

---

# THE INFORMATION CONTENT OF AN OPEN LIMIT-ORDER BOOK

---

**CHARLES CAO\***  
**OLIVER HANSCH**  
**XIAOXIN WANG**

Using data from the Australian Stock Exchange, the authors assess the information content of an open limit-order book with a particular focus on the incremental information contained in the limit orders behind the best bid and offer. The authors find that the order book is moderately informative—its contribution to price discovery is approximately 22%. The remaining 78% is from the best bid and offer prices on the book and the last transaction price. Furthermore, the authors find that order imbalances between the demand and supply schedules along the book are significantly related to future short-term returns, even after controlling for the autocorrelations in return, the inside spread, and the trade imbalance.  
© 2008 Wiley Periodicals, Inc. *Jrl Fut Mark* 29:16–41, 2009

The authors thank Geoffrey Booth, Tarun Chordia, Ian Domowitz, Joel Hasbrouck, Kenneth Kavajecz, Bruce Lehmann, Mark Lipson, Chris Muscarella, Gideon Saar, Chester Spatt, Avanidhar Subrahmanyam, Robert Webb (the Editor), an anonymous referee, and seminar participants at Penn State, Rutgers, George Washington Universities, the Western Finance Association Meetings, the NBER-JFM Microstructure Conference, the American Finance Association Meetings, and the 15th Annual Asia Pacific Futures Research Symposium for their helpful comments. The authors are grateful to the Australian Stock Exchange and the Securities Industry Research Centre Asia-Pacific for providing the data used in this study. The authors also thank Morgan Stanley for providing the equity market microstructure research grant for this study.

\*Correspondence author, Smeal College of Business Administration, Pennsylvania State University, 609 BAB, University Park, PA 16802. Tel: +1-814-865-7891, Fax: +1-814 865 3362, e-mail: charles@loki.smeal.psu.edu  
*Received November 2006; Accepted February 2008*

- 
- *Charles Cao is a Smeal Chair Professor of Finance and Oliver Hansch is Assistant Professor of Finance at Smeal College of Business Administration, Pennsylvania State University, University Park, Pennsylvania.*
  - *Xiaoxin Wang is Assistant Professor of Finance, College of Business Administration, Southern Illinois University, Carbondale, Illinois.*

## INTRODUCTION

A majority of equity and derivative markets around the world are organized as electronic limit-order books. Such equity markets include the Electronic Communication Networks (ECNs) in the United States, the Toronto Stock Exchange, and the Hong Kong Stock Exchange. Electronic trading platforms in derivative markets have also gained popularity in recent years over the traditional open-outcry auctions. Chicago Mercantile Exchange's (CME) Globex platform, International Securities Exchange's electronic option trading platform, and the single centralized limit-order book offered by Euronext.liffe are all good examples. Electronic limit-order book has stepped up to the center stage of the change in financial market structure. The London International Financial Futures and Options Exchange Connect System has been adopted by several derivative markets [e.g., the Chicago Board of Trade (CBOT), etc.]. Other markets, including the New York Mercantile Exchange, have migrated their trading to the CME Globex electronic platform. Many derivative markets have fully converted to an electronic trading system (e.g., the Sydney Futures Exchange, the International Petroleum Exchange of London, and the Hong Kong Futures Exchange).

With the growing popularity of electronic limit-order books in equity and derivative markets, studying the trade prices and best quotes is no longer enough for investors. Some derivative markets have released real-time limit-order books to investors for a monthly fee. Today, investors can view the top five levels of depth for contracts listed on the CME (e.g., the popular E-Minis futures) through the CME Globex Level II data feed, whereas the CBOT Level II data feed allows traders to view the top ten levels of depth for contracts listed on the CBOT (e.g., the Mini-Sized Dow Jones Industrial Average Futures).<sup>1</sup>

Among the reasons put forth for the popularity of an open limit-order book is the greater transparency offered by these systems when compared with dealer market settings. Although dealer markets usually rely upon dissemination of only the dealers' best quotes, a limit-order-book system allows its users to view the depth at a number of price levels away from the market. These displayed prices and quantities are instantaneously executable, typically. In dealer markets, however, prices for trades beyond the quoted size must be assessed through negotiations with market makers. In addition, market makers may choose to offer price improvement for trades up to a quoted size. In this study, the authors study the information content of an open book, and examine whether pre-trade transparency facilitates price discovery and helps investors in predicting short-term price movements.

<sup>1</sup>This difference in reporting depth for stock index futures contracts between CME and CBOT still exists despite the merger of the two exchanges in 2007.

Although the information content of a limit-order book has been the subject of an active debate in the price discovery literature, there has been no consensus about whether or not the order book is informative. On the one hand, Glosten (1994), Rock (1996), and Seppi (1997) incorporated informed traders into their models, assuming that they favor and actively submit market orders. This suggests that the order book beyond the best bid and offer contains little, if any, information. On the other hand, several recent studies suggest that the order book is informative. Using an experimental design, Bloomfield, O'Hara, and Saar (2005) found that in an electronic market, informed traders submit more limit orders than market orders. Using SuperDot limit orders in the TORQ database, Harris and Panchapagesan (2005) showed that the order book is informative, and that New York Stock Exchange (NYSE) specialists use the book information in ways that favor them over the limit-order traders. Kaniel and Liu (2006) showed that informed traders prefer limit orders, and that limit orders convey more information than market orders when the private information is long-lived.

Regarding the question of whether pre-trade transparency facilitates price discovery, the extant literature offers mixed evidence. Baruch (2005) provided a theoretical model showing that an open limit-order book improves liquidity and information efficiency of prices. Consistent with this prediction, Boehmer, Saar, and Yu (2005) found that the deviations of transaction prices from the efficient prices became smaller after the NYSE's adoption of the open book system. In contrast, Madhavan, Potter, and Weaver (2005) found larger spreads and higher volatility after the Toronto Stock Exchange disseminated the top four price levels of the limit-order book in April 1990.

This study is part of the emerging literature on the information content of the limit-order book. The authors are particularly interested in the incremental information content of the book, over and above the information traditionally available in a dealer market: i.e., the best bid and offer prices along with their respective depths. Using order-book information from the Australian Stock Exchange (ASX), the authors empirically assess the information content of an open limit-order book from two perspectives.

First, the authors explore whether the limit-order book allows better estimations of a security's value than simply the best bid and offer. If it does, how much additional information can be gleaned from the book? By investigating the relation among the midpoint of the best bid and ask, the transaction price, and a quantity-weighted price based on limit orders, the authors find that the order book beyond the first step is modestly informative and that price discovery measures suggest that the contribution of the order book beyond the best bid and offer is approximately 22%. Next, the authors examine whether the order-book information is associated with future returns. In a repetitive regression

framework, the authors show that the imbalance in the limit buy and sell orders beyond the inside market is useful and has additional power in explaining future short-term returns.

This study is organized as follows: the second section develops testable hypotheses. The third section describes the institutional environment of the ASX and the sample data. The information content of the order book is assessed in the fourth section, and the fifth section provides evidence of the association between returns and lagged imbalances between demand and supply. Concluding remarks are provided in the sixth section.

## TESTABLE HYPOTHESES

Price discovery is one of the principal functions of a financial market. The authors' first hypothesis pertains to the price discovery function in a limit-order book. Private information is incorporated into prices through trading by informed traders, who can choose to employ market or limit orders in their dynamic strategies. If informed traders use limit orders, especially limit orders away from the inside market, their information is presumably reflected in the book. If, however, informed traders use market orders, the orders in the book may not contain any of their private information.

Models of limit-order books featuring privately informed traders typically assume that they will favor market orders (Glosten, 1994; Rock, 1996; Seppi, 1997). Rock (1996) argued that, with short-lived private information, informed traders will prefer market orders because they guarantee immediate execution. Furthermore, given the direction of price movements conditional on the private information, the execution of limit orders designed to exploit a trader's informational advantage is unlikely. For example, an informed trader who knows that the current market price is too high will expect the price to go downward in the future, especially when the other traders learn the same information. Thus, the likelihood of achieving execution for a limit sell order is relatively small in this situation. Similarly, Angel (1997) and Harris (1998) argued that informed traders are more likely to use market orders than are liquidity traders.

In contrast, Kaniel and Liu (2006) considered the case where private information is long-lived and the number of traders who may discover the private information is small. They showed that informed traders may well prefer limit over market orders, such that, in equilibrium, limit orders may convey more information than market orders. One reason why informed traders are reluctant to submit market orders is that submitting market orders signals impatience and reveals too much information. Although their execution is certain and immediate, market orders incur higher trading costs than limit orders. The two competing strands of models lead to the authors' first hypothesis:

*Hypothesis 1:* The Orders Behind the Best Bid and Ask Prices Contribute to Price Discovery.

Evidence consistent with this hypothesis supports the prediction that informed traders use limit orders as part of their trading strategies. On the contrary, if informed traders prefer to use market orders over limit orders, it will be expected that their market orders may pick off any mis-priced limit orders until the book reflects all available information.

The authors second hypothesis is concerned with the implications of the order-book shape for short-term price dynamics, even in the absence of asymmetric information. The shape of the order book (i.e., the number of shares on each price step and how far away price steps are from each other) gives investors a concurrent picture of the market demand and supply. Specifically, the asymmetries between the buy and sell side of the book indicate shifts in the supply and demand curves caused by unobservable exogenous factors that affect the price of a stock. Through observing the demand and supply, investors have a better chance of guessing what these factors are and of predicting the future price movements.

Harris (1990) considered two types of limit-order traders: pre-committed and value-motivated traders. The former submit limit orders to reduce trading costs, but will switch to using market orders if their orders remain unfilled for too long. The latter express their valuations of the asset through their choice of limit price. Both types of limit orders can convey information about future price movements. A book imbalance caused by pre-committed traders may signal future price movements owing to these traders having to convert their unfilled limit orders into market orders. For example, a heavier buy than sell side would indicate a future price increase. An imbalance owing to the presence of value-motivated traders will reveal their valuations which will then become impounded in prices. Harris (1990) also discusses a so-called “quote-matching” strategy in which traders extract options values from the standing limit orders by trading ahead of the heavier side of the book. The presence of quote-matchers may thus create a link between asymmetries in the shape of the two sides of the book and price dynamics.

Empirically, Huang and Stoll (1994) found that differences in quoted depth between the bid and ask side do predict future price changes, especially for price changes over short intervals. Recently, Chordia, Roll, and Subrahmanyam (2002), Chordia and Subrahmanyam (2004), and Boehmer and Wu (2006) documented the relation between order imbalance and future price movements. The “order imbalance” in these studies refers to the difference between quantity bought and sold. The Lee and Ready algorithm is usually adopted to learn if a trade is buyer or seller initiated. As the authors study a

limit-order-book market with clearly stated buy and sell orders on each side of the book, the ambiguity in classifying trade directions is avoided. The authors broaden the scope of the investigation by including information on orders beyond the best bid and offer. Kavajecz and Odders-White (2004) demonstrated that support and resistance levels coincide with peaks in depth on the limit-order book and that moving average forecasts reveal information about the relative position of depth on the book. These issues lead to the following hypothesis:

*Hypothesis 2:* Limit orders behind the best bid and ask prices contain information about short-term future price movements.

A rejection of the hypothesis indicates that the shape and make-up of the limit-order book is not informative about short-term price movements beyond those conveyed by the inside spread and depth.

## DATA

The data used in this study are provided by the ASX via the Securities Industry Research Centre Asia-Pacific. The ASX uses the fully computerized Stock Exchange Automated Trading System (SEATS), which is based on the Toronto Stock Exchange's Computer Assisted Trading System.<sup>2</sup> There are no dealers or other designated liquidity providers on the ASX so that public orders can interact with each other directly. The ASX operates an open limit-order book, and the order book is widely disseminated. Depending on the chosen level of detail, SEATS Trading Screen users can view details of individual buy and sell orders along the book or the aggregate depth at multiple price levels in real time, generally for at least the first ten steps. For example, E\*trade-Australia and TDwaterhouse-Australia provide their clients with the book truncated at the tenth best step, and National Online Trading provides the book truncated at the 20th step. More details of the order-book information are available to brokers and institutional investors, whereas the aggregate version is representative of what online traders would be able to see. The openness of the book allows the authors to examine whether the book information is valuable to investors.

Each day, the market goes through several stages. During the pre-opening period from 7:00 A.M. to 10:00 A.M. (Eastern Standard Time), orders can be entered into the system but no matching takes place. The ASX opens at 10:00 A.M. with a procedure aimed principally at maximizing traded volume at the chosen

<sup>2</sup>In 2006, ASX changed its name to Australian Securities Exchange after its merge with Sydney Futures Exchange and SEATS was updated to the Integrated Trading System (ITS) workstation that can provide more transactions per second; yet the feature of a pure open limit-order book remains.

opening price. Stocks open sequentially in five groups, based on the alphabetical order of the ticker symbol, and normal trading begins immediately after the conclusion of the opening algorithm for their group. This phase, during which the vast majority of trading takes place, lasts until 4:00 P.M. Orders entered during normal trading hours are matched, resulting in trades, or they are stored in the order book automatically. At 4:00 P.M., a five-minute period of “pre-close” begins which is followed by the official single-price closing auction.

The authors obtain data covering the 100 most actively traded ASX stocks for the month of March 2000 from the ASX intraday data set that provides historical details of all individual orders placed on SEATS as well as any of its resulting trades. Each order and trade record includes information on the price, size, and direction, time-stamped to the nearest one-hundredth of a second. Crucially, the data allow for the reconstruction of the limit-order book for each stock at any point during the sample period. The authors’ procedure for building the book is similar to the method described by Bessembinder and Venkataraman (2004), who reconstruct the limit-order book on the Paris Bourse. To avoid confounding effects from the opening and pre-close procedures, the authors restrict their attention to the period from 10:15 A.M. to 4:00 P.M. Because most investors at ASX can access the price and aggregate number of shares up to the tenth step of the book, the authors construct the limit-order book up to ten price steps away from the inside of the market.

Table I presents summary statistics for the 100 sample stocks. Reported are cross-sectional distributions of daily trading activity. The average daily trading volume is 1.783 million shares, and the average number of trades is 349 per day, with an average trade size of 5,279 shares. The average dollar spread is 2.82 cents and the relative spread is 0.28%. There is some cross-sectional difference among the stocks. The most active stock has an average 1,323 transactions per day, whereas the least active has only 63, and the daily share volume varies

**TABLE I**  
Characteristics of the 100 Most Actively Traded Stocks on the Australian Stock Exchange

	<i>Daily Share Volume (Million)</i>	<i>Trade Size (Share)</i>	<i>Number of Trades Per Day</i>	<i>Dollar Spread (Cents)</i>	<i>Relative Spread (%)</i>
Mean	1.783	5,279	349	2.82	0.280
SD	1.967	5,158	266	2.28	0.179
Min.	0.127	1,186	63	0.53	0.060
Max.	12.791	37,532	1,323	11.26	0.838

*Note.* This table reports summary statistics for the 100 most actively traded stocks on the Australian Stock Exchange. The sample period spans from March 1 to March 31, 2000. Reported above are cross-sectional mean, standard deviation, minimum and maximum of daily share volume, trade size, daily number of trades, and time-weighted average of dollar (inside) bid–ask spread and relative (inside) bid–ask spread.

from 127,000 to 12,791,000 shares. Liquidity measured as the inside bid–ask spread shows an average range between 0.53 and 11.26 cents.

## PRICE DISCOVERY IN THE LIMIT-ORDER BOOK

### Summary Statistics of Order-Book Shape

At any point in time, a limit-order book contains a large number of buy and sell orders. The aggregate market demand and supply are represented by these orders as step functions of the accumulated number of shares offered at each price level. To make the order-book data amenable for empirical analysis, the authors use two summary measures designed to capture the most important features of the step functions.

- The height of a step  $i$  on the demand side is  $\nabla_i^d = P_i^d - P_{i-1}^d$ . It is the price difference between the  $i$ th and the  $(i - 1)$ th best price (regardless of the number of shares) on the buy side of the book. To compute the height of the first step of the book, the authors denote the average of the best bid and offer in the book by  $MID = P_0^d$ .
- The length of a step  $i$  on the demand side of the book,  $Q_i^d$ , is the aggregate number of shares across all orders at price  $P_i^d$ .
- The heights and lengths of steps on the supply side,  $\nabla_i^s$  and  $Q_i^s$ , are defined analogously.

The authors then normalize a step height by the cumulative price difference between the tenth step and  $MID$ . The authors also normalize a step length by the cumulative length from the first step to the tenth step.

Table II reports the cross-sectional average of the step heights and lengths. For both buys and sells, steps closer to the top of the book are generally longer. Each of the first five steps in the book represents more than 10% of the respective side's length, whereas the last step accounts for less than 4%. The second step offers more depth than the first step for both the buy and sell sides. More than 88% of the shares reside on the second step and beyond. This feature warrants a particular interest studying the orders that are located behind the top of the book. The top steps are also lower (price increments are smaller) than the away steps, and the step height increases monotonically along the order book. The first step on each side is 1.41 cents—half the inside spread. The second step goes up to 2.47 cents for buys and 2.39 cents for sells, and it increases to the tenth step at 4.06 cents for buys and 4.88 cents for sells. In terms of percentage, the first step makes up about 5% of the total height from the center top of the book to the end of step 10, whereas each step from the fifth on accounts for more than 10% of the price gap. These results suggest that the shape of the book is

**TABLE II**  
Summary Statistics of the Shape of the Limit-Order Book

Steps	Length (%)		Height (%)		Height (Cents)	
	Buy	Sell	Buy	Sell	Buy	Sell
1	11.97	11.94	5.07	5.23	1.41	1.41
2	13.23	12.67	8.55	8.97	2.47	2.39
3	12.06	11.56	9.25	9.59	2.85	2.85
4	11.04	10.77	9.84	10.12	3.18	3.33
5	10.59	10.49	10.24	10.45	3.45	3.69
6	9.91	10.16	10.63	10.62	3.66	4.05
7	9.55	9.87	10.95	10.71	3.66	4.33
8	9.06	9.58	11.28	10.88	3.76	4.44
9	8.80	9.52	11.71	11.38	3.92	4.65
10	3.80	3.43	12.48	12.06	4.06	4.88

*Note.* Reported above are, respectively, the cross-sectional averages of relative step length and height of the order book from steps 1 to 10. Relative step length is measured as the number of shares on a step as a fraction of the total number of shares in the first ten steps. Relative step height is measured as the price difference between a step and its previous step, divided by the price gap between step 10 and mid inside price (*MID*). The cross-sectional average of the absolute step height in cents is also reported. For the first step, the price of its previous step is set to be *MID*.

dense from the beginning to step 5. A denser section of the book represents higher liquidity in that section as more shares are offered with less price gaps. Figure 1 demonstrates the shape of the average book in the authors' sample.

To combine the price aspect and the quantity aspect of the book, the authors use a weighted price defined as

$$WP^{n_1-n_2} = \frac{\sum_{j=n_1}^{n_2} (Q_j^d P_j^d + Q_j^s P_j^s)}{\sum_{j=n_1}^{n_2} (Q_j^d + Q_j^s)}, \quad n_1 \leq n_2. \quad (1)$$

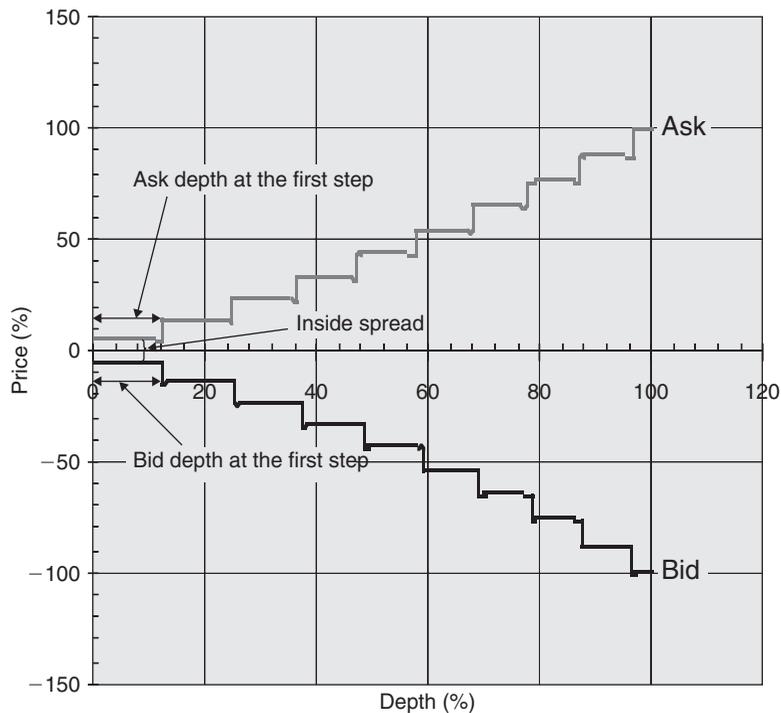
When  $n_1 = n_2 = 1$ ,  $WP^{n_1-n_2}$  is  $WP^1$ , the weighted average mid-quote:

$$WP^1 = \frac{Q_1^d P_1^d + Q_1^s P_1^s}{Q_1^d + Q_1^s}. \quad (2)$$

A measure similar to  $WP^1$  is the simple average of the best bid and ask, i.e., the mid-quote *MID*:

$$MID = \frac{P_1^d + P_1^s}{2}. \quad (3)$$

The value of  $WP^1$  will change if the height or length of the first step of the book changes, whereas  $WP^{n_1-n_2}$  changes if any of the steps between  $n_1$  and  $n_2$



**FIGURE 1**

The shape of the average limit-order book for the sample stocks. This figure shows the shape of the average limit-order book based on relative step length and height (in %) for the 100 most actively traded stocks on the Australian Stock Exchange. The sample period spans from March 1 to March 31, 2000. Relative step length (in %) is measured as the number of shares on a step as a fraction of the total number of shares in the first ten steps. Relative step height (in %) is measured as the price difference between a step and its previous step divided by the price gap between step 10 and mid inside price (*MID*). For the first step, the price of its previous step is set to be *MID*.

experience a change in either height, length, or a combination of both. The authors compute  $WP^1$  and  $MID$  by using information that has traditionally been available to traders, namely the best bid and offer prices with their associated depths.  $WP^{n_1-n_2}$  is an average price between any two steps on the book, weighted by the number of shares at each step. It summarizes all information contained in the order book from step  $n_1$  to  $n_2$ .<sup>3</sup>

### **An Error Correction Model and the Information Share of the Order Book**

To examine the marginal contribution of the book beyond the first step toward price discovery, the authors employ the methods developed by Hasbrouck

<sup>3</sup>Although  $WP^{n_1-n_2}$  is designed to summarize the book information, it is not perfect: Two books with different combinations of prices and depths may have the same value of  $WP^{n_1-n_2}$ .

(1995) to assess the information content of different price series. Specifically, the authors consider three prices,  $MID$ ,  $P$ , and  $WP^{2-10}$ , that are derived from the same limit-order book.<sup>4</sup> The last transaction price,  $P$ , is included to further alienate the information share of the book from what has traditionally been available to investors. The authors examine whether the order book provides any additional information after controlling for  $MID$  and the last transaction price  $P$ . The methodology used to estimate the information share of different price series is well established. For details, the reader is referred to Hasbrouck (1995), DeJong (2002), and Huang (2002).

Let  $X_t$  be a vector of the three related prices,  $MID$ ,  $P$ , and  $WP^{2-10}$ . Although each individual series is nonstationary, they are cointegrated and share a common stochastic trend because they are all prices of the same underlying security. As a result, the difference between any two prices is stationary.<sup>5</sup> The multivariate price process can be written in an error correction form as

$$\Delta X_t = \alpha z_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \varepsilon_t \quad (4)$$

where  $z_{t-1} = \beta' X_{t-1}$  are the error correction terms and  $k$  is the order of the original value at risk (VAR) of  $X_t$ . The authors do not include deterministic terms for clarity of exposition. The Granger representation theorem by Engle and Granger (1987) proved that cointegrated variables can be represented as a vector moving average through the use of the Wold decomposition theorem:

$$\Delta X_t = \Psi(1)\varepsilon_t + (1 - L) \sum_{i=0}^{\infty} \left( - \sum_{j=i+1}^{\infty} \Psi_j \right) L^i \varepsilon_t \quad (5)$$

where  $\Psi(1) = (I_n + \sum_{i=1}^{\infty} \Psi_i)$ . The sums of all the moving average coefficients are in the matrix  $\Psi(1)$ . Hasbrouck (1995) defined the measure of information share that can be attributed to price series  $j$  as

$$S_j = \frac{\Psi_j^2 \Sigma_{jj}}{\Psi \Sigma \Psi'} \quad (6)$$

<sup>4</sup> $MID$  is widely used as a measure of the true value of a stock. As a robustness check, the  $MID$  was replaced with  $WP^1$  and it was found that the results were qualitatively similar. To save space, these results are omitted but are available from the authors.

<sup>5</sup>The authors tested and found that the individual series were nonstationary and cointegrated using augmented Dickey Fuller tests and Johansen's (1988) procedure.

where  $\Psi_j$  is the  $j$ th element of  $\Psi$  and  $\Psi$  is any row in  $\Psi(1)$ .  $\Sigma_{jj}$  is the  $j$ th diagonal element of  $\Sigma$  and  $\Sigma$  is the variance–covariance matrix of the innovations  $\varepsilon_t$ . This information share is uniquely defined if the price series do not exhibit contemporaneous correlation, i.e., if  $\Sigma$  is a diagonal matrix. In the presence of contemporaneous correlation, Hasbrouck (1995) proposed a Cholesky decomposition procedure for  $\Sigma$ , similar to the prediction error variance decomposition in VARs:

$$\Sigma = FF' \quad (7)$$

where  $F$  is a uniquely defined triangular matrix that orthogonalizes the price innovations to

$$\varepsilon_t = F\eta_t \quad (8)$$

where  $\eta_t$  has zero mean and identity variance–covariance matrix. The information share of a price  $j$  thus computed depends on the ordering of the list of the prices. It is bounded between a minimum and a maximum value, conditional on whether  $j$  is listed as the last or first variable in  $X_t$ . The use of high-frequency intraday data in the estimation can reduce the gap between the minimum and maximum information share. Following Hasbrouck (2003), the authors sample their order-book data at second-by-second intervals in this section, and compute the average of the minimum and maximum information share. The procedure of using the average of the minimum and maximum information share as a proxy for the information share is widely adopted in the literature (see Booth, Lin, Martikainen, & Tse, 2002).

## Results

For each of the 100 stocks, the authors estimate the error correction model and obtain the maximum and minimum of the Hasbrouck information share. Table III reports the cross-sectional distributions of these estimates. Panel A reports results based on the three price series *MID*, *WP<sup>2-10</sup>*, and *P*. It shows that the information share of *MID* is the largest—the cross-sectional average is 54.50%, whereas the average information shares of *WP<sup>2-10</sup>* and *P* are 22.47 and 23.15%, respectively. Although the top of the book as represented by *MID* contributes the most to price discovery, the additional information share of *WP<sup>2-10</sup>* beyond *MID* and *P* is considerable. Judged by the reported standard deviations, the authors note that the estimates of information shares are statistically significant.

*WP<sup>2-10</sup>* aggregates nine price steps of the book and the follow-up question is whether these nine steps contribute evenly, or if there is a specific section of the book that is more informative than the others. To answer this question, the

**TABLE III**  
Estimates of the Hasbrouck (1995) Information Shares

MID			P			WP <sup>2-10</sup>		
Min.	Ave.	Max.	Min.	Ave.	Max.	Min.	Ave.	Max.
<i>Panel A: Estimates of information share (in %) using MID, P, and WP<sup>2-10</sup></i>								
Mean	53.40	55.60	22.33	23.15	23.31	21.62	23.31	22.47
SD	11.14	10.19	7.02	8.16	11.03	10.50	11.03	19.68
Median	53.95	56.02	22.33	22.98	20.63	18.72	20.63	
MID			P			WP <sup>2-4</sup>		
Min.	Ave.	Max.	Min.	Ave.	Max.	Min.	Ave.	Max.
<i>Panel B: Estimates of information share (in %) using MID, P, WP<sup>2-4</sup>, and WP<sup>5-10</sup></i>								
Mean	49.19	51.27	20.55	21.38	12.29	9.76	11.03	17.60
SD	12.36	11.35	6.62	7.68	6.48	4.84	8.7	10.47
Median	50.53	51.68	20.69	21.56	10.95	9.24	10.10	15.55

Note. For each of the 100 sample stocks, the Hasbrouck (1995) information shares are estimated. Reported above are the cross-sectional average, standard deviation, and median of the information-share estimates for two sets of price series. The estimated error correction model is

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

where  $X = (MID, P, WP^{2-10})$ , in Panel A and  $X = (MID, P, WP^{2-4}, WP^{5-10})$ , in Panel B.  $k$  is the number of lags determined by AIC information criterion.  $MID$  is the bid-ask mid-price,  $WP^{n_1-n_2}$  is the volume-weighted average price from steps  $n_1$  to  $n_2$  of the limit-order book, and  $P$  is the last transaction price. The maximum (minimum) contribution of a price to the variance of the common factor's innovation is when it is the first (last) variable in the Cholesky factorization. The reported numbers are in percentages. AIC, Akaike Information Criterion.

authors partition  $WP^{2-10}$  into  $WP^{2-4}$  and  $WP^{5-10}$ , and re-examine the error correction model using a vector of four variables  $X = (MID, WP^{2-4}, WP^{5-10}, P)$ . Panel B of Table III reports that the results for  $P$  change little—its information share is 21.38%. The information share of  $MID$  is still the largest (50.23%), whereas the information share of  $WP^{2-4}$  and  $WP^{5-10}$  are 11.03 and 17.60%, respectively.

The larger information share of the later section of the book (e.g.,  $WP^{5-10}$ ) reaffirms the importance of evaluating the orders along the limit-order book. Order activities at or near the top of the book could be unrelated to new information. Hasbrouck and Saar (2002) documented the presence of orders used to fish for hidden orders or to spoof the traders on the other side of the market. A seller can submit a sell order priced just above the best bid price fishing for trading against any hidden buy orders at the better price level. Other times, a seller may submit a large buy order that betters the current best buy price hoping to entice other buyers to match or even outbid so as to sell to them. Whether somebody on the buy side falls into the trap or not, the “faked” buy order will be cancelled by the seller within a second or two. This spoofing strategy was a heavily debated practice on the Eurex system in 2004, as a flipper profited handsomely from posting flipping bids and offers on the Bobl and Schatz interest rate futures contracts (*Financial Times*, 2004/4/19). Hence the top of the book may be very active but polluted by these “fishing” and/or spoofing actions. The later section of the order book may contain orders with more “stable” and less “noisy” information.

As a robustness check, the authors re-estimate the Hasbrouck information share for each stock using the order-book data, sampled every five minutes. The authors’ results indicate that the average information share of  $WP^{2-10}$  is 25.02%, which is close to the 22.47% reported in Panel A of Table III. As a result of sampling at a lower frequency, the difference between the minimum and maximum information share is slightly larger. Overall, the authors results are consistent with Hypothesis 1 and suggest that orders behind the best bid and ask make a modest contribution to the price discovery.

## **ORDER-BOOK INFORMATION AND SHORT-TERM STOCK RETURNS**

### **Methodology**

In this section, the authors investigate whether the order-book information (e.g., the demand and supply schedule) is associated with future short-term returns. In the literature, there is ample evidence of short-term predictability in stock returns. It is conceivable that some information contained in the order book (e.g., the imbalance between demand and supply schedules) may be useful

in forecasting short-term returns. For dealer markets, the theoretical model of Cao, Evans, and Lyons (2006) showed that a dealer's superior knowledge of market demand and supply conditions helps them forecast price changes. However, in their study, the information on order flows is private, whereas the order book under the authors' investigation is public. The question the authors ask is whether the public information on the book is associated with short-term future returns.

The authors first examine the relation between five-minute returns and lagged order-book statistics that are constructed from the demand and supply schedules. Snapshots of the order book at five-minute intervals are taken to strike a balance between the desire to have a large number of observations and the need to allow the share price to experience a meaningful change between any two subsequent observations. The authors experiment with one-second, one-minute, five-minute, and ten-minute samples and find that the five-minute sample strikes the best balance. As short-term returns are serially correlated, the authors pre-whiten returns to focus on the innovation in returns. The return innovation is used as the dependent variable in the regression analysis because the innovation captures the unpredictable component of the return. Specifically, the authors obtain innovation in return by using an AR(5) model for each sample firm<sup>6</sup>:

$$r_t = \alpha_0 + \sum_{i=1}^5 \alpha_i r_{t-i} + \varepsilon_t^* \quad (9)$$

where  $r_t$  is the five-minute mid-price (*MID*) return at  $t$  and  $\varepsilon_t^*$  the return innovation. The independent variables include the inside spread (*Spread*), the trade imbalance (*Timb*), and the imbalance in the length (quantity) and height (price) of the order book. *Timb* is measured as the imbalance between the executed buy and sell volumes in a five-minute period. Buy volume and sell volume are classified when marketable buy orders and sell orders consume limit orders sitting on the other side of the book. In a sense, *Timb* aggregates the net transaction volume every five minutes. The authors include *Spread* and *Timb* to control for effects of the top of the book that investors usually have access to. For each step  $j$ , let  $QR_j$  and  $HR_j$  denote the imbalance in the length and height:

$$QR_j = \frac{Q_j^s - Q_j^d}{Q_j^s + Q_j^d}, \quad j = 1, \dots, 10 \quad (10)$$

<sup>6</sup>On the basis of the Akaike Information Criterion, the authors find that an AR model with five lags is sufficient for the sample firms.

$$HR_j = \frac{(P_j^s - P_{j-1}^s) - (P_j^d - P_{j-1}^d)}{(P_j^s - P_{j-1}^s) + (P_j^d - P_{j-1}^d)}, \quad j = 2, \dots, 10. \quad (11)$$

Intuitively,  $QR$  (and  $HR$ ) can be interpreted as the step-wise scaled imbalance in quantity (and price) between the demand and supply schedules.

The authors' first set of regressions contains ten regressions; each includes the book information up to step  $n$  where  $n = 1, 2, \dots, 10$ :

$$\varepsilon_t^* = \alpha_0 + \delta_0 Timb_{t-1} + \beta_0 Spread_{t-1} + \gamma_1 QR_{1,t-1} + \sum_{j=2}^n \beta_j HR_{j,t-1} + \sum_{j=2}^n \gamma_j QR_{j,t-1} + \eta_t. \quad (12)$$

The list of the independent variables starts with  $Timb_{t-1}$  and  $Spread_{t-1}$ . The authors add  $QR_1$  when  $n = 1$ , and  $QR_n$  and  $HR_n$  for  $n = 2, 3, \dots, 10$ . The regression on a stock-by-stock basis is run and results are reported based on the cross-sectional averages. As the objective is to investigate whether the book beyond the first step helps to predict future returns, the fact that whether the average *Adjusted R*<sup>2</sup> increases as more steps of the book are included in the regression is examined. The *Adjusted R*<sup>2</sup> is reported because its change summarizes the role that any additional book information may have in explaining the dependent variable, while controlling for the number of independent variables.

The authors briefly discuss the anticipated relation between stock returns and book imbalance. Intuitively, return is predicted to be negatively related to  $QR_{j,t-1}$  as the excess supply drives the share price down when more shares are submitted to the supply side. Furthermore, return is expected to increase with  $HR_{j,t-1}$  because smaller price increments on the demand side suggest more interest in purchasing, rather than selling shares. Several studies on order placement strategies document a relation between the status of the order book and future order submissions (see Biais, Hillion, & Spatt, 1995; Cao, Hansch, & Wang, 2008; Foucault, 1999; Griffiths, Smith, Turnbull, & White, 2000; Hollifield, Miller, & Sandas, 2004; Parlour, 1998; Rinaldo, 2004). These studies find that a large depth at the buy (sell) side of the book encourages more market buy (sell) orders. These market orders consume limit orders on the other side of the order book and drive the price higher (lower). It is the link between the book and order placement strategies that connects future price movement and current book status.<sup>7</sup>

Another way to capture the imbalance in book length and height is to consider the demand and supply side of the book separately. The authors use

<sup>7</sup>The snapshot of the limit-order book represents the static view of the book. It is an aggregation of the order flows within a certain time-frame. The higher the frequency at which the snapshots are taken, the closer they represent the dynamic order flows.

the height and length of steps on the demand and supply sides (e.g.,  $Q_j^d, Q_j^s, \nabla_j^d, \nabla_j^s$ ) as independent variables and run a second set of regressions for  $n = 1, 2, \dots, 10$ :

$$\begin{aligned} \varepsilon_t^* = & \alpha_0 + \delta_0 \text{Timb}_{t-1} + \beta_0 \text{Spread}_{t-1} + \sum_{j=1}^n \gamma_{d,j} Q_{j,t-1}^d \\ & + \sum_{j=1}^n \gamma_{s,j} Q_{j,t-1}^s + \sum_{j=2}^n \beta_{d,j} \nabla_{j,t-1}^d + \sum_{j=2}^n \beta_{s,j} \nabla_{j,t-1}^s + \eta_t. \end{aligned} \quad (13)$$

In addition to the imbalance in book length and height, the authors use price impact measures as independent variables. Price impact is measured ex ante as a cost of trading for a hypothetical trade size of  $q$  shares. Given the fact that the average trade size varies across firms, the authors choose  $q$  as multiples of the firm-specific average trade size  $\bar{Q}$  (e.g.,  $q = 1.0\bar{Q}, 1.5\bar{Q}, \dots, 5.0\bar{Q}$ ). Let  $LD(q)$  be the price impact measure on the demand side, which is defined as the discount per share that a market-seller gets below the midpoint of the best bid and ask:

$$LD(q) = 0.5(P_1^d + P_1^s) - \frac{\sum_{j=1}^{m_1-1} P_j^d Q_j^d + P_{m_1}^d Q_{m_1}^d}{q} \quad \text{and} \quad \sum_{i=1}^{m_1-1} Q_i^d + Q_{m_1}^d = q \quad (14)$$

where the step  $m_1$  is determined according to  $\sum_{i=1}^{m_1-1} Q_i^d < q \leq \sum_{i=1}^{m_1} Q_i^d$  and  $Q_{m_1}^d$  is the number of shares on step  $m_1$  to fulfill the order of  $q$  shares after the first  $m_1 - 1$  steps are filled. Similarly, the price impact on the supply side  $LS(q)$  is the premium per share a market-buyer needs to pay above the midpoint of the best bid and ask:

$$LS(q) = \frac{\sum_{j=1}^{m_2-1} P_j^s Q_j^s + P_{m_2}^s Q_{m_2}^s}{q} - 0.5(P_1^d + P_1^s) \quad \text{and} \quad \sum_{i=1}^{m_2-1} Q_i^s + Q_{m_2}^s = q. \quad (15)$$

The step  $m_2$  is determined according to  $\sum_{i=1}^{m_2-1} Q_i^s < q \leq \sum_{i=1}^{m_2} Q_i^s$  and  $m_2$  is not necessarily equal to  $m_1$ . Intuitively,  $LS(q)$  and  $LD(q)$  are inverse measures of liquidity. The authors compute the scaled imbalance in price impact,  $LR(q)$ , as the following:

$$LR(q) = \frac{LS(q) - LD(q)}{LS(q) + LD(q)}, \quad q = g^*(\bar{Q}/2) \quad \text{and} \quad g = 2, 3, \dots, 10 \quad (16)$$

where the hypothetical trade size  $q$  takes value of  $q = 1.0\bar{Q}, 1.5\bar{Q}, \dots, 5.0\bar{Q}$ .

The third set of regressions relies on the scaled price impact measure  $LR(q)$ , and the fourth set of regressions uses the separated price impact on the demand and supply side as independent variables ( $g = 2, 3, \dots, 10$ ):

$$\varepsilon_t^* = \alpha_0 + \delta_0 Timb_{t-1} + \beta_0 Spread_{t-1} + \sum_{j=2}^g \beta_j LR(j\bar{Q}/2)_{t-1} + \eta_t \quad (17)$$

$$\varepsilon_t^* = \alpha_0 + \delta_0 Timb_{t-1} + \beta_0 Spread_{t-1} + \sum_{j=2}^g \beta_{d,j} LD(j\bar{Q}/2)_{t-1} + \sum_{j=2}^g \beta_{s,j} LS(j\bar{Q}/2)_{t-1} + \eta_t. \quad (18)$$

As smaller price impact implies better liquidity, the authors predict that the return at  $t$  is positively related to  $LR(q)_{t-1}$  (or  $LS(q)_{t-1}$ ) and negatively related to  $LD(q)_{t-1}$ . In other words, when there is more liquidity on the demand side, more limit buy orders are placed at crowded buy steps, which attracts more market buy orders, and these market orders drive up the price.

Although the authors' empirical design intends to uncover the dynamics between the return and lagged book information in general, it is of potential importance to examine the association between return and the book when there is a large asymmetry between the supply and demand. If the book is severely imbalanced, it may reveal the excess demand or supply that allows traders to better infer future movements of the price. To investigate this question, each firm's observations by the imbalance in price impact  $LR(q)$  is sorted with  $q = 2.5\bar{Q}$ <sup>8</sup> and the top and bottom 5% observations from each firm is used to re-estimate the above four sets of regressions.

## Empirical Results

Table IV presents the results of the first set of regressions specified in Equation (12). The dependent variable is the return innovation from an  $AR(5)$  model, and independent variables include trade imbalance, inside spread, and the scaled imbalance in quantity and price. For each firm the regression is estimated ten times; each regression includes the book information up to step  $n$  ( $n = 1, 2, \dots, 10$ ). Two samples are considered: (1) the full sample and (2) a restricted sample with the top and bottom 5% observations sorted by  $LR(2.5\bar{Q})$ .

First the cross-sectional average of *Adjusted R*<sup>2</sup> is examined for the full sample. With the trade imbalance and spread as the independent variables, the average *Adjusted R*<sup>2</sup> is merely 0.59%. After adding  $QR_1$  as an independent variable, the average *Adjusted R*<sup>2</sup> increases to 5.05%. When the lagged book imbalances

<sup>8</sup>The authors experimented with other choices of  $q$  and found that their conclusion is quantitatively similar.

**TABLE IV**  
Regression Analysis of Returns and Scaled Order Imbalances

Steps	Full Sample			Restricted Sample		
	Adjusted R <sup>2</sup>	% of Significant Coefficients With Anticipated Signs	% of Firms That Pass F-Test	Adjusted R <sup>2</sup>	% of Significant Coefficients With Anticipated Signs	% of Firms That Pass F-Test
<i>Timb and Spread</i>	0.59	–	100	2.37	–	100
1	5.05	100	100	17.40	97	97
2	6.55	72	82	19.87	47	37
3	6.76	51	27	20.87	31	16
4	6.83	39	16	21.08	23	5
5	6.91	31	11	21.55	19	7
6	7.00	27	15	21.79	15	5
7	7.03	24	6	22.17	14	9
8	7.08	21	11	22.76	13	12
9	7.16	19	14	23.13	11	5
10	7.22	18	9	24.15	11	17

Note. For each of the 100 sample stocks, the following regression model is estimated ten times by using the lagged order-book information up to step  $n$  ( $n = 1, 2, \dots, 10$ ):

$$\varepsilon_t^* = \alpha_0 + \delta_0 \text{Timb}_{t-1} + \beta_0 \text{Spread}_{t-1} + \gamma_1 \text{QR}_{1,t-1} + \sum_{j=2}^n \beta_j \text{HR}_{j,t-1} + \sum_{j=2}^n \gamma_j \text{QR}_{j,t-1} + \eta_t$$

where  $\varepsilon^*$  is the innovation in returns estimated by using an  $AR(5)$  model, *Timb* is the trade imbalance, *Spread* is the inside spread,  $\text{QR}_j$  is the scaled imbalance in quantity at step  $j$ , and  $\text{HR}_j$  is the scaled imbalance in price at step  $j$ . Each regression model is estimated for the full sample and a restricted sample. The restricted sample is based on the top and bottom 5% of imbalanced price impact measured at  $2.5 \cdot \bar{Q}$ , where  $\bar{Q}$  is the stock-specific average trade size. Reported above are the cross-sectional averages of the *Adjusted R<sup>2</sup>* (in %), the percentage of coefficient estimates with anticipated signs, and the percentage of regressions (among 100) that pass the *F*-test at 5% significance level. For a given step  $n$ , the null hypothesis of the *F*-test is that coefficients of  $\text{HR}_n$  and  $\text{QR}_n$  are jointly zero.

from steps 2 to 10 are included, the average *Adjusted R<sup>2</sup>* goes up to 7.22%, whereas the greatest increase occurs at step 2.

When using the top and bottom 5% observations according to  $LR(2.5\bar{Q})$ , a stronger result is found. For example, the cross-sectional average of *Adjusted R<sup>2</sup>* increases from 2.37 to 17.40% for  $n = 1$ , and then to 24.15% for  $n = 10$ . Unlike the full sample result, where the majority of the improvement occurs at steps 1 and 2, the authors find a continued improvement in fitting result along the book for the restricted sample. In comparison with the full sample result, the explanatory power is higher for the restricted sample. In summary, the order-book information is useful in explaining short-term future returns. This result is particularly true when the order book shows a large imbalance.

The results reported in Table IV rely on 20 regressions for each of the 100 sample firms. Owing to space limitation, the regression coefficients and their *T*-statistics are not reported. Instead, the results are summarized by presenting the percentage of significant coefficients at the 5% significance level with the

expected signs. The authors find that the range of anticipated signs of estimated coefficients spans from 100% (if the book information in step 1 is used) to 72% (if the book information from steps 1 and 2 is used). According to the signs of estimated coefficients, future returns tend to be lower if more shares are submitted to the supply side. On the other hand, if buy orders are submitted at smaller price increments than sell orders, the demand for the stock increases and future returns tend to be higher. This latter finding is unique, and it highlights the importance of the price dimension of the supply and demand schedule. Without viewing the book information behind the best bid and ask, investors may forever miss the value buried in the price dimension of the order book.

Although the percentage of significant coefficients with the anticipated signs drops as more variables along the book are included in the regression, the authors find that the near orders tend to conform to the anticipation. The faraway orders have the opposite effect but the majority of their coefficients are insignificant. The accumulation of buy (sell) limit orders near the top of the book attracts buy (sell) market orders as traders become aggressive, to get their orders executed first. As a result, price is driven up (down). Therefore, the order book, especially the section of the book that is near the top, helps investors to predict future short-term returns. This effect persists after controlling for the autocorrelations in returns and other commonly observed variables such as inside spread and trade imbalance from past transactions. Chordia et al. (2002) documented a positive association between aggregate order imbalance and market return, and Chordia and Subrahmanyam (2004) showed the same cross-sectional results in stocks. Boehmer and Wu (2006) found that institutional investors' trade imbalance (buy–sell) positively affects future price movements, whereas the opposite impact holds for individual investors' trade imbalance. These trade imbalance effects are controlled for and the authors' investigation is extended beyond the top of the book and it is found that imbalance in the existing limit orders along the book also affects future price positively.

To test for the joint significance of the coefficients added in each step, an  $F$ -test is conducted using 5% significance level. For a given step  $n$ , the null hypothesis of the  $F$ -test is that coefficients of  $HR_n$  and  $QR_n$  are jointly zero. Table IV reports the percentages of the 100 firms that pass the  $F$ -test in each step. For the full sample, 82% of the sample firms pass the  $F$ -test when the book information from step 2 is included in the model. This result suggests that the null hypothesis that coefficients of  $HR_2$  and  $QR_2$  are jointly zero for 82 firms (among 100) should be rejected. When the book information from step 3 is added, the null hypothesis that coefficients of  $HR_3$  and  $QR_3$  are jointly zero for 27% of the sample firms is rejected. The drop in the  $F$ -test passing rate

from step 1 to 10 is consistent with the drop in the percentage of the significant coefficients and is also consistent with the gradual but small increase in the *Adjusted R*<sup>2</sup>. Beyond step 4, the percentage of firms that pass the *F*-test is below 20%, indicating that coefficients associated with  $HR_n$  and  $QR_n$  ( $n = 5, 6, \dots, 10$ ) are generally insignificant.

Table V presents estimation results of the second set of regressions (the model is provided in Equation (13)). When using the height and length of steps on the demand and supply sides as independent variables, the null hypothesis of the authors' *F*-test is that all coefficients added in step  $n$  are jointly zero. The authors find that the results reported in Table V are consistent with those in Table IV. Specifically, the average *Adjusted R*<sup>2</sup> for the full sample increases from 0.59 to 6.31% (when  $n = 10$ ). The change in the average *Adjusted R*<sup>2</sup> is from 2.37 to 28.02% (when  $n = 10$ ) for the restricted sample. Turning to the percentage of coefficient estimates with anticipated signs, the authors find that

**TABLE V**  
Regression Analysis of Returns and Order-Book Step Length (and Height)

Steps	Full Sample			Restricted Sample		
	Adjusted <i>R</i> <sup>2</sup>	% of Significant Coefficients With Anticipated Signs	% of Firms That Pass <i>F</i> -Test	Adjusted <i>R</i> <sup>2</sup>	% of Significant Coefficients With Anticipated Signs	% of Firms That Pass <i>F</i> -Test
<i>Timb and Spread</i>	0.59	–	100	2.37	–	100
1	2.94	93	96	12.85	93	88
2	4.52	59	76	18.11	59	51
3	4.88	41	25	19.79	41	17
4	5.09	31	17	21.31	31	18
5	5.24	25	17	22.92	25	10
6	5.38	22	18	23.58	22	10
7	5.57	20	21	24.75	20	12
8	5.74	18	19	25.48	18	13
9	5.96	17	25	26.65	17	12
10	6.31	16	28	28.02	16	18

*Note.* For each of the 100 sample stocks, the following regression model is estimated ten times by using the order-book information up to step  $n$  ( $n = 1, 2, \dots, 10$ ):

$$\varepsilon_t^* = \alpha_0 + \delta_0 \text{Timb}_{t-1} + \beta_0 \text{Spread}_{t-1} + \sum_{j=1}^n \gamma_{d,j} Q_{j,t-1}^d + \sum_{j=1}^n \gamma_{s,j} Q_{j,t-1}^s + \sum_{j=2}^n \beta_{d,j} \nabla_j^d + \sum_{j=2}^n \beta_{s,j} \nabla_j^s + \eta_t$$

where  $\varepsilon^*$  is the innovation in returns estimated by using an *AR*(5) model, *Timb* is the trade imbalance, *Spread* is the inside spread,  $Q_j^d(Q_j^s)$  is the length of step  $j$  on the demand (supply) side, and  $\nabla_j^d(\nabla_j^s)$  is the height of step  $j$  on the demand (supply) side of the order book. Each regression model is estimated for the full sample and a restricted sample. The restricted sample is based on the top and bottom 5% of imbalanced price impact measured at  $2.5^* \bar{Q}$ , where  $\bar{Q}$  is the stock-specific average trade size. Reported above are the cross-sectional averages of the *Adjusted R*<sup>2</sup> (in %), the percentage of coefficient estimates with anticipated signs, and the percentage of regressions (among 100) that pass the *F*-test at 5% significance level. For a given step  $n$ , the null hypothesis of the *F*-test is that all coefficients added in step  $n$  are jointly zero.

the results reported in Table V are comparable to those in Table IV. The percentages of firms that pass the  $F$ -test are larger than those in Table IV for both the full and restricted samples. Taking the restricted sample as an example, when the book information from step 4 is added, the authors reject the null hypothesis that all coefficients added in step 4 are jointly zero for 18% of the sample firms, whereas the same test in Table IV has only a 5% passing rate.

The results of the third set of regressions are provided in Table VI. These regressions intend to uncover the relation between returns at  $t$  and price impact (liquidity) measures along the book at  $t - 1$ . For a given trade size  $q = g^* \bar{Q}/2$ , the null hypothesis of the  $F$ -test is that  $\beta_g$  is zero. For the full sample, the authors show that the average *Adjusted R*<sup>2</sup> is 0.59% with trade imbalance and spread as the only two independent variables. It increases to 4.75% at  $q = \bar{Q}$  and then to 6.56% at  $q = 5\bar{Q}$  when the scaled price impact measures are used as additional explanatory variables. Using the restricted sample, the authors find that the average *Adjusted R*<sup>2</sup> increases from 2.37 to 20.40%. The majority of the coefficient estimates have the anticipated signs. This result

**TABLE VI**  
Regression Analysis of Returns and Scaled Imbalance in Price Impact

$q$	Full Sample			Restricted Sample		
	<i>Adjusted R</i> <sup>2</sup>	% of Significant Coefficients With Anticipated Signs	% of Firms That Pass $F$ -Test	<i>Adjusted R</i> <sup>2</sup>	% of Significant Coefficients With Anticipated Signs	% of Firms That Pass $F$ -Test
<i>Timb and Spread</i>	0.59	–	100	2.37	–	100
$1.0\bar{Q}$	4.75	100	100	17.48	98	98
$1.5\bar{Q}$	5.42	38	71	18.73	13	24
$2.0\bar{Q}$	5.88	28	64	19.74	9	18
$2.5\bar{Q}$	6.20	16	44	19.82	5	9
$3.0\bar{Q}$	6.46	13	40	19.95	5	8
$3.5\bar{Q}$	6.63	12	34	20.18	5	7
$4.0\bar{Q}$	6.78	10	27	20.16	4	5
$4.5\bar{Q}$	6.90	9	23	20.44	4	7
$5.0\bar{Q}$	6.98	7	13	20.40	4	3

*Note.* For each of the 100 sample stocks, the following regression model is estimated nine times by using the lagged price impact measures:

$$\varepsilon_t^* = \alpha_0 + \delta_0 \text{Timb}_{t-1} + \beta_0 \text{Spread}_{t-1} + \sum_{j=2}^g \beta_j \text{LR}(g^* \bar{Q}/2)_{t-1} + \eta_t$$

where  $\varepsilon^*$  is the innovation in returns estimated by using an  $AR(5)$  model, *Timb* is the trade imbalance, *Spread* is the inside spread, and  $\text{LR}(g^* \bar{Q}/2)$  is the scaled imbalance in price impact for a hypothetical trade size  $q = g^* \bar{Q}/2$  and  $g = 2, 3, \dots, 10$ . Each regression model is estimated for the full sample and a restricted sample. The restricted sample is based on the top and bottom 5% of imbalanced price impact measured at  $2.5^* \bar{Q}$ , where  $\bar{Q}$  is the stock-specific average trade size. Reported above are the cross-sectional averages of the *Adjusted R*<sup>2</sup> (in %), the percentage of coefficient estimates with anticipated signs, and the percentage of regressions (among 100) that pass the  $F$ -test at 5% significance level. For a given trade size  $q = g^* \bar{Q}/2$ , the null hypothesis of the  $F$ -test is that  $\beta_g$  is zero.

**TABLE VII**  
Regression Analysis of Returns and Price Impact Measures

$q$	Full Sample			Restricted Sample		
	Adjusted $R^2$	% of Significant Coefficients With Anticipated Signs	% of Firms That Pass $F$ -Test	Adjusted $R^2$	% of Significant Coefficients With Anticipated Signs	% of Firms That Pass $F$ -Test
<i>Timb and Spread</i>	0.59	–	100	2.37	–	100
$1.0\bar{Q}$	5.05	96	99	17.44	96	93
$1.5\bar{Q}$	6.01	29	72	19.80	29	31
$2.0\bar{Q}$	6.55	21	61	21.64	21	29
$2.5\bar{Q}$	6.88	14	37	22.59	14	21
$3.0\bar{Q}$	7.21	12	38	23.12	12	13
$3.5\bar{Q}$	7.48	11	36	23.64	11	8
$4.0\bar{Q}$	7.68	10	21	24.37	10	17
$4.5\bar{Q}$	7.90	9	28	24.73	9	10
$5.0\bar{Q}$	8.02	9	18	25.38	9	15

*Note.* For each of the 100 sample stocks, the following regression model is estimated nine times by using the lagged price impact measures:

$$\varepsilon_t^* = \alpha_0 + \delta_0 \text{Timb}_{t-1} + \beta_0 \text{Spread}_{t-1} + \sum_{j=2}^g \beta_{d,j} LD(g^* \bar{Q}/2)_{t-1} + \sum_{j=2}^g \beta_{s,j} LS(g^* \bar{Q}/2)_{t-1} + \eta_t$$

where  $\varepsilon^*$  is the innovation in returns estimated by using an  $AR(5)$  model, *Timb* is the trade imbalance, *Spread* is the inside spread, and  $LD(g^* \bar{Q}/2)$  and  $LS(g^* \bar{Q}/2)$  are price impact measures on the demand and supply side for a hypothetical trade size  $q = g^* \bar{Q}/2$  and  $g = 2, 3, \dots, 10$ . The regression model is estimated for the full sample and a restricted sample. The restricted sample is based on the top and bottom 5% imbalanced price impact measured at  $2.5^* \bar{Q}$ , where  $\bar{Q}$  is the stock-specific average trade size. Reported above are the cross-sectional averages of the *Adjusted  $R^2$*  (in %), the percentage of coefficient estimates with anticipated signs, and the percentage of regressions (among 100) that pass the  $F$ -test at 5% significance level. For a given trade size  $q = g^* \bar{Q}/2$ , the null hypothesis of the  $F$ -test is that  $\beta_{d,g}$  and  $\beta_{s,g}$  are jointly zero.

suggests that stock return increases when the demand side is more liquid than the supply side. The better liquidity on the demand side is the result of demand outweighing supply. Finally, the percentages of firms that pass the  $F$ -test are larger than 25% for the full sample for all trade sizes (with two exceptions). Taking the trade size  $q = 2\bar{Q}$  as an example, the authors reject the null hypothesis that  $\beta_2$  is zero for 64% of sample firms. Therefore, the coefficient of  $\beta_2$  is significant for majority of the sample firms.

The authors present the results of the fourth set of regressions in Table VII. In this table, measures of the demand- and supply-side price impact enter the regression model as separate variables. These results are slightly stronger than those reported in Table VI and reinforce the authors' earlier findings. Overall, the regression results indicate that returns are significantly related to the lagged order-book information after controlling for autocorrelations in return, the inside spread, and the trade imbalance. Consistent with Hypothesis 2, the book information, especially the book information derived from steps 2 to 4, is useful and does provide additional explanatory power. Though not as strong as

the results reported in Table III, the fact that the information-share analysis relates to the process of price discover, which may take longer than five minutes, is understood, whereas the price prediction is for the next five-minute interval.

## CONCLUSION

In spite of the success of limit-order books in financial markets in the United States and around the world, empirical research on the information content of the order book is sparse and only emerging. The academic literature has customarily used the quoted mid-price as a proxy for the true asset value when considering dealer markets, or when order-book information is not available. The value of the information content beyond the first price step of the book is largely unexplored.

This article examines the value of the order book from two perspectives: (1) In comparison with the best bid and ask prices and their depths, the authors ask whether investors can make better estimations about the true value of the underlying stock by using book information beyond the first step, and (2) they ask whether future short-term returns are associated with current demand and supply schedules.

Our empirical evidence indicates that the order book beyond the first step is moderately informative about the true value of the asset. According to the authors' estimates of the Hasbrouck (1995) information share for the 100 sample stocks, it is found that the contribution of the order book beyond the best bid and offer is significant and is approximately 22%. In addition, the authors find that the lagged book information behind the best bid and ask is significantly related to future returns. This is particularly true for a restricted sample with an extremely imbalanced order book.

The authors' results have important implications given the substantial changes that financial markets have experienced in their regulatory framework and competitive landscape in the recent past. The reduction in the minimum price variations from eighths to sixteenths and subsequently to decimals leads to a reduction in the quoted depth as a large number of orders have been shifted to hide behind the quotes. As a result, pre-trade transparency has been demanded by investors. ECNs successfully satisfy investors' request by disclosing the order book (to a certain degree), and so they are becoming more and more popular in the trading landscape. Recently, in 2005, the NYSE–Archipelago merger and the Nasdaq/INET acquisition once again pushed the two most popular ECNs on the same stage with the two biggest traditional market places. The pre-trade transparency offered by many of these open book trading platforms will certainly be exploited by sophisticated traders. The order book's

information share and its predictive power for short-term price movements can help accelerate the overall process of price discovery and increase the overall welfare for investors at large.

## BIBLIOGRAPHY

- Angel, J. (1997). Limit versus market orders (working paper). Georgetown University.
- Baruch, S. (2005). Who benefits from an open limit-order book? *Journal of Business*, 78, 1267–1306.
- Bessembinder, H., & Venkataraman, K. (2004). Does an electronic stock exchange need an upstairs market? *Journal of Financial Economics*, 73, 3–36.
- Biais, B., Hillion, P., & Spatt, C. (1995). An empirical analysis of the limit order book and the order flow in the Paris Bourse. *Journal of Finance*, 50, 1655–1689.
- Bloomfield, R., O'Hara, M., & Saar, G. (2005). The “make or take” decision in an electronic market: Evidence on the evolution of liquidity. *Journal of Financial Economics*, 75, 165–199.
- Boehmer, E., Saar, G., & Yu, L. (2005). Lifting the veil: An analysis of pre-trade transparency at the NYSE. *Journal of Finance*, 60, 783–815.
- Boehmer, E., & Wu, J. (2006). Order flow and prices (working paper). Texas A&M University.
- Booth, G., Lin, J.-C., Martikainen, T., & Tse, Y. (2002). Trading and pricing in upstairs and downstairs markets. *Review of Financial Studies*, 15, 1111–1135.
- Cao, C., Hansch, O., & Wang, X. (2008). Order placement strategies in a pure limit order book market. *Journal of Financial Research*, 31, 113–140.
- Cao, H., Evans, M., & Lyons, R. (2006). Inventory information. *Journal of Business*, 79, 325–364.
- Chordia, T., Roll, R., & Subrahmanyam, A. (2002). Order imbalance, liquidity and market returns. *Journal of Financial Economics*, 65, 111–130.
- Chordia, T., & Subrahmanyam, A. (2004). Order imbalance and individual stock returns: Theory and evidence. *Journal of Financial Economics*, 72, 485–518.
- DeJong, F. (2002). Measures of contributions to price discovery: A comparison. *Journal of Financial Markets*, 5, 323–328.
- Foucault, T. (1999). Order flow composition and trading costs in a dynamic limit order market. *Journal of Financial Markets*, 2, 99–134.
- Glosten, L. (1994). Is the electronic open limit order book inevitable? *Journal of Finance*, 49, 1127–1161.
- Griffiths, M., Smith, B., Turnbull, A., & White, R. (2000). The costs and determinants of order aggressiveness. *Journal of Financial Economics*, 56, 65–88.
- Harris, L. (1990). Liquidity, trading rules, and electronic trading systems, New York University. Monograph series in finance and economics, Monograph 1990–4.
- Harris, L. (1998). Optimal dynamic order submission strategies in some stylized trading problems. *Financial Markets, Institutions and Instruments*, 7, 1–76.
- Harris, L., & Panchapagesan, V. (2005). The information-content of the limit order book: Evidence from NYSE specialist trading decisions. *Journal of Financial Markets*, 8, 25–67.
- Hasbrouck, J. (1995). One security, many markets: Determining the contributions to price discovery. *Journal of Finance*, 50, 1175–1199.

- Hasbrouck, J. (2003). Intraday price formation in the market for U.S. equity indexes. *Journal of Finance*, 58, 2375–2400.
- Hasbrouck, J., & Saar, G. (2002). Limit orders and volatility in a hybrid market: The island ECN (working paper). New York University.
- Hollifield, B., Miller, R., & Sandas, P. (2004). Empirical analysis of limit order markets. *Review of Economic Studies*, 71, 1027–1063.
- Huang, R. (2002). The quality of ECN and Nasdaq market maker quotes. *Journal of Finance*, 57, 1285–1319.
- Huang, R., & Stoll, H. (1994). Market microstructure and stock return predictions. *Review of Financial Studies*, 7, 179–213.
- Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12, 231–254.
- Kaniel, R., & Liu, H. (2006). So what orders do informed traders use? *Journal of Business*, 79.
- Kavajecz, K., & Odders-White, E. (2004). Technical analysis and liquidity provision. *Review of Financial Studies*, 17, 1043–1071.
- Madhavan, A., Potter, D., & Weaver, D. (2005). Should securities markets be transparent? *Journal of Financial Markets*, 8, 265–287.
- Parlour, C. (1998). Price dynamics in limit order markets. *Review of Financial Studies*, 11, 789–816.
- Ranaldo, A. (2004). Order aggressiveness in limit order book markets. *Journal of Financial Markets*, 7, 53–74.
- Rock, K. (1996). The specialist's order book and price anomalies (working paper). Harvard University.
- Seppi, D. (1997). Liquidity provision with limit orders and a strategic specialist. *Review of Financial Studies*, 10, 103–150.