies of specialists in New York Stock Exchange (NYSE), confirm that limit order book contain much information about future price movements and thus the specialists may gain profits through the information,

which is more obvious for active stocks under fierce competition [5]. Cao, Hansch and Wang, who try to assess the information content of open limit order book of Australian Stock Exchange (ASX), examine the value of limit order book from two aspects: firstly, in comparison with the best quote and the depth, whether the investors can better estimate the true value of potential stocks with the information in limit order book other than the best quote; secondly, whether future short-term stock returns are related with current supply and demand schedule [6]. They find that the limit order book was moderately informative even after controlling return autocorrelation, internal spread and order imbalance, and the imbalance between the buy and sell sides is significantly correlated with future short-term stock returns. Hillman and Salmon use the high-frequency data of USD/DM rate to study the information content of the limit order book variorum techniques [7]. They show clear in-sample ability to explain very short-run movements in the rate using a range of measures of order book structure. Hellstrom and Simonsen also show that both the change and the imbalance of the order book statistically significantly explain future price changes and suggest that the information content of the order book is very short-term [8].

Some studies try to exploit similar issue by out-of-sample methods. Huang and Stoll (1994) find that differences in quoted depth predict future returns at five-minute intervals out-of-sample [9]. Froot and Ramadorai find that limit order book may contain some information for future exchange rate returns in low frequency data [10]. Rime, Sarno, and Sojli confirm these findings in three major exchange rates [11]. However, Danielsson, Luo, and Payne and Sager and Taylor find limit or no evidence of the predicting ability of order books at different forecast horizons [12-13]. Using a simple linear model, Kozhan and Salmon find that despite the in-sample statistical significance of variables describing the structure of the limit order book in explaining tick-by-tick returns, they do not consistently add significant economic value out-of- sample [14].

The studies on the limit order book of China's stock market are still limited. Chen studies the adverse selection cost of limit order book in China's stock market [15]. They described the statistical property of limit order book in China's stock market. However, these studies do not involve the analysis on the information content of limit order book. Kim, Yu and Yang employ the goodness-of-fit $R2$ of regression equation to examine whether the information in limit order book other than the best quote helps to predict future returns and find that the second and third quotes also contain the information related to short-term stock returns but it is unlikely to precisely predict future returns according to limit order book [16-17].

In this paper, according to the definition of information content by Hasbrouck [18], the authors try to analyze a set of measurements on the information content of limit orders based on an error correction model and finally explore the relation between limit order book and short-term stock returns with order imbalance and price impact taken into consideration. According to our findings, limit order book plays an important role in price discovery and the investors can predict future short-term returns through limit order book.

2. Hypotheses and Model

2.1. Basic Hypotheses

Kaniel and Liu suggest that informed traders prefer limit orders to market orders if the private information is long-lived [19]. On one hand, informed traders would bear high cost if market orders, though instantaneously executable, are submitted. On the other hand, the behavior to submit market orders may reveal a lack of patience and even more information. In contrast, some scholars suggest that informed traders will submit market orders which could

be quickly put into effect if the private information is short-lived. According the above literature, the first hypothesis is proposed as below:

Hypothesis 1: Limit order book is informative and therefore contributes to price discovery.

Limit order book, like providing a picture of the market supply and demand for the investors, indicates the willingness of the investors to buy and/or sell. If the demand overweighs the supply, namely the quantity of the buy side is higher; share price will go up, which shows that limit order book can signal future short-term price movements. According to the so-called "quote matching strategy" by Harris, traders can obtain option value through the side which is favorable in limit order book and traders who make matched quotes can gain profits through the relation between order imbalance and price dynamics [5]. Consequently, the second hypothesis is proposed as follows:

Hypothesis 2: Limit order book contains the information regarding future short-term price movements.

2.2. Research Methods

2.2.1. The Information Content of Limit Order Book

According to the definition of information content by Hasbrouck [18] and the measurement for stock price in formativeness proposed by Cao, Hansch and Wang [6], we define three measurement of prices related to limit order book as below:

① Average price of best bid and ask quote (MID[1]) refers to the average value of the best bid price and the best ask price in limit order book, namely

$$MID = (a_1 + b_1)/2 \tag{1}$$

Where a_1 and b_1 are the best bid and ask prices in limit order book respectively. The average price of best bid and ask quote is the most typical indicator to measure the information content of limit order book.

② Last transaction price (P) is brought in for the comparison of the contribution of different price indicators to the information content as it is the best expected stock value.

③ Weighted average price (WP) is introduced to integrate the price and quantity in limit order book other than the best quotes:

$$WP = \frac{\sum_{j=2}^{5} (b_j * bv_j + s_j * sv_j)}{\sum_{j=2}^{5} (bv_j + sv_j)} \tag{2}$$

Where b_j is the bid price and bv_j is the quantity that the buy side is willing to buy at the jth step in limit order book; s_j is the ask price and sv_j is the quantity that the sell side is willing to sell at the jth step in limit order book[2]. This indicator sums up the information content from the second step to fifth step in limit order book with a special focus on the marginal contribution of quotation other than the best bid and ask quote to the information content of limit order book.

To examine the contribution of these three prices to the information content, a vector $X_t = (MID, P, WP)$ is selected and an error correction model (ECM) is applied for comparative analysis, thus obtaining the contribution of these three prices to the information content of

---

[1] This variable has been used to measure the expectation of true stock value in many literatures.
[2] For simplicity, the first to fifth best quotes are ranked and called the first to fifth steps.

limit order book. The error correction model applied to describe the multiple price process is as follows:

$$\Delta X_t = \alpha z_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \varepsilon_t \qquad (3)$$

Where $z_{t-1} = \beta' X_{t-1}$ is the error correction term; k is the order of original value at risk (VAR) of X_t), namely the optimal lag period determined in accordance with Akaike information criterion (AIC) [6]. Hasbrouck defines the information content of price series j as [18]:

$$S_j = \frac{\varphi_j^2 \Omega_{jj}}{\varphi \Omega \varphi'} \qquad (4)$$

Where Ω is covariance matrix; φ is column vector of the moving average coefficients and φ_j is the jth element of φ, with standardization conducted to ensure the summation of all information content is 1. Given contemporaneous correlation, an innovation in prices (*i.e.,* the information content) is orthogonalized with triangular matrix F:

$$\varepsilon_t = Fz_t \qquad (5)$$

Where z_t is the unit covariance matrix with zero mean and F is the triangular matrix $\Omega=FF'$ after Cholesky decomposition of Ω, *i.e.,* $\Omega = FF'$. Here ε_t is also called innovative component in price process, which stands for the contribution of innovations to stock value.

2.2.2. The Relation between the Information Content of Limit Order Book and Short-Term Stock Returns

According to Hypothesis 2, limit order book helps to predict short-term stock price movements. However, sometimes stock price fluctuations are unpredictable or generally beyond predictions, which are called innovation component of short-term stock return. Is limit order book correlated with the innovation component of short-term stock return so that it helps to predict stock price fluctuations? To explore the problem, whether order imbalance can explain the innovation component of short-term stock return shall be studied.

In this paper, the short-term stock returns is represented by the average return of the best bid and ask prices in every five minutes. Therefore,

$$r_t = \frac{MID_t - MID_{t-1}}{MID_{T-1}} \qquad (6)$$

The frequency for time t is 5 minutes. Since short-term returns are serial correlated, pre-whiten shall be performed to eliminate serial correlation in advance. In the correction of serial correlation with the auto regression model, stepwise selection is performed to each stock and the lagged terms are automatically selected, thus making the coefficients of all independent variables significant and the intercept usually insignificant. As a result, the innovations of stock returns can be reflected with the auto regression model below:

$$r_t = \sum_{i=1}^{k} \alpha_i r_{t-i} + \varepsilon_t^{*3} \qquad (7)$$

---

[3] The residual term obtained after regression will be used as the dependent variable for regression examination and

Where k is determined through the stepwise method and ε_t^* is short-term return innovation.

We use the following variables to construct the model and examine the relationship between limit order book and short-term stock returns: trade imbalance (*Timb*), which measures the total net transactions of the buy and sell sides in five minutes; inside spread (*Spread*), which refers to the difference between the highest bid price and the lowest ask price; the length (quantity) imbalance of limit order book (*QR*), which measures the relative value of the imbalance between the quantities of the buy side and the sell side; the height (price) imbalance of limit order book (*HR*), which measures the relative value of the imbalance between the bid and ask quotes. All these variables are defined as follows:

$$Timb_t = (ts_t * bsratio_t - ts_{t-1} * bsratio_{t-1}) - ((ts_t - ts_{t-1}) - (ts_t * bsratio_t - ts_{t-1} * bsratio_{t-1}))$$

$$= 2 * (ts_t * bsratio_t - ts_{t-1} * bsratio_{t-1}) - (ts_t - ts_{t-1}) \tag{8}$$

Where ts_t is the accumulated trading volume from the opening time to time t that day; $bsratio_t$ is the bid competition ratio at the time t. Since the frequency is five minutes, $Timb_t$ is the summation of net balance of trading volume between the buy and sell sides every five minutes. As $Timb_t$ grows, the capacity of the buy side is improved and stock price goes up. Consequently, the short-term stock returns will be increased.

$$Spread = s_1 - b_1 \tag{9}$$

Where s_1 and b_1 are the best bid and ask quotes respectively, and is also called inside spread. In the quotation system of limit order book, the best ask price is higher than the best bid price, so *Spread* is a non-negative value. *Spread* and $Timb_t$ are used to indicate the top effects of limit order book that can be observed by the investors[4].

$$QR_j = \frac{sv_j - bv_j}{sv_j + bv_j} \tag{10}$$

Where $j=1\text{-}5$. QR_j is applied to measure the length imbalance of limit order book. When QR_j rises, the capacity of the sell side is strengthened and the stock price goes down. As a result, the short-term stock returns will be decline.

$$HR_j = \frac{(s_j - s_{j-1}) - (b_{j-1} - b_j)}{(s_j - s_{j-1}) + (b_{j-1} - b_j)} \tag{11}$$

Where $j=2\text{-}5$. HR_j is applied to measure the height imbalance of limit order book. When HR_j goes up, the growth of the bid price is smaller than that of the ask price. Along with the higher stock demand, the short-term stock returns will go up and the expected symbol are positive.

Above all, the theoretical model can be expressed as follows:

$$\varepsilon_t^* = f(\overset{+}{timb}, \overset{-}{spread}, \overset{-}{QR}, \overset{+}{HR}) \tag{12}$$

---

therefore estimation coefficients of dependent variables shall be significant when auto regression is performed to each stock.

[4] The order to buy and sell at market price can counteract corresponding limit order to buy or sell that rank on the top of limit order book.

Where ε_t^* an innovation of future short-term stock is returns (namely unpredictable short-term stock returns) and above explanatory variables are the expected symbols of their regression coefficients. The regression model is set as below:

$$\varepsilon_t^* = \alpha_0 + \delta_0 Timb_{t-1} + \beta_0 Spread_{t-1} + \gamma_1 QR_{1,t-1} + \sum_{j=2}^{n} \beta_j HR_{j,t-1} + \sum_{j=2}^{n} \gamma_j QR_{j,t-1} + \eta_t \quad (13)$$

Where null hypothesis of F-test H_0: $\beta_n = \gamma_n = 0$ and $n = 1\text{-}5$, *i.e.,* five regression equations are included in the set of regressions to examine the effects of the nth step of limit order book on future short-term stock returns.

Apart from net transaction (*Timb*) and inside spread (*Spread*), the length of the buy and sell sides (bv_j) and (sv_j) as well as the height of the buy and sell sides (hb_j) and (hs_j) are introduced in the model.

The variables above are calculated as follows: $hb_j = b_{j-1} - b_j$ and $hs_j = s_j - s_{j-1}$. As the length of the buy side rises, the capacity of the buy side is improved and the stock price goes up, thus increasing future short-term stock returns. However, as the length of the sell side rises, the capacity of the sell side is strengthened and the stock price goes down, thus reducing future short-term stock returns. On the contrary, as the height of the sell side rises, the stock demands grow and future short-term stock returns will be increased. As the height of the buy side rises, the stock supply grows and future short-term stock returns will be decreased.

Above all, the theoretical model can be expressed as follows:

$$\varepsilon_t^* = f(timb^+, spread^+, bv^-, hb^-, sv^-, hs^+) \quad (14)$$

The buy and sell sides are considered separately and the effects of the buy and sell sides on future short-term stock returns can be compared through the regression of independent variables. The regression model is as below:

$$\varepsilon_t^* = \alpha_0 + \delta_0 Timb_{t-1} + \beta_0 Spread_{t-1} + \sum_{j=1}^{n} \gamma_{d,j} bv_{j,t-1} +$$
$$\sum_{j=1}^{n} \gamma_{s,j} bs_{j,t-1} + \sum_{j=2}^{n} \beta_{d,j} hb_{j,t-1} + \sum_{j=2}^{n} \beta_{s,j} hs_{j,t-1} + \eta_t \quad (15)$$

Apart from the length and height, price impact is further introduced to describe the imbalance of limit order book and examine the impacts of price on future short-term stock returns. The price impact of the buy side ($LD(q)$) is applied to describe the discounts which can be obtained by the sell side for each stock when their quote are lower than the middle value of the best bid and ask prices (*MID*), which also reflects the liquidity of the buy side. The price impact of the sell side ($LS(q)$) is used to describe the premium which should be paid by the buy side for each stock when their quotes are higher than the middle value of the best bid and ask prices(*MID*), which stands for the liquidity of the sell side. Through the comparison of $LD(q)$ and $LS(q)$, the imbalance between the buy and sell sides can be obtained and the effects of limit order book on future short-term stock returns can be analyzed. They are calculated by the following equation:

$$LD(q) = 0.5 * (s_1 + b_1) - \frac{\sum_{j=1}^{m-1_1} b_j bv_j + b_{m_1} q_{m_1}}{q} \quad (16)$$

Where q is the multiple of average trade size ($\overline{Q}$) of each stock; m_1 is the steps it takes for the buy side to reach the accumulated trading volume q from the first step; q_{m_1} is the trading volume submitted by the buy side at the m_1 step to reach the accumulated trading volume q. q is calculated as follows:

$$\sum_{i=1}^{m_1-1} \mathrm{bv}_i < q \leq \sum_{i=1}^{m_l} bv_i \, , \, q_{m_l} = q - \sum_{i=1}^{m_1} bv_i \tag{17}$$

Therefore, $LD(q)$ is equal to the difference between the average values of the best bid and asks prices and the average weighted price when the buy side tries to realize the accumulated trading volume q. When $LS(q)$ decreases, the demand side becomes more liquid and the future short-term stock returns grow.

Similarly,

$$LS(q) = \frac{\sum_{j=1}^{m_2-1} s_j sv_j + s_{m_2} q_{m_2}}{q} - 0.5 * (s_1 + b_1) \tag{18}$$

Where $LS(q)$ is the difference between weighted price when the sell side tries to realize accumulated trading volume q and the mean value of the best bid and ask prices. When $LS(q)$ decreases, the supply side becomes more liquid and the future short-term stock returns falls.

Above all, the theoretical model can be expressed as follows:

$$\varepsilon_t^* = f(timb^+, spread^-, LD(q), LS^+(q)) \tag{19}$$

The regression model is shown as below:

$$\varepsilon_t^* = \alpha_0 + \delta_0 Timb_{t-1} + \beta_0 Spread_{t-1} + \sum_{j=2}^{g} \beta_{d,j} LD(g\overline{Q}/2)_{t-1}$$

$$+ \sum_{j=2}^{g} \beta_{s,j} LS(g\overline{Q}/2)_{t-1} + \eta_t \tag{20}$$

We further introduce the aggregate imbalance of price impacts ($LR(q)$) to measure the imbalance of price impacts at different trading volumes between the buy and sell sides, which is expressed as follows:

$$LR(q) = \frac{LS(q) - LD(q)}{LS(q) + LD(q)} \tag{21}$$

Where $q = g*(\overline{Q}/2)$, $g = 2\text{-}5$. When the demand side has better liquidity than the supply side, the stock returns will go up.

Above all, the theoretical model can be expressed as follows:

$$\varepsilon_t^* = f(timb^+, spread^+, LR(q)) \tag{22}$$

The empirical model is shown as below:

$$\varepsilon_t^* = \alpha_0 + \delta_0 Timb_{t-1} + \beta_0 Spread_{t-1} + \sum_{j=2}^{g} \beta_j LR\, (g\,\overline{Q}/2)_{t-1} + \eta_t \qquad (23)$$

3. Data and Descriptive Statistics

3.1. Sample Selection and Processing

In empirical study, we take the tick-by-tick high frequency data of the components of Shanghai Securities Exchange 180 Index (SSE 180) from September 2009 to December 2010 as the sample. Raw data are processed according to the following rules:

(1) Daily trading volume=Monthly total trading volume/Accumulated stock exchange days of that month.

(2) Average daily transaction number=Monthly total transaction number/Accumulated stock exchange days of that month.

(3) Average trade size ($\overline{Q}$) =Daily trading volume/Daily transaction number or Monthly trading volume/Monthly transaction number. Both calculation methods are the same and average trade size is used to indicate trade activity and capital flow strength.

(4) The height of the buy side in $hb_j=b_{j-1}-b_j$. hb_1 can be calculated as the difference between the bid price at the first step and the average of best bid and ask prices, namely, $MID-b_1$, which is used to measure the imbalance of the buy side. Similarly, the height of the sell side is $hs_j=s_j-s_{j-1}$.

(5) Relative height of the buy side at the ith step is the height at i^{th} step divided by the difference between the middle value of best bid and ask prices and the bid price at the fifth step (namely the sum of the height of the buy side at five steps) and the equation is $RBheight_i=(b_i-b_{i-1})/(b_5-MID)$. Therefore the percentage of the height at each step can be obtained. The relative height of the sell side $RSheight_i$ can be obtained similarly.

Relative length of the buy side is the length at the ith step divided by the sum of the length of the buy side at five steps and the equation is $RBlength_i=bv_i/\sum_{i=1}^{5} bv_i$. Therefore the percentage of the length at each step can be obtained. Similarly, the relative length of the sell side $RSlength_i$ can be calculated.

(6) In our regression analysis, two kinds of samples are considered. One is the full sample and the other is the restricted sample with the top and bottom 5% observations sorted by $LR(2.0\overline{Q})$.

3.2. Descriptive Statistics

According to the statistical results in Table 1, daily trade volume of 180 constituent stocks covered in SSE 180 Index is 302, 679 shares. Although they are individual shares of large scale and good liquidity, there still exists great difference among stocks. Average trade size of those 180 stocks is 1422 shares; average daily transaction number is 20302; average internal spread is 1.76 cents and average relative spread is 0.124%.

Table 1. Basic Descriptive Statistics of Samples

	Daily Trade Volume (100 shares)	Average Trade Size (shares)	Daily Transaction Number	Inside Spread (cents)	Relative Spread (%)
Average	3026.79	1422	20302	1.76	0.124
Standard Deviation	3197.96	546	17572	1.68	0.044
Minimum	147.61	375	1818	1.01	0.050
Maximum	18765.41	4107	133436	20.40	0.258

Limit order book is quite informative. Since market liquidity is completely provided by limit order book, the bid and asks sides represent the market demand and supply respectively. The length and height as well as relative length and height of the buy and sell sides can be used to measure the imbalance of limit order book, which may contribute to price discovery according to Hypothesis 1.

According to Table 2, relative length of the buy and sell sides increases monotonously from the first to fifth step but the growth rate is limited. The relative length of the buy side rises from 15.55% at the first step to 22.06% at the fifth step, while the relative length of the sell side rises from 16.44% at the first step to 22.19% at the fifth step. The trade volume of both the buy and sell sides at the first step is over 83%, which indicates the orders other than the first step are meaningful. However, the relative height of the buy and sell sides is maximized at the second step and decreases monotonously from the second to fifth step with slight decrease rate which means a dense limit order book and better market liquidity. From data processing method, we can find the height of the buy and sell sides are the same which is a half of internal spread in Table 2 and is maximized at the second step, whose changes are similar to those of relative height.

Table 2. Basic Shape of Limit Order Book

Step	Relative Length (%)		Relative Height (%)		Absolute Height (cent)	
	Buy Side	Sell Side	Buy Side	Sell Side	Buy Side	Sell Side
1	15.55	16.44	13.78	13.50	0.88	0.88
2	19.84	19.50	22.46	22.69	1.33	1.41
3	20.98	20.59	21.14	21.52	1.22	1.27
4	21.56	21.28	21.20	21.20	1.20	1.24
5	22.06	22.19	21.12	21.10	1.24	1.22

4. Empirical Results and Analysis

4.1. The Information Content of Limit Order Book

According to the error correction model in equation (3), a set of cross section results can be obtained after regression for each stock, as shown in Table 3.

Table 3. Estimations of the Information Content of Limit Order Book

Variables	Observation	Mean (%)	Std Dev.	Min (%)	Max (%)
WP	180	40.58	4.97	27.97	64.42
MID	180	55.53	5.35	32.87	69.17
P	180	3.89	1.43	0.05	7.36

Table 3 shows that the cross section average information content of *WP* is 40.58%, slightly lower than the cross section average information content of *MID* of 55.53%. The last transaction price (*P*) contains the least information content, with the cross section average information content of 3.89%, and is greatly lower than the former two. Therefore, it can be found that limit order book is informative, with the contribution to price discovery of over 40%.

In most studies, the average value of the best bid and ask prices is employed to measure stock value, which has been confirmed by the results of Table 3 to some extent. However, the information other than the best bid and asks prices have also played an inevitable role in price discovery, which is definitely of great importance for investment practice.

4.2. The Relationship between Limit Order Book and Short-Term Equity Returns

4.2.1. Future Short-Term Stock Returns with Aggregate Buy and Sell Imbalance

To examine whether the information in limit order book other than the best bid and ask prices helps to predict future short-term stock returns, we can observe whether adjusted R^2 increases with the growth of steps incorporated in regression equation. The changes of adjusted R^2 reflect the explanatory ability of additional information of limit order book to the dependent variables. In full sample analysis, average adjusted R^2 is 10.71% when there are only two dependent variables of *Timb* and *Spread* and rises to 10.97% when QR_1 is introduced and 12.4% when QR_2 and HR_2 are taken into consideration. Finally average adjusted R^2 reaches 16.72% at the fifth step. Since many complicated factors are involved in asset pricing model, the fitting degree is not good enough but acceptable, which exactly reveals the explanatory ability of limit order book to significantly improve asset pricing model. When 5% of the top and bottom observations sorted by price impacts are taken for regression analysis, the results are similar.

Table 4. Regression Analysis with Aggregate Buys and Sell Imbalance

Step	Full Sample			Restricted Sample		
	Adjusted R^2 (%)	Percentage of stocks that pass F-test (%)	Percentage of stocks consistent with expected signal (%)	Adjusted R^2 (%)	Percentage of stocks that pass F-test (%)	Percentage of stocks consistent with expected signal (%)
0	10.71	-	-	11.18	-	-
1	10.97	69.7	91.2	12.29	20.6	63.5
2	12.40	62.4	75.3	13.35	19.6	55.4
3	13.45	55.2	66.1	14.92	16.1	36.2
4	15.86	48.0	59.2	16.36	16.0	33.8
5	16.72	35.3	45.6	18.84	14.4	30.9

According to Table 4, in full sample analysis, when the information of the first step of limit order book is utilized, the result of regression analysis shows that 91.2% of stocks of length imbalance are consistent with expected signals. As is mentioned above, when the supply side (namely the sell side) provides larger trading volume and QR_j increases, the future short-term stock returns decreases. When the information of the second step is utilized, the result of regression analysis shows that 75.3% of stocks of length imbalance and height imbalance are consistent with expected signals, which indicates future short-term stock returns further decreases when QR_j continues to increase. In addition, future short-term stock

returns grows when the growth rate of the bid price is larger than that of the ask price, *i.e.,* HR_j increases. When the first three steps of limit order book are introduced into the model, stocks that pass F-test at 5% level account for over 60%, which indicates that the a majority of constituent stocks of SSE 180 Index reject the joint null hypothesis. With the fourth and fifth steps introduced into the model, the stocks with expected signal decreases monotonously and the stocks that pass F-test gradually descends to lower than 40%. A basic conclusion can be drawn that the quoted prices that rank on the top of limit order book helps to predict future short-term stock returns.

4.2.2. Future Short-Term Stock Returns with Separate Buy and Sell Imbalance

Table 5 provides empirical results when buy and sell sides' imbalance are considered separately in the regression.

Table 5. Regression Analysis with Separate Buys and Sell Imbalance

Step	Full Sample			Restricted Sample		
	Adjusted R^2 (%)	Percentage of stocks that pass F-test (%)	Percentage of stocks consistent with expected signal (%)	Adjusted R^2 (%)	Percentage of stocks that pass F-test (%)	Percentage of stocks consistent with expected signal (%)
0	10.35			13.12		
1	11.15	51.1	71.4	14.52	20.8	47.7
2	11.37	48.2	59.6	15.37	26.5	38.0
3	12.65	42.8	42.0	16.00	18.4	30.1
4	14.83	29.1	39.2	17.05	15.5	25.8
5	15.77	15.7	25.5	18.20	7.3	7.9

On the whole, the results in Table 5 are basically similar as those of Table 4. In Table 5, the buy and sell sides are considered separately and thus four signals are examined to check if the length and height coefficients are consistent with expectations. As a result, the percentage of stocks with expected signals in Table 5 is lower than that in Table 4, but it shows the same variation trend, that is to say, the larger the percentage of stocks with expected signals is, the better the prediction of future short-term stock returns can be made. The percentage of stocks that pass F-test in Table 5 is basically the same as that in Table 4. Since the buy and sell side are considered separately in Table 5 and there are more independent variables in regression model in Table 5, adjusted R^2 in Table 5 is larger but the difference is not so obvious.

4.2.3. Future Short-Term Stock Returns with Book Imbalance and Separate Buy and Sell Price Impacts

The regression model relied on by Table 6 reveals the effects of lagged price effect on future short-term stock returns. In full sample analysis, average adjusted R^2 is 15.25% when there are only two dependent variables of *Timb* and *Spread* and reaches a maximum of 20.85% when price effects of the buy and sell sides $LR(q)$ and $LR(s)$ are taken into consideration separately in different trade sizes. While in restricted sample analysis, average adjusted R^2 is 16.80% when there are only two dependent variables of *Timb* and *Spread* and

reaches 21.54% when the trade size is $3.0^{\bar{Q}}$. Different from full sample analysis, average adjusted R^2 doesn't show monotonous changes.

The statistical results of stocks that pass F-test and stocks with expected signals are similar to Table 4 and 5. Stocks with expected signals account for a maximum of 79.2% in full sample analysis while the maximum and minimum of stocks with expected signals are 77.2% and 56% respectively in restricted sample analysis, which indicates the returns of most stocks increases when the demand side has better liquidity. Therefore, lagged price effect of the buy and sell sides are helpful to predict future short-term stock returns.

Table 6. Regression Analysis with Book Imbalance and Separate Buy and Sell Price Impacts[5]

Trade Size	Full Sample			Restricted Sample		
	Adjusted R^2 (%)	Percentage of stocks that pass F-test (%)	Percentage of stocks consistent with expected signal (%)	Adjusted R^2 (%)	Percentage of stocks that pass F-test (%)	Percentage of stocks consistent with expected signal (%)
0	15.25			16.80		
1.0 $\bar{Q}$	16.46	71.7	79.2	18.09	58.8	74.4
1.5 $\bar{Q}$	17.89	59.1	71.5	17.97	35.0	77.2
2.0 $\bar{Q}$	19.17	42.0	68.2	19.86	13.3	61.6
2.5 $\bar{Q}$	20.21	24.1	48.9	20.81	8.6	56.4
3.0 $\bar{Q}$	20.85	8.6	46.5	21.54	5.5	55.9

4.2.4. Future Short-Term Stock Returns with Book Imbalance and Aggregate Price Impacts

When price effect is introduced and order imbalance of the buy and sell sides is taken into consideration, the effects of limit order book on future short-term stock returns are as shown in Table 7. In full sample analysis, average adjusted R^2 is 15.25% when there are only two dependent variables of *Timb* and *Spread* and reaches 20.63% when price effects of the buy and sell sides $LR(q)$ are considered comprehensively in different trade sizes. In restricted sample analysis, the maximum of average adjusted R^2 is 20.62%. It can also be found in Table 7 that the stocks that pass F-test account for over 50% with given changes of trade size, which indicates over a half of the stocks reject null hypothesis.

---

[5] When trade size is one to three times of $\bar{Q}$, because there are 387 observations whose total trade volumes for five steps fail to reach 3.0 $\bar{Q}$, namely $\sum_{i=1}^{5} bv_i < q$, and m_1 decided by $\sum_{i=1}^{m_1-1} bv_i < q \leq \sum_{i=1}^{m_1} bv_i$ has no solution. If the trade size is further increased, m_1 of more observations have no solutions. It is the same to the sell side. When the trade size is 3.0 $\bar{Q}$, 311 observations whose total trade volumes cannot reach 3.0 $\bar{Q}$, namely $\sum_{i=1}^{5} sv_i < q$, and m_2 decided by $\sum_{i=1}^{m_2-1} sv_i < q \leq \sum_{i=1}^{m_2} sv_i$ has no solution either. Consequently, m_2 of more observations have no solutions with the growth of trade size.

Table 7. Regression Analysis with Book Imbalance and Aggregate Price Impacts

Trade Size	Full Sample			Restricted Sample		
	Adjusted R^2 (%)	Percentage of stocks that pass F-test (%)	Percentage of stocks consistent with expected signal (%)	Adjusted R^2 (%)	Percentage of stocks that pass F-test (%)	Percentage of stocks consistent with expected signal (%)
0	15.25	-	-	16.81	-	-
1.0 $\bar{Q}$	16.30	70.9	88.8	17.03	57.1	47.9
1.5 $\bar{Q}$	16.94	79.4	89.9	17.14	59.0	39.4
2.0 $\bar{Q}$	17.45	72.2	74.1	17.31	46.2	32.2
2.5 $\bar{Q}$	19.53	61.3	63.3	19.16	45.7	25.6
3.0 $\bar{Q}$	20.63	55.7	70.3	20.62	32.2	17.8

As is shown in Table 7, the percentage of stocks with expected signal reaches nearly 90% when the trade size $q=1.5\bar{Q}$ in full sample analysis and the minimum is over 60% with given changes of trade size, which indicates that the stock returns expands when the demand overweighs the supply and the buy side has better liquidity than the sell side.

Eventually, the conclusion that limit order book does offer the information regarding future short-term stock returns can be drawn, which is consistent with Hypothesis 2.

5. Conclusion

It is considered in some literature that the research on the information content of limit order book is of great theoretical and practical significance, which helps the investors to obtain the true value of stocks and predict future short-term stock price movements to some extent. To this end, an empirical study on the information content of limit order book for the constituent stocks of Shanghai Securities Exchange 180 Indexes is carried out. In accordance with two basic hypotheses, the analysis falls into two aspects: firstly, whether the investors obtain the true stock value based on the information in limit order book other than the best bid and ask quotes, namely the role of limit order book in price discovery; secondly, whether the real-time information of the buy and sell sides is related to future short-term stock returns.

The empirical findings show that the information in limit order book other than the best bid and ask prices can offer true stock value, with the contribution to price discovery of nearly 41%. Meanwhile, after repeated regressions, it can be confirmed that lagged information other than the best bid and ask prices is quite helpful to predict future short-term stock returns.

This study can provide implications for the investment practice. For the investors of stock market, limit order book can actually provide real-time will of the supply and demand sides, which is an important clue to determine the true stock value. Study on the specific pattern of limit order book helps to gain short-term investment returns as well.

Acknowledgements

References

[1] K. Cohen, S. Maier and R. Schwartz, J. Polit. Econ., vol. 89, no. 2, **(1981)**, pp. 287-305.
[2] L. Harris, "Liquidity, trading rules, and electronic trading systems", Monograph series in finance and economics, New York University, **(1990)**.
[3] P. Handa and R. A. Schwartz, J. Financ., vol. 51, **(1996),** pp. 1835-1861.
[4] R. Bloomfield, M. O'Hara and G. Saar, J. Financ. Econ., vol. 75, **(2005),** pp. 165-199.
[5] L. Harris and V. Panchapagesan, J. Financ. Mark, vol. 8, **(2005)**, pp. 25-67.
[6] C. Cao, O. Hansch and X. Wang, J. Futures. Markets, vol. 29, no. 1, **(2009),** pp. 16-41.
[7] R. Hillman, M. Salmon, J. Financ. Fore, vol. 1, **(2007)**, pp. 3-43.
[8] J. Hellstrom and O. Simonsen, "Does the open limit order book reveal information about short-run stock price movements"? Unpublished manuscript, **(2009)**.
[9] R. Huang and H. Stoll, Rev. Financ. Stud., vol. 7, **(1994),** pp. 179-213.
[10] K. Froot, T. Ramadorai, J. Financ., vol. 60, **(2005)**, pp. 1535-1566,
[11] D. Rime, L. Sarno and E. Sojli, J. Int. Econ., vol. 80, **(2010)**, pp. 72-88.
[12] J. Danielsson, J. Luo and R. Payne, "Exchange rate determination and inter-market order flow effects", Unpublished manuscript, **(2002)**.
[13] M. Sager and M. Taylor, J. Money. Credit. Bank, vol. 40, **(2008)**, pp. 583-625.
[14] R. Kozhan and M. Salmon, J. Financ. Mark, vol. 15, **(2012)**, pp. 1-28.
[15] Y. Chen, "The Research on the volume price relation and adverse selection cost of limit order book of China's stock market", Master Thesis of University of Electronic Science and Technology of China, **(2008)**.
[16] S. Kim and P. Sohn, Int. J. Multimedia Ubiquitous Eng., vol. 8, no. 2, **(2013)**, pp. 197-204.
[17] Y. Yu and Y. Yang, Stat. & Deci., vol. 7, **(2006),** pp. 19-21.
[18] J. Hasbrouck, Financ., vol. 50, **(1995)**, pp. 1175-1199.
[19] R. Kaniel and H. Liu, J. Bus, vol. 79, **(2006)**, pp. 1867-1913.

Authors

Yaqing Liu, is with the School of Economics and Management, North China University of Technology, Shijingshan, Beijing, China 100144 (email: liu_yaqing@yeah.net). Yaqing Liu got Ph.D in economics in School of Economics, Huazhong University of Science and Technology, and used to be post-doctoral fellowship of School of Economics and Management, Tsinghua University (2009-2012).

Hongbing Ouyang, is with the School of Economics, Huazhong University of Science and Technology, Luo Yu Road 1037#, Wuhan, China 430074 (email: ouyang_hb@yeah.net). Hongbing Ouyang got Ph.D in finance in City University of Hong Kong, and was a visiting professor of Sogang University, Korea, in 2005.