The Informational Content of an Open Limit Order Book

Charles Cao Oliver Hansch Xiaoxin Wang*

March 15, 2004

Abstract

We assess the informational content of an open limit order book from three directions: (1) Does the limit order book allow better inferences about a security's value than simply the best bid and offer prices from the first step of the book? If it does, how much additional information can be gleaned from the book? (2) Are imbalances between the demand and supply schedules informative about future price movements? and (3) Does the shape of the limit order book impact traders' order submission strategies? Our empirical evidence suggests that the order book beyond the first step is informative–its information share is about 30%. The imbalance between demand and supply from step 2 to 10 provides additional power in explaining future short-term returns. Finally, traders do use the available information on the state of the book, not only from the first step, but also from other steps, when developing their order submission strategies.

JEL **Classification Numbers**: G10, G12, G13

Keywords: Limit order book, price discovery, order placement strategy, demand and supply schedule

*Cao and Hansch are at the Department of Finance, Smeal College of Business Administration, Penn State University, University Park, PA 16802, and Wang is at the Department of Finance, College of Business, Southern Illinois University. We are grateful to the Australian Stock Exchange and the Securities Industry Research Centre Asia-Pacific for providing the data used in this paper. We would like to thank Tarun Chordia, Ian Domowitz, Mark Lipson, Chris Muscarella, and Gideon Saar for helpful comments. Special thanks to Ian Domowitz who provided the data used in an earlier version of the paper. The authors are especially grateful to Joe Hasbrouck whose suggestions have substantially improved this paper.

The Informational Content of an Open Limit Order Book

Abstract

We assess the informational content of an open limit order book from three directions: (1) Does the limit order book allow better inferences about a security's value than simply the best bid and offer prices from the first step of the book? If it does, how much additional information can be gleaned from the book? (2) Are imbalances between the demand and supply schedules informative about future price movements? and (3) Does the shape of the limit order book impact traders' order submission strategies? Our empirical evidence suggests that the order book beyond the first step is informative–its information share is about 30%. The imbalance between demand and supply from step 2 to 10 provides additional power in explaining future short-term returns. Finally, traders do use the available information on the state of the book, not only from the first step, but also from other steps, when developing their order submission strategies.

JEL **Classification Numbers**: G10, G12, G13

Keywords: Limit order book, price discovery, order placement strategy, demand and supply schedule

A majority of stock markets around the world are organized as electronic limit order books. Among the reasons advanced for their popularity is the greater transparency offered by these systems compared to dealer market settings. A typical limit order book system allows its users to view depth at a number of price levels at and away from the market, while dealer markets usually rely on dissemination of only the dealers' best quotes. In a limit order book, all displayed prices and quantities are typically executable instantaneously, whereas in dealer markets prices for trades beyond the quoted size need to be assessed through negotiations between traders and market makers. Also, market makers may offer price improvement for trades up to quoted size.

Although limit order book trading systems have been successful around the world, little research has been done to address the value of the information contained in the order book.[1] In particular, one important question that remains unanswered is whether the demand and supply schedules as expressed in the limit orders to buy and sell contain information beyond the best bid and offer prices. Clearly, the immediate benefit of being able to view the book is the ability to assess the cost of buying or selling a certain number of shares. Unlike dealer markets, where orders may receive price improvement or have to be negotiated if they exceed quoted depth, limit order books allow almost perfect assessment of an order's price.[2] The purpose of this paper is to address three related questions. First, does the limit order book allow better inferences about a security's value than simply the best bid and offer prices? If it does, how much additional information can be gleaned from the book? Second, are imbalances between the demand and supply schedules informative about future price movements? Finally, does the shape of the limit order book impact traders' order submission strategies?

In the price discovery literature, it is a common belief that limit orders are not as informative as market orders. Well-known models of limit order books by Glosten (1994), Rock (1996), and Seppi (1997) incorporate informed traders but assume that they always use market orders. For instance, Rock's (1996) argument is that, with short-lived private information, informed traders will prefer market orders because they guarantee immediate execution and because, given the direction of price movements conditional on the private information, the execution of limit orders designed to exploit the trader's informational advantage is unlikely. For example, if the informed trader knows that the current market price is too high, the price will be headed downward in the near future, especially if other traders learn the same information. Thus, the likelihood of achieving execution for a limit sell order is remote. Similarly, Angel (1994) and Harris (1998) argue that informed traders are more likely to use market orders than liquidity traders. Our primary

[1] Exceptions include Ahn, Bae, and Chan (2001), Irvine, Benston, and Kandel (2000), and Harris and Panchapagesan (2003).

[2] One feature that may limit the accuracy of this assessment is the presence of hidden orders.

objective is to empirically assess the informational content of the limit order book.

Our research is particularly relevant considering the substantial change that U.S. stock markets have experienced in their regulatory framework and competitive landscape in the recent past. Reductions in minimum price variations, from eighths to sixteenths and subsequently to decimals, as well as the rise of Alternative Trading Systems such as Electronic Communication Networks (ECNs) are arguably the most important. The response to these changes by the traditionally dominant market places, namely the New York Stock Exchange and Nasdaq, has been varied.

The NYSE faced pressures from institutional traders who, in a post-decimals world, were concerned about their inability to accurately assess liquidity based on the best bid and offer alone. On January 24, 2002, in an effort to increase the pre-trade transparency of its market, the NYSE began to publish aggregate depths at all price levels on either side of the specialists' books under what is known as the NYSE OpenBook program. Currently the NYSE is seeking to launch another program, called LiquidityQuote, which will display a bid and offer price, potentially different from the best quotes in the market, specifically aimed at block traders.

Nasdaq, responding to the same pressures faced by the NYSE and in an effort to emulate the transparency and speed of its new competitors, introduced its new trading system, SuperMontage, during the fourth quarter of 2002. In a departure from a system dominated by competing market makers, SuperMontage's design closely follows the blueprint of an electronic limit order book, albeit with some exceptions to an overall price/time priority rule for orders. Nevertheless, the immediate result of this change has been a tremendous increase in traders' ability to monitor and assess liquidity in the Nasdaq market. The SuperMontage displays aggregate depths at the five best price levels on either side, and offers a "scan" function which allows traders to assess liquidity even further along the book.[3] In summary, traders in the U.S. stock markets find themselves endowed with a large amount of market data previously inaccessible to most of them.

This paper is part of an emerging literature about the informational content of the limit order book. We are particularly interested in the *incremental* informational content of the order book over and above the information traditionally available in the U.S. and other markets, i.e., the best bid and offer quotes along with their respective depths.[4] Using the order book information from the Australian Stock Exchange, we examine the limit order book from three principal directions. First, we assess the value of order book information in determining the true value of the stock

[3] Of course, the ECNs have long offered users a much more comprehensive view of the available liquidity in their systems. Most notably, the Island ECN has been publishing the top of its order books over the Internet. Many market observers have attributed part of the ECNs success to their greater transparency.

[4] In the U.S., the National Best Bid and Offer (NBBO) has long been the most widely available piece of information, and regulatory rules often refer to the NBBO as a reference price pair.

by examining the lead-lag relationship between WP^1 and WP^{1-10}, where WP^1 and WP^{1-10} are quantity-weighted prices using order book information in step one and from step 1 to 10 of the order book. We find that the order book beyond the first step is informative for a significant fraction of our sample firms. Estimates of Hasbrouck (1995) information share suggest that the contribution of the order book beyond the first step is about 30%.

Next, we ask the related question of whether order book information is associated with future returns. In a regression framework, we investigate the relation between returns and the lagged imbalance between supply and demand as expressed in the book. Our results show that the imbalance information derived from step 2 to 10 is useful and provides additional explanatory power.

Finally, we examine whether and to what extent the state of the order book impacts traders' order submission strategies. The theoretical literature offers some testable hypotheses regarding the influence of the best bid and offer in the book on order choice which have been confirmed in recent empirical work (Ellul et al. (2002)). As before, our focus is on the additional impact of the order book over and above the best bid and offer. Thus, we go further than previous work in that our analysis not only considers the best bid and offer but extends along the book.

Our empirical evidence is in contrast to the predictions of theoretical models of Angel (1994), Glosten (1994), Rock (1996), and Seppi (1997). However, it is consistent with recent findings of Bloomfield, O'Hara, and Saar (2003) who investigate the use of limit and market orders by informed and uninformed traders in a laboratory market setting. They find that informed traders make extensive use of limit orders, especially once most of the uncertainty around the true value of the security has been resolved.

The paper is organized as follows. Section 1 describes the institutional environment of the Australian Stock Exchange and the sample data. Section 2 empirically assesses the informational content of the limit order book using the error correction model approach. In Section 3, we examine the information share of the first step and the rest step of the order book. Section 4 provides evidence of the association between returns and lagged imbalances between demand and supply. Section 5 examines the relation between order book information and order submission strategies. Concluding remarks are provided in Section 6.

1 The Australian Stock Exchange and Sample Data

The data used in this paper was provided by the Australian Stock Exchange (ASX) via the Securities Industry Research Centre Asia-Pacific (SIRCA).[5] The ASX uses the fully computerized

[5]We would like to thank both SIRCA and the ASX for the provision of the data.

Stock Exchange Automated Trading System (SEATS) which is modeled on Toronto's Computer Assisted Trading System (CATS) popular around the world. Although limit order book data is now available from a number of sources including the New York Stock Exchange's OpenBook program, Nasdaq's SuperMontage, Island ECN ITCH datafeed, Paris Bourse, etc., there are several reasons why ASX data is particularly suitable for the purposes of our study.

First, there are no dealers or other designated liquidity providers on the ASX, so that most public orders can interact with each other directly. This is in contrast to the New York Stock Exchange where the specialist continues to play an important role in the provision of liquidity and maintenance of market integrity, and the Paris Bourse, which has a number of market makers assigned to its stocks. In the case of NYSE's OpenBook, the data consists of aggregate depths at each price level, whereas the ASX data contains detailed information about each order. Further, the prices and depths shown by OpenBook are not actually executable – they mostly serve (pre-trade) informational purposes.

Second, using data from SuperMontage or from ECNs such as Instinet or Island would be problematic for our study because the market for Nasdaq-listed stocks is highly fragmented. SuperMontage is estimated to account for only 20% of volume in Nasdaq stocks as of December 2002. Instinet and Island, whose combined market share is quite large, maintain separate books despite their recent merger. Consequently, there is no single, centralized limit order book for Nasdaq stocks as there is for ASX stocks.

1.1 Institutional Details of the ASX Trading System

Each day, the market goes through several stages. During the pre-opening period from 7:00AM to 10:00AM Eastern Standard Time orders can be entered into the system but no matching takes place. The ASX opens at 10:00AM with a procedure aimed principally at maximizing traded volume at the chosen opening price. Stocks open sequentially in five groups, based on the alphabetical order of the ticker symbol, and begin normal trading right after the conclusion of the opening algorithm for their group. This phase, during which the vast majority of trading takes place, lasts until 4:00PM. Orders entered during normal trading hours are matched, resulting in trades, or stored in the order book automatically. At 4:00PM, a 5-minute period of so-called "pre-close" begins which is followed by the official single-price closing auction. Afterwards, the system accepts new orders until 5:00PM but does not match them. Interested parties can engage in so-called late-trading, still governed by price/time priority, until 7:00PM. Following this, traders can engage in overnight trading until 7:00AM. During this time, the SEATS system is unavailable and trades go through at mutually agreed prices.

Limit and market orders are entered into the system by dual-capacity brokers. The order book enforces strict price/time priority. Orders stored in the book can be amended by either changing the limit price or reducing the number of shares offered or demanded. While a size reduction does not affect the order's priority, a price-changed order has to join the back of the queue of orders at the new price level. There is no size priority rule as, for example, in Nasdaq's SuperMontage. The only exception to the priority rules are two types of crossings. Special crossings require a minimum consideration of between AUS$1m and AUS$5m and can take place at any time during the day. The market only becomes aware of a special crossing after it has occurred. On-SEATS crossings take place within the system. The crossing broker needs to enter a limit order at the desired price (which gains time but not price priority) into the system which is one price step or less away from the best price on the other side of the market and then executes the opposing side of the crossing order against herself. Limit orders for at least AUS$100,000 can include an undisclosed reserve size which replenishes the visible order size upon partial execution.

Order book information 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 aggregate depth at multiple prices. The former, more detailed format is usually available to brokers, while the latter is representative of what on-line traders would be able to see.

1.2 Data

We obtained data covering the 21 constituent stocks of the ASX-20 index for the month of March 2000 from the ASX Intra-Day data set which provides historical details of all individual orders placed on SEATS as well as any 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 the re-construction of the limit order book for each stock at any point during the sample period. Since our subject is the information content of the limit order book, we focus on normal trading. To avoid confounding effects from the opening procedure, we further restrict our attention to the period from 10:15AM to 4:00PM. The details about the construction of the limit order book are provided in Appendix A.

Table 1 presents summary statistics for the sample stocks. They are among the largest on the ASX, with an average market value of AUS$17.8 billion as of December 31, 1999. They also represent a variety of industry sectors, including Banks, Energy, Insurance, Materials, Media, Real Estate, and Telecommunication Services. The average share price is AUS$14.6. All ASX-20 index constituent stocks are actively traded. The average number of trades is 619 trades per day and the average trade size is 3,596 shares. The daily average dollar and share volume are AUS$32 million and 2.2 million shares, respectively. It is noted that the daily average number of trades is 65% of the number of orders placed and trade size is about 60% of the order size.

2 How Informative is the Order Book?

2.1 Summary Statistics of Order Book Shape

At any point in time the limit order book contains a large number of different buy and sell orders. In order to make this order book data amenable to empirical analysis, we introduce two summary measures which are designed to capture the most important features of an order book. In actual limit order markets, the demand and supply side of the book can be represented as step functions. Thus, we propose to summarize the shape of the order book using the notion of *steps*:

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

Building on these definitions of step height and length it is straightforward to define the cumulative height and length of (part of) each side of the book. Specifically, height and length of the demand side of the book up to step n can be expressed as $H_n^d = \sum_{i=1}^n \Delta_i^d$ and $L_n^d = \sum_{i=1}^n Q_i^d$, respectively.

Table 2 reports the cross-sectional average of heights and lengths up to step ten. The order book is quite long with a mean (median) cumulative number of shares of 106,609 (74,291) and 121,236 (82.237) for the buy and sell side, respectively. Since the average order size is 6,085 shares and the average trade size is 3,596 shares, it appears that the first ten steps in the order book will capture most of the relevant activity in the book. Consequently, we will truncate the reconstructed order book at ten price steps away from the market for all of the subsequent analyses.

For both buys and sells, steps close to the market are generally longer (offer more depth) and lower (price increments are smaller) than those far away. The first three steps in the book each represent 12-13% of the respective side's length, while the last two only account for 7-8%. The height of the first step is 1.5 cents on both sides, while that for step 10 is 9.1 (5.5) cents on the demand (supply) side. The shape of the book appears to be asymmetric, with the number of shares quoted on the sell side being significantly greater than that on the buy side. In contrast, the height of the book on the demand side is significantly larger.

We measure the change in the shape of the book in step 1 and from step n_1 to n_2 by the change in the respective weighted prices, WP^1 and $WP^{n_1-n_2}$, which are defined as:

$$\begin{aligned} WP^1 &= \frac{Q_1^d P_1^d + Q_1^s P_1^s}{Q_1^d + Q_1^s}, \\ WP^{n_1-n_2} &= \frac{\sum_{j=n_1}^{n_2}(Q_j^d P_j^d + Q_j^s P_j^s)}{\sum_{j=n_1}^{n_2}(Q_j^d + Q_j^s)}, \qquad n_1 < n_2. \end{aligned}$$

WP^1 will change if the height or the length of the first step of the book changes, while $WP^{n_1-n_2}$ changes if any of the steps between n_1 and n_2 experience a change in either height, length, or a combination of both. WP^1 can be computed using only information that historically has been available to traders in the U.S. equity markets, namely the best bid and offer and their respective depths. To compute $WP^{n_1-n_2}$ information from the order book is necessary. Our primary interests are WP^1, WP^{1-10} and WP^{2-10}, where WP^{2-10} summarizes the marginal information in the book beyond that of step 1.

Table 3 reports the cross-sectional distribution of the fraction of intervals during which there are changes in book shape as measured by WP^1 and WP^{2-10} for 1-minute, 5-minute, 10-minute and 15-minute intervals. For minute-by-minute snapshots, WP^1 (WP^{2-10}) changes 70% (52%) of the time, on average. For five-minute intervals, this figure increases to 97% (89%), while the median shows an increase to 98% (90%). Further increasing the interval length to 15 minutes reduces the average number of no-change intervals to essentially zero. The cross sectional distributions of order submissions and the number of trades show that, on average, 14 new orders arrive and 9 trades occur during each 5-minute interval. The most active stock has twice the average number of orders and trades, while the least active stock has roughly a third.

In choosing the appropriate time interval for our subsequent work we need to compromise between the desire to have a large number of observations and the need to allow the order book to experience a meaningful change between subsequent observations. Based on the evidence in Table 3, it appears that 5-minute intervals best strike that balance and we will thus use them as the base for our analyses. Nonetheless, to ensure that our findings are robust with respect to the choice of the interval length, we replicate all analyses using observations sampled every 10 minutes. The results from the 10-minute sample are qualitatively similar and are available from the authors upon request.

2.2 Order Book Information and Stock Values

To address the question of whether the order book provides valuable information, we compare the informational content of WP^1 and WP^{1-10}. Both can be thought of as proxies for the true value of the underlying stock. If the order book beyond the first step contains value relevant information WP^{1-10} can be expected to be a better indicator of value than WP^1. Alternatively, if the book does not contain such information, introduces noise, or reacts sluggishly to new information because of stale limit orders in the book, WP^1 will be more informative. Note that while WP^1 is always between the best bid and offer in the book, WP^{1-10} can take values below (above) the best bid (offer). In this section, we first perform unit root and cointegration tests for each stock and then employ an Error Correction Model (ECM) to assess the contribution of each price series to price discovery. The adoption of this methodology is motivated by recent work of Hasbrouck (1995), Harris et al. (1995), Huang (2002), and Eun and Sabherwal (2003), who examine the information content of prices of the same security observed in different markets. Our analysis focuses on prices of a stock estimated from different steps of the order book.

2.2.1 Unit Root and Cointegration Tests

Since the true value V of a stock is not observable, we cannot directly compute the distance between WP^1 (WP^{1-10}) and V. However, as long as we believe that prices converge to true value, we can compare WP^1 and WP^{1-10} and examine whether there exists a lead-lag relationship between them. The lead-price converges faster and is therefore closer to the true value. A common approach to estimating such relationships is to apply the Granger Representation Theorem in an error correction model setting, where cointegrated variables can be represented as an ECM.

We begin by investigating whether WP^1 and WP^{1-10} are non-stationary and cointegrated. Assume that x_t is a bivariate vector time series. The components of x_t are cointegrated of order one if each of the series taken individually is I(1), that is non-stationary with a unit root, while a linear combination of the series is stationary.

For each stock and each time series of WP^1 and WP^{1-10}, we apply the Augmented Dickey-Fuller (ADF) unit root test. The null hypothesis is that each price series contains a unit root and is thus non-stationary. Three alternative hypotheses are considered: (1) zero-mean stationarity; (2) stationarity with a non-zero mean; and (3) linear trend stationarity. Given the space constraint, we briefly summarize results without presenting a table. Our results indicate that the ADF test fails to reject the null hypothesis at the 5% level for each price series for all stocks, except for the firm WPL. These results hold regardless of the alternative hypothesis used.

Having demonstrated that both WP^1 and WP^{1-10} are non-stationary, we use Johansen's (1988) trace test to check for the presence of a cointegration vector between WP^1 and WP^{1-10} for each stock. The null hypothesis is that there are r cointegrating vectors between WP^1 and WP^{1-10}, and the alternative hypothesis is that there are more than r cointegrating vectors. In our implementation, r starts from zero and then increases to 1. Results from the test reject the null of no cointegration as the test statistic is greater than the critical value 12.21 for each firm. Further, the test fails to reject the null hypothesis of a single cointegrating vector between WP^1 and WP^{1-10}. In summary, the empirical evidence suggests that WP^1 and WP^{1-10} are cointegrated of order one. Unreported results also show that WP^1 and WP^{2-10} are cointegrated of order one.

2.2.2 Error Correction Models

We now turn to the analysis of the lead-lag relationship between WP^1 and WP^{1-10}. Our approach utilizes the error correction models that have been used extensively in the price discovery literature on information content of prices of the same security observed in different markets. Examples include, but are not limited to, Hasbrouck (1995, 2002), Harris et al. (1995), Huang (2002), and Eun and Sabherwal (2003).

Before we turn to our specification of the ECM, a brief discussion of the economic forces behind the cointegration relationship between WP^1 and WP^{1-10} is in order. The above-mentioned papers typically invoke arbitrage arguments to explain why prices for a single security observed in two (or more) different markets cannot diverge substantially in the long run. For example, if the market prices for a stock on the New York Stock Exchange and the Boston Stock Exchange diverged, one would expect arbitrageurs to buy shares at the lower price in one market and sell shares at the higher price in the other market, thereby exerting upward (downward) price pressure in the lower (higher)-priced market, and ultimately reducing the price differential to within transaction cost bounds.

In our setting, no such arbitrage opportunities exist, since neither WP^1 nor WP^{1-10} represent tradeable prices. However, we argue that a non-trivial divergence between WP^1 and WP^{1-10} induces actions by traders which tend to lead to a reduction in that difference. Suppose, for example, that $WP^1 \gg WP^{1-10}$, i.e., the lengths or heights (or both) of the steps on the demand side of the book are much greater than those on the supply side. Also, for the moment, ignore possible changes to the first step on either side of the book, i.e., assume that WP^1 remains unchanged. In this case, buyers have incentives to place their limit orders aggressively, thus increasing the relative weights of price steps closer to the market and increasing WP^{1-10}. Turning to the sellers, observe that any limit sell order entered into the book will have the effect of increasing WP^{1-10}. Thus, limit orders submitted beyond the first step are expected to lead to a reduction in the difference between WP^1 and WP^{1-10} through an increase in the latter, i.e., WP^{1-10} adjusts towards WP^1.

Now, consider limit orders submitted *at* the best bid. These orders will not only increase WP^{1-10} but also decrease WP^1, as the best bid receives a relatively larger weight than the best offer, i.e., in this case, both WP^1 and WP^{1-10} adjust to reduce the divergence between them. Limit orders submitted at the best offer price will increase both WP^1 and WP^{1-10}, with a stronger effect on WP^1, and will thus further widen the discrepancy between the two prices. However, recall that WP^1 is bounded from above by P_1^s, so that WP^1 cannot increase too much. Thus, overall, the divergence should tend to be reduced by traders' actions.

Another case to consider is limit orders which *improve* the current best bid or offer. Now, the effects are reversed, as limit orders to sell reduce the difference between the two prices by decreasing WP^1 and increasing WP^{1-10} simultaneously, while limit orders to buy will increase both prices. As before, recall that WP^1 is bounded from above by P_1^s, so that the potential for further divergence is clearly limited.

Finally, consider market orders. These will always reduce the length of the first step in the book and will thus always have a greater impact on WP^1 than WP^{1-10}. Market buys will reduce the weight given to the best ask price, and thus reduce both WP^1 and WP^{1-10}, implying a reduction in the divergence, while market sells will have the opposite effect. However, given the book's assumed state, sellers have less incentives to use market orders than buyers, so that overall we would expect to observe a reduction in the divergence.

In summary, we can conclude that economic forces together with the simple mechanics of the book will ensure that our proxies for the true value of the stock, WP^1 and WP^{1-10}, will not diverge over time. Based on this rationale for cointegration, we specify the error correction models as follows:

$$\Delta WP_t^1 = \alpha_{1,0} - \alpha_1(WP_{t-1}^1 - \beta\, WP_{t-1}^{1-10}) + \sum_{i=1}^{p}(\gamma_{1,i}\Delta WP_{t-i}^1 + \delta_{1,i}\Delta WP_{t-i}^{1-10}) + \eta_{1,t} \quad (1)$$

$$\Delta WP_t^{1-10} = \alpha_{2,0} + \alpha_2(WP_{t-1}^1 - \beta\, WP_{t-1}^{1-10}) + \sum_{i=1}^{p}(\gamma_{2,i}\Delta WP_{t-i}^1 + \delta_{2,i}\Delta WP_{t-i}^{1-10}) + \eta_{2,t}, \quad (2)$$

where α_1 and α_2 are error correction coefficients, β reflects the long-run equilibrium relation between WP^1 and WP^{1-10}, and $WP_{t-1}^1 - \beta\, WP_{t-1}^{1-10}$ denotes the error correction term. We expect β to be unity for two reasons: (1) WP^1 and WP^{1-10} are alternative proxies for the true value and both are derived from the same order book for a given stock; (2) WP^1 and WP^{1-10} are bounded not to deviate from each other by too much as we argued above. The terms $\sum_{i=1}^{p}(\gamma_{k,i}\Delta WP_{t-i}^1 + \delta_{k,i}\Delta WP_{t-i}^{1-10})$, k=1,2, reflect short-term dynamics between WP^1 and WP^{1-10}, with p being the appropriate number of lags as determined by the Akaike information criterion. In the specification of the ECM, we normalize the cointegration vector and set the coefficient in front of WP^1 to be one.

We have explained above why the two series can be expected to be cointegrated. We now focus on the question of whether WP^1 tends to adjust towards WP^{1-10} or vice versa. On the one hand, WP^{1-10} aggregates information from all ten steps along the book, and provides a more comprehensive picture of demand and supply than WP^1. Therefore, WP^1 is expected to adjust to WP^{1-10} and correct the deviation. On the other hand, WP^1 represents the most aggressively priced orders which may be expected to capture most if not all relevant information. For this reason, one may also expect WP^{1-10} to adjust to WP^1.

In terms of our ECM specification in Eq. (1)-(2), which allows both WP^1 and WP^{1-10} to respond to deviations from the long-term equilibrium relation between the two price series, we can find the answer by examining the coefficients α_1 and α_2. Both coefficients are expected to be non-negative, i.e., the two price series do not diverge in the long-run as per our arguments above. Following Harris et al. (1995), we categorize each stock to one of the four cases based on the significance of coefficients α_1 and α_2:

- Case 1: α_1 is insignificant, but α_2 significant $\Longrightarrow$ WP^1 leads WP^{1-10};
- Case 2: α_1 is significant, but α_2 insignificant $\Longrightarrow$ WP^{1-10} leads WP^1;
- Case 3: Both α_1 and α_2 are significant $\Longrightarrow$ both WP^1 and WP^{1-10} adjust towards long-term equilibrium; and
- Case 4: Neither α_1 nor α_2 is significant $\Longrightarrow$ neither WP^1 nor WP^{1-10} adjust towards long-term equilibrium.

Classifying a stock as belonging to case 1 means that WP^1 does not respond to the divergence, while WP^{1-10} does. Thus, WP^1 leads WP^{1-10}, and is concluded to be a better approximation of the true value. The reverse is true if a stock belongs to case 2. Overall, based on the frequency with which we observe each case, we can determine whether the order book beyond the first step contains relevant information for estimating the true value of the stock.

We estimate the error correction model using the maximum likelihood method for each stock. Table 4 presents our estimates of β, α_1, and α_2. The reported estimates are quite informative about the internal working of the ECM and the dynamic interaction between WP^1 and WP^{1-10}, and we can make several observations. First, as expected, the estimated β is very close to unity, with an average value of 0.999, and not significantly different from unity for all sample firms. This result implies that the normalized cointegration vector is $(1, -1)'$, and that the long-term equilibrium relation between WP^1 and WP^{1-10} is indeed their difference. Another conclusion we can draw is that, as we argued above, an imbalance in the order book will encourage traders to submit orders in such a way as to reduce the imbalance, and thus in the long-run, WP^1 and WP^{1-10} are fairly close to each other.

Second, the estimated α_1 and α_2 have the expected signs and are positive for all firms. Among the 21 sample firms, α_1 is significant at the 5% level for only 7 firms. In contrast, α_2 is significant for all but one firm (CCL). To appreciate these estimates, recall that α_1 and α_2 reveal whether WP^1 adjusts to the departure from WP^{1-10} or visa versa. The estimation results imply that WP^1 does not adjust to WP^{1-10} for a majority of sample firms, while WP^{1-10} does adjust to WP^1 for virtually all firms. Using a 5% significance level, we compare the significance of α_1 and α_2 and assign each stock to one of the four cases above: Fourteen stocks belong to case 1, six stocks belong to case 3 and one belongs to case 2. For about two-thirds of the stocks, WP^{1-10} responds to the divergence between WP^1 and WP^{1-10}, while WP^1 does not. This finding provides evidence that WP^1 generally leads WP^{1-10} and appears to be a better indicator of the underlying true value. On the other hand, coefficients α_1 and α_2 are jointly significant for one-third of sample firms, indicating that the order book beyond the first step is informative for a significant fraction of our sample firms.

3 Information Share of the Order Book

We have shown that, although WP^1 generally leads WP^{1-10}, the book beyond the first step is somewhat informative. The natural follow-up question is: What is the marginal contribution of the prices and quantities posted beyond the first step towards price discovery? Or, in other words, how much information does the book beyond the first step provide about the true value of the stock? Since WP^{1-10} aggregates information in all ten steps, including information in step one, it is logical to separate the first step of the book from the other steps in order to answer the above question. In this section, we employ the methods developed by Hasbrouck (1995) and Gonzalo and Granger (1995) to assess the information content of WP^1 and WP^{2-10} by estimating their respective information shares.

The Hasbrouck information share and Gonzalo-Granger common factor methods explore the mechanics of price discovery process between markets where the same (or near-identical) security is traded. These methods share the same price dynamics–the error correction model. The difference between them is the definition of price discovery. Since our objective is to evaluate the informativeness of each price, WP^1 and WP^{2-10}, we are concerned about the robustness of the results with respect to the chosen methodology. Therefore, while we have selected Hasbrouck's method for the primary analysis, we use the Gonzalo-Granger method to cross-validate our results. The next section briefly reviews the Hasbrouck and Gonzalo-Granger methods.

3.1 The Hasbrouck and Gonzalo-Granger Methods

The literature on the Hasbrouck information share and Gonzalo-Granger common factor method is extensive. For detailed discussion, we refer readers to Hasbrouck (1995, 2002), Baillie et al. (2002), Booth et al. (2002), Gonzalo and Granger (1995), Harris et al. (2002), Huang (2002), and Lehmann (2002). In contrast to the usual multiple markets setup, we consider two prices, WP^1 and WP^{2-10} derived from the same order book. Specifically, consider the following error correction model:

$$\Delta WP_t^1 = \alpha_{1,0} - \alpha_1(WP_{t-1}^1 - \beta\, WP_{t-1}^{2-10}) + \sum_{i=1}^{p}(\gamma_{1,i}\Delta WP_{t-i}^1 + \delta_{1,i}\Delta WP_{t-i}^{2-10}) + \eta_{1,t} \quad (3)$$

$$\Delta WP_t^{2-10} = \alpha_{2,0} + \alpha_2(WP_{t-1}^1 - \beta\, WP_{t-1}^{2-10}) + \sum_{i=1}^{p}(\gamma_{2,i}\Delta WP_{t-i}^1 + \delta_{2,i}\Delta WP_{t-i}^{2-10}) + \eta_{2,t}, \quad (4)$$

where each error term has a zero-mean and is serially uncorrelated. Let ρ denote the correlation between $\eta_{1,t}$ and $\eta_{2,t}$. The covariance matrix of the error terms is:

$$\Omega = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

The Hasbrouck information share builds on the decomposition of the variance of the common factor innovations, where the common factor is the permanent impact of new information on the stock price. In our context, the information share of WP^1 (or, WP^{2-10}) is its proportional variance to the variance of the common factor innovations. If the two errors are uncorrelated (i.e., $\rho = 0$), the matrix Ω is diagonal, and the information shares of WP^1 and WP^{2-10} are:

$$S_1 = \frac{\alpha_1^2\sigma_1^2}{\alpha_1^2\sigma_1^2 + \alpha_2^2\sigma_2^2} \quad (5)$$

$$S_2 = \frac{\alpha_2^2\sigma_2^2}{\alpha_1^2\sigma_1^2 + \alpha_2^2\sigma_2^2}. \quad (6)$$

Given that WP^1 and WP^{2-10} contain order book information from step one and step 2-10, respectively, and that the book information in various steps is highly correlated, the correlation between $\eta_{1,t}$ and $\eta_{2,t}$ is likely non-negligible. In the presence of contemporaneous correlation, Hasbrouck (1995) proposes to account for this correlation by finding the Cholesky factorization of Ω, with $\Omega = FF'$ and

$$F = \begin{pmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sigma_2(1-\rho^2)^{1/2} \end{pmatrix}.$$

Armed with the lower triangular matrix F, we can orthogonalize the correlated error terms and obtain the information shares of WP^1 and WP^{2-10} as

$$S_1 = \frac{(\alpha_1\sigma_1 + \alpha_2\sigma_2\rho)^2}{(\alpha_1\sigma_1 + \alpha_2\sigma_2\rho)^2 + \alpha_2^2\sigma_2^2(1-\rho^2)} \quad (7)$$

$$S_2 = \frac{(\alpha_2^2\sigma_2^2(1-\rho^2)}{(\alpha_1\sigma_1 + \alpha_2\sigma_2\rho)^2 + \alpha_2^2\sigma_2^2(1-\rho)}. \quad (8)$$

It is noted that S_1 and S_2 depend crucially on the ordering of the price series in the factorization. Further, the first price in the ordering corresponds to the maximum information share while the second price corresponds to the minimum. The information shares given in Eq. (7)-(8) are based on the assumption that WP^1 is the first and WP^{2-10} the second price in the factorization. By rotating the sequence of WP^1 and WP^{2-10}, we obtain the lower and upper bounds of the information share for each price. The difference in lower and upper bounds might be substantial if the correlation coefficient ρ is large. In this situation, we use the midpoint of the lower and upper bounds as a sensible estimate of the information share. Finally, given that the distributions and test statistics of these bounds are difficult to obtain, we follow Hasbrouck (1995), Booth et al. (2002) and Huang (2002) to assess the statistical significance of the bounds by using the cross-sectional standard errors.

In contrast to Hasbrouck's method, the Gonzalo and Granger approach defines the common factor as $s_1 WP^1 + s_2 WP^{2-10}$, which is a linear combination of WP^1 and WP^{2-10}. The common factor weights s_1 and s_2 denote the respective contribution of WP^1 and WP^{2-10} to the efficient price:

$$s_1 = \frac{\alpha_2}{\alpha_1 + \alpha_2} \quad (9)$$

$$s_2 = \frac{\alpha_1}{\alpha_1 + \alpha_2}. \quad (10)$$

As such, the Gonzalo and Granger method relies solely on the error correction coefficients.

3.2 Results

For each stock, we estimate the error correction model given in Eq. (3)-(4) and obtain the maximum and minimum of the Hasbrouck information share and the Gonzalo and Granger common factor weights for WP^1 and WP^{2-10}. It is worthwhile to point out that the estimated error correction coefficients α_1 and α_2 are different from those in Table 4 as we have replaced WP^{1-10} with WP^{2-10}. Table 5 reports the estimation results and summary statistics for each price contribution statistic. Take WP^1 as an example, the average of the maximum and minimum information shares are 0.93 and 0.47. The corresponding statistics for WP^{2-10} are 0.52 and 0.07. The t-test and non-parametric sign test show that the maximum information share for WP^1 (0.93) is significantly larger than that for WP^{2-10} (0.52), with a t-statistic and p-value being 6.58 and 0.00. The large deviation between the maximum and minimum information share indicates a substantial contemporaneous correlation between innovations in ΔWP_t^1 and ΔWP_t^{2-10}.

According to Baillie et al. (2002) and Booth et al. (2002), in the presence of contemporaneous correlation, a sensible approach to extract one point estimate of the information share is to use the average of the maximum and minimum as an approximation to the center of the information share distribution. The cross-sectional average of the maximum-minimum midpoint is 0.70 for WP^1 and 0.30 for WP^{2-10}, hence, the information share of WP^{2-10} is material.

Next we examine the Gonzalo and Granger common factor weights reported under the column heading "G-G" in Table 5. Cross-sectionally, the common factor weight assigned to WP^1 is four times as large as that assigned to WP^{2-10} (0.79 versus 0.21). The cross-sectional average of G-G estimates, 0.79 versus 0.21, are broadly in line with estimates of the Hasbrouck information share, 0.70 versus 0.30.

Using the G-G estimates, we perform two specification tests: (1) $H_0 : s_1 = 1, s_2 = 0$, and (2) $H_0 : s_1 = 0, s_2 = 1$. The first hypothesis states that the underlying true value is determined completely by the information in the first step of the book, and the rest of the order book is uninformative. The implication of the second hypothesis is exactly opposite: the first step of the book is useless while the rest steps of the book is informative. We use unrestricted and restricted estimation results to compute the likelihood ratio test statistic. This statistic follows a χ^2 distribution with one degree of freedom. The p-values of the test for each firm are presented in Table 5. For the first hypothesis, we reject the null for seven firms at 5% significance level and for ten firms at 10% level. Our specification test rejects the second hypothesis for all firms.

Recall that the Hasbrouck information share and Gonzalo and Granger common factor weights differ in two important aspects: (1) the Hasbrouck approach decomposes the variance of innovations in the common factor; while the G-G approach decomposes the common factor; (2) the Hasbrouck method takes the contemporaneous correlation into account while the G-G method does not. Although the two approaches are different in principle, we reach a qualitatively similar conclusion–the order book information beyond the first step is important in determining the true value of the underlying stock.

4 Relation of Returns and Order Book Information

So far, the results of the error correction model show that WP^1 often leads WP^{1-10} and thus can be thought of as a better, more informative indicator of a stock's true value. In this section, we investigate whether information on the shape of the order book is associated with future short-term returns.

If the market is efficient, the publicly available order book information will be orthogonal to future returns. However, there is ample evidence of short-term predictability in stock returns. It is conceivable that some information contained in the order book, such as the imbalance between demand and supply schedules, can be used in forecasting short-term returns. For dealer markets, the theoretical model of Cao and Lyons (2000) shows that a dealer's superior knowledge of market demand and supply conditions helps them forecast price. However, in their study, the information on order flows is private, while the order book information is public. Further, Kavajecz and Odders-White (2002) show that technical analysis indicators are related to the liquidity supply in a limit order market. The question we are interested in is whether the public information on the book is associated with future returns in the short-run.

To assess the informativeness of the order book with respect to future returns, we first examine the relation between five-minute returns and lagged variables constructed from the demand and supply schedule. Because short-term returns are serial correlated over time, we pre-whiten returns so that we can focus on the innovation in returns. Experimenting with various specifications, we find that the AR(5) model is sufficient to eliminate predictable component of returns for all firms. Innovations in returns are from the following time-series model:

$$r_t = \alpha_0 + \sum_{i=1}^{5} \alpha_i r_{t-i} + \epsilon_t^* \tag{11}$$

where r_t is the 5-minute return at t, and ϵ_t^* the return innovation. For simplicity, the subscript for each firm is suppressed. To ensure that the variables are comparable across firms, all innovations are normalized by the mean and standard deviation of that series. Observations from sample firms are then pooled together prior to estimation.

The independent variables include the inside spread, and the imbalance in the length and height of the order book from step 1 to 10. For each step j, let QR_j and HR_j denote the imbalance in length and height defined as:

$$\begin{aligned} QR_j &= \frac{Q_j^s - Q_j^d}{Q_j^s + Q_j^d}, \qquad j = 1, \ldots, 10 \\ HR_j &= \frac{(P_j^s - P_{j-1}^s) - (P_j^d - P_{j-1}^d)}{(P_j^s - P_{j-1}^s) + (P_j^d - P_{j-1}^d)}, \qquad j = 2, \ldots, 10 \end{aligned}$$

Intuitively, QR and HR can be interpreted as scaled imbalance in quantity and price between the supply and demand. Standardizing these variables guarantees the consistency of the variable across firms. We consider ten regressions, each includes the book information up to step n where

$n = 1, \ldots, 10$. The nth estimated model is:

$$\epsilon_t \quad = \quad \alpha_0 + \alpha_1 Spread_{t-1} + \beta_1 QR_{1,t-1} + \sum_{j=2}^{n} \beta_j \, QR_{j,t-1} + \sum_{j=2}^{n} \gamma_j \, HR_{j,t-1} + \eta_t, \qquad (12)$$

where ϵ_t is the standardized innovation at t in 5-minute return obtained from ϵ_t^*, and $Spread_t$ the inside relative spread at t on a firm-by-firm basis. Our objective is to check whether the book beyond the first step helps to predict future returns. We report the adjusted R^2 of each regression and see if it increases as more steps of the book are included in the regression.

In addition to the imbalance in book length and height, we also consider using liquidity measures as independent variables. Given the fact that the average trade size varies across firms, our liquidity measure is defined for a given trade size of q shares, where q is a multiple of the average trade size. Each stock, let $\overline{Q}$ denote the average trade size. For a given trade size of q ($q = 1.0\overline{Q}, \ldots, 5.0\overline{Q}$), let $LD(q)$ be a liquidity measure on the demand side, defined as the discount per share below the midpoint of the best bid and ask:

$$LD(q) \quad = \quad 0.5(P_1^d + P_1^s) - \frac{\sum_{j=1}^{m_1-1} P_j^d \, Q_j^d + P_{m_1}^d \, Q_{m_1}'^d}{q}, \qquad \sum_{i=1}^{m_1-1} Q_i^b + Q_{m_1}'^b = q, \qquad (13)$$

where the step m_1 is determined according to $\sum_{j=1}^{m_1-1} Q_i^d < q \leq \sum_{j=1}^{m_1} Q_i^d$. Similarly, the liquidity on the supply side $LS(q)$ is the premium per share above the midpoint of the best bid and ask:

$$LS(q) \quad = \quad \frac{\sum_{j=1}^{m_2-1} P_j^s \, Q_j^s + P_{m_2}^s \, Q_{m_2}'^s}{q} - 0.5(P_1^d + P_1^s), \qquad \sum_{i=1}^{m_2-1} Q_i^s + Q_{m_2}'^s = q. \qquad (14)$$

The step m_2 is determined according to $\sum_{j=1}^{m_2-1} Q_i^s < q \leq \sum_{j=1}^{m_2} Q_i^s$ and m_2 is not necessarily equal to m_1. The independent variable is the imbalance in liquidity, denoted by $LR(q)$:

$$LR(q) = \frac{LD(q) - LS(q)}{LD(q) + LS(q)}, \quad q = 1.0\overline{Q}, \ldots, 5.0\overline{Q}. \qquad (15)$$

This regression model is estimated for each trade size q ($q = 1.0\overline{Q}, \ldots, 5.0\overline{Q}$):

$$\epsilon_t \quad = \quad \alpha_0 + \alpha_1 Spread_{t-1} + \beta LR(q)_{t-1} + \eta_t. \qquad (16)$$

While our empirical design intends to uncover the dynamics between the return and liquidity in general, there is one potentially important question: What is the association between the return and liquidity in particular whenever there is a large asymmetry in liquidity alone the book? If the book is symmetric in quantity and price dimension, it is unlikely to contain information about future value of the stock. On the other hand, if the book is severely imbalanced, it reveals the excessive demand or supply and allows traders to better infer the true value of the stock. To investigate this question, we sort each firm's observations by the imbalance in liquidity LR(q) with q being $5\overline{Q}$. The top and bottom 5% observations from each firm are pooled together and used to re-estimate Eq. (12) and (16).

Table 6 presents results for the analysis of returns at t and order book length/height imbalance at $t-1$ (see Eq (12)). If only the information in the first step is used, the adjusted R^2 is 10.78%. Including length/height imbalance from step 2 to 10 increases the R^2 by 11% to 17% in relative terms. Initially, the R^2 increases sharply in step two, and then becomes flat after step 5. Using top and bottom 5% observations with the largest liquidity imbalance in either direction, we find a similar pattern in the adjusted R^2: It goes up by 11% from step one ($R^2 = 18.33\%$) to step two ($R^2 = 20.03\%$), and is flat beyond step 5, while the magnitude of R^2 is larger for the observations with large liquidity imbalance.

Now turning to Table 7, we report results for the analysis of returns at t and liquidity imbalance at $t-1$. The results are presented for several chosen trade size q ($q = 1.0\overline{Q}, 1.5\overline{Q}, \ldots, 5.0\overline{Q}$). Using all observations, we find that the adjusted R^2 increases gradually when the liquidity imbalance information for a larger trade size is incorporated. For example, the adjusted R^2 is 6.68% if $LR(1.0\overline{Q})$ is the independent variable. This figure increases to 9.98% when $LR(5.0\overline{Q})$ is the independent variable. Focusing on the sample of the top and bottom 5% liquidity-imbalanced observations, the R^2 also shows an increasing pattern, from 17.75% to 19.61%.

Overall, these results indicate that returns are strongly associated with the lagged order book information. Although most explanatory power comes from the information in the first step, the liquidity information derived from step 2 to 10 is useful and does provide additional explanatory power on the margin. Finally, the relation between returns and lagged order book length/height (or the lagged liquidity imbalance) is particularly strong for those observations with the largest liquidity imbalance.

5 Book Information and Order Choice

Next, we examine whether and to what extent the state of the limit order book influences traders' order submission strategies. Generally, we would expect traders to make their decision using all available information, including the state of the order book. This would be particularly true for traders who submit large orders, e.g., large or institutional traders whose liquidity demands exceed the depths offered at the very top of the book, and who would benefit from assessing the state of the book when determining their (possibly dynamic) trading strategies. On the other hand, the state of the book may have a limited effect on order submission strategies for the following reasons: (1) Monitoring and analyzing the book is prohibitively costly; (2) traders believe that limit orders outside the best bid and offer are non-informative and simply ignore them; and (3) traders are overconfident and give little weight to order book information relatively to their own assessment of the stock's true value.

Theoretical work on order choice due to Cohen et al. (1981), Harris (1998), and Foucault (1999) predicts that market orders become less and limit orders become more attractive as the spread increases. Another empirical prediction, due to Parlour's (1998) model, is that depth on either side of the market has an impact on traders' choice between market and limit orders. In Parlour's model, though, market depth is modeled only for the first step of the order book, whereas we aim to go further and see whether the steps along the book have similar effects as the first. Further, more recent work by Foucault, Kadan, and Kandel (2003) predicts that limit order aggressiveness either increases or decreases with the inside spread, depending on whether patient or impatient traders dominate the trading population.

Empirical work on order submission strategies examining questions similar to ours includes Biais, Hillion, and Spatt (1995), Griffiths et al. (2000), Hollifield et al. (2003), and Ranaldo (2003). The general finding in this literature is that the rate of limit order submissions increases with the size of the spread, and that depth at the top of both sides of the book affects order choice, as predicted by the theoretical literature. A recent study similar in spirit to ours is Ellul et al. (2002) which examines limit order choice on the New York Stock Exchange. They use data from the NYSE's hybrid dealer market and focus on SuperDOT orders which capture 50% of NYSE's share volume.[6] Their evidence is broadly consistent with Parlour's (1998) predictions in that depth on either side of the book as well as the size of the spread are important. Ranaldo (2003) investigates order aggressiveness on the Swiss Stock Exchange using a model very similar to the one we propose below, albeit with a data set restricted to the best bid and offer. Our analysis differs from both Ellul et al(2002) and Ranaldo (2003) in that we use a set of complete order data from a pure limit order market, and examine the impact of the state of the book beyond the best bid and offer.

5.1 An Ordered Probit Model of Order Choice

We study the order submission strategies of traders using an ordered probit modeling approach because the different actions available to traders are inherently ordered in terms of their aggressiveness. A market order has no price limit and can therefore be considered the most aggressive type of order as their submitters are prepared to transact at any price. Limit orders are clearly less

[6]This is chiefly because order book information was not available to traders during their sample period. NYSE's OpenBook program only began in January 2002.

aggressive, as traders are only prepared to trade over a bounded range of possible prices. Among limit orders, a ranking can be established in terms of their limit price, in fact, that is precisely how the limit order book is made up. Finally, limit order cancellations can be regarded as the least aggressive, as traders who cancel their orders as they remove liquidity from the book.

Consistent with the extant theoretical literature, our approach implicitly assumes that traders arriving in the market have made the decision whether they want to be active on the buy or sell side of the market independently of the state of the book. In other words, we do not model the possibility that the decision to place an order as well as the choice of order direction, as opposed to just the order type, is influenced by traders' view of the order book. However, this potential shortcoming is irrelevant for the question we seek to address. In effect, we investigate whether the state of the book, conditional on order direction, has any impact on the type of the order. Evidence supportive of such an influence would also be supportive of the more general conclusion that the state of the book influences traders' decisions.

Specifically, we distinguish six types of actions by traders and rank them in terms of their aggressiveness. Let R_i denote the rank of action i. As discussed above, the most aggressive action is the submission of a market order or a marketable limit order which consumes liquidity on the opposite side of the market. We assign the highest rank $R_i = 6$ to this type of action. Only slightly less aggressive is the submission of a limit order which improves the current best bid or offer, and we assign rank $R_i = 5$ to this type of action. Next, we distinguish between three types of limit orders depending on the step at which the order enters the book. We assign a rank of $R_i = 4$ to an order which matches the best bid or offer, a rank of $R_i = 3$ to an order which is submitted behind the first step but below or at the third step in the book. Submissions of limit order at any price step behind the third step in the book are assigned rank $R_i = 2$. We chose the third step as a cutoff because it ensures that there is a roughly equal number of observations classified with ranks 3 and 2. Finally, the least aggressive type of trader action is the cancellation of a limit order as it removes liquidity from the book without a transaction. We assign a rank of $R_i = 1$ to this type of action.

The ordered probit model consists of two parts. The first part relates the observable action type to the latent linking variable Z_i which is continuous and whose domain is the set of real numbers:

$$R_i = \begin{cases} 6, & \text{if } Z_i \in [\mu_5, \infty), \\ 5, & \text{if } Z_i \in [\mu_4, \mu_5), \\ 4, & \text{if } Z_i \in [\mu_3, \mu_4), \\ 3, & \text{if } Z_i \in [\mu_2, \mu_3), \\ 2, & \text{if } Z_i \in [\mu_1, \mu_2), \\ 1, & \text{if } Z_i \in (-\infty, \mu_1), \end{cases}$$

and where the μ_j are partition points to be estimated. The second part of the model relates the latent variable to the underlying observed variables:

$$\begin{aligned} Z_i &= \beta'\mathbf{X_i} + \epsilon_i, \\ &= \beta_1 Spread + \beta_2 Q_1^{own} + \beta_3 \sum_{k=2}^{3} Q_k^{own} + \beta_4 \sum_{k=4}^{10} Q_k^{own} + \beta_5 Q_1^{other} + \beta_6 \sum_{k=2}^{3} Q_k^{other} + \beta_7 \sum_{k=4}^{10} Q_k^{other} \\ &\quad + \beta_8 |P_1^{own} - P_{10}^{own}| + \beta_9 |P_1^{other} - P_{10}^{other}| + \epsilon_i. \end{aligned}$$

The explanatory variables will be discussed below. The error terms ϵ_i's are assumed to be normally distributed across observations with mean and variance normalized to 0 and 1, respectively. With this formulation we have

$$\text{Prob}(R_i = 6) = 1 - \Phi(\mu_5 - \beta'\mathbf{X}_i), \tag{17}$$

$$\text{Prob}(R_i = j) = \Phi(\mu_j - \beta'\mathbf{X}_i) - \Phi(\mu_{j-1} - \beta'\mathbf{X}_i), \tag{18}$$

$$\text{Prob}(R_i = 1) = \Phi(\mu_1 - \beta'\mathbf{X}_i). \tag{19}$$

Estimation of the model is accomplished by the maximum likelihood method.

We include a number of order book statistics as explanatory variables in the ordered probit model. The chosen specification is designed to address the question of whether and to what extent the state of the order book beyond the first step influences traders' order choice. The first variable is the relative inside spread *Spread*. We expect that a wider inside spread leads to fewer market orders and more limit orders or cancellations. A wide spread ought to discourage traders from placing costly market orders and create incentives to place limit orders instead. We further include the number of shares offered or demanded at the first step of the book Q_1^s and Q_1^d. Together, these three variables capture the state of the book using only information available from the first step. According to Parlour (1998), we expect a longer first step on the trader's own side, e.g., the demand side for buyers, to lead to more market orders. Similarly, we expect that a longer first step on the opposite side, e.g., the supply side in the case of buyers, will lead to more limit orders.

To gauge the influence of subsequent steps in the book on trader actions, we include a number of variables designed to summarize the state of the book beyond the first step. Specifically, we include the aggregate number of shares available at steps two and three, $\sum_{i=2}^{3} Q_i^d$ and $\sum_{i=2}^{3} Q_i^s$ as well as the aggregate number of shares available between steps four and ten, $\sum_{i=4}^{10} Q_i^d$ and $\sum_{i=4}^{10} Q_i^s$. To capture the price dimension of the book we also include the cumulative heights of the book from step one to step ten, $P_1^d - P_{10}^d$, and $P_{10}^s - P_1^s$. Together, these variables capture the state of both sides of the order book beyond the first step. To make observations comparable across stocks, we scale all size and price variables by stocks' average trade size and the current midprice between the best bid and offer, respectively. Finally, all explanatory variables are standardized and observations from each firm are pooled together before the estimation.

5.2 Results

The estimation results are presented in Table 8, where we report coefficient estimates for the full sample as well as demand and supply side events separately. Several observations are in order. First, note that our results are consistent with previous empirical research of the impact of the inside spread and the depth available on the first step on either side of the book. All of the associated coefficients are significant at 5% level. The positive estimates on relative inside spread ($Spread$) imply that larger spreads reduce the likelihood of observing a market order event, since market orders are relatively costly. The negative estimates on own side depth, e.g., demand side for buy orders, at the first step of the book Q_1^{own} imply an increase in the likelihood of observing a highly-ranked event such as a market order or aggressive limit order presumably since the expected time to execution for a less aggressive order type is relatively long. Likewise, the positive signs for Q_1^{other} suggest that traders choose less aggressive actions in the expectation that traders on the opposite side of the market will be induced to choosing more aggressive order types.

Second, notice that many of the additional variables we include in order to capture the state of the order book beyond the best bid and offer are also highly significant, consistent with the notion that traders use the available information on the state of the book when developing their order submission strategies.

Now, we turn to examining the signs of coefficients and interpreting their economic meaning. Note that none of the estimates associated with the aggregate own side depth at the second and third steps are significant. We interpret this to be the result of a tension between newly arriving traders' incentives to "jump the queue" by submitting more aggressive orders, and previously active traders' incentives to cancel their existing orders and perhaps resubmit them at a later time. If aggregate depth of the second and third own side step is high, some traders will use market orders or post limit orders ahead of the crowd, while those traders with orders either in or behind the second and third step are encouraged to cancel their orders. In the estimation, both effects cancel each other out. Consistent with this explanation, aggregate own side depth four or more steps away from the market enters the model with a positive sign, i.e., if the own side book is very thick, traders respond by canceling orders. Since this variable covers six steps representing many orders, the cancellation effect dominates the "jumping the queue" effect.

The coefficients associated with depths on the opposite side of the market are close to a mirror image of those associated with own side depths. Coefficients of aggregate depths at steps 2 and 3 are positive (as for first step depth), while coefficients for aggregate depths beyond step 3 have a negative sign and are highly significant. This is consistent with the notion that long steps at the top of the opposite side induce traders to use less aggressive orders in the expectation of impatient traders on the other side, whereas long steps further away from the top make very aggressive orders from the other side less likely, thus inducing traders to be more aggressive themselves.

Turning to book height we note that both sides of the book have a significant impact on order choice. Recall that book height proxies for the potential price impact of a market order "walking up" the book. As we would expect, opposite side height, i.e., the price difference between orders on the first and tenth step, has a positive sign indicating that higher potential costs for market orders lead to a reduction in the likelihood of their use. Own side height has a negative coefficient, implying the use of more aggressive orders when the difference between the first and tenth step prices is large. This is intuitive since traders on the opposite side will be discouraged to use market orders by their high implied costs. Recognizing this, traders become more aggressive with their own orders.

6 Conclusion

In spite of the success of limit order book trading systems in the U.S. and around the world, little research has been done to assess the informational content of the order book (Jain (2002)). The academic literature has customarily used the quoted midprice 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 step of the book is largely unexplored. This article examines the value of the order book from three directions: (1) In comparison to the best bid and ask prices and corresponding depths, can we make better inference about the true value of the underlying stock using book information beyond the first step? (2) Are future short-term returns associated with imbalances between the demand and supply schedules? and (3) What is the connection between the shape of the order book and order submission strategies?

Our empirical evidence indicates that, although the information from the first step is dominating, the order book beyond the first step is informative about the true value of the underlying stock. According to our estimates of the Hasbrouck (1995) information share, the contribution of the order book beyond the first step is 30%. Using the Gonzalo and Granger (1995) common factor approach leads to a qualitatively similar conclusion. In terms of the relation between future returns and imbalances between the demand and supply, we find that imbalance information from step 2 to 10 leads to an 11-18% increase in adjusted R^2 in comparison to the result using imbalance from step one only. Finally, by examining the association between the order book information and order submission strategies, we establish that traders do use the available information on the state of the book, not only from the first step, but also from other steps, when developing their order submission strategies.

Appendix A: Reconstruction of the limit order book

The dataset provided by SIRCA is divided into self-contained, separate files for each stock and each day. There are three different types of records, discussed below. Each record is time-stamped to the nearest hundredth of a second and carries a transaction identifier which is based on the sequence of events occurring in ASX's system on a particular day.

1. Entered and deleted orders

   These records contain comprehensive information about orders entered into or deleted from the book. Each new order is assigned its record's unique transaction identifier. As advised by SIRCA, we use the order's direction, price, and position in the file to determine its rank in the book according to price/time priority. Any special order conditions, such as "fill-or-kill", "all-or-nothing", "crossing", or "hidden volume", are flagged in the order record. We do not include hidden volume in our step lengths computations, so that the reconstructed book conforms to the information set of traders at the time.

2. Trades and trade cancellations

   Trade records refer to both the buy and the sell orders which interacted in the trade by their identifiers. Special conditions, such as whether the trade was the result of a crossing or a block, are flagged in the trade record.

3. Amendments to orders

   An amendment refers to the order by the identifier. Two amendments are possible: size reductions and price changes. Only the latter affect the priority of the order. In this case, we delete the order from the book, amend its price, and subsequently reenter it into the book, basing the order's ranking on the new price and the transaction identifier associated with the amendment rather than the original identifier assigned on order arrival.

A stock's daily file usually begins with orders remaining on the book from the previous day. Those and any newly arriving orders are entered into the book, according to their priority.

References

Ahn, H.-J., Bae, K.-H., and K. Chan, 2001, Limit orders, depth and volatility: Evidence from the Stock Exchange of Hong Kong, *Journal of Finance* 56, 767-788.

Angel, J., 1994, Limit versus market orders, Working paper, Georgetown University

Baillie, R., G. Booth, Y. Tse and T. Zabotina, 2002, Price discovery and common factor models, *Journal of Financial Markets* 5, 309-321.

Biais, B., P. Hillion, and C. Spatt, 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., M. O'Hara, and G. Saar, 2003, The "Make or Take" decision in an electronic market: Evidence on the evolution of liquidity, Working paper, Cornell University

Booth, G., J. Lin, T. Martikainen, and Y. Tse, 2002, Trading and pricing in upstairs and downstairs stock markets, *Review of Financial Studies* 15, 1111-1135.

Cao, H., and R. Lyons, 2000, Inventory Information, Working paper, UC-Berkeley.

Cohen, K., S. Maier, R. Schwartz, and D. Whitcomb, 1981, Transaction costs, order placement strategy, and existence of the bid-ask spread. *Journal of Political Economy* 89, 287-305.

Ellul, A., C. Holden, P. Jain, and R. Jennings, 2002, Determinates of order choice on the New York Stock Exchange, Working paper, Indiana University.

Eun, C., and S. Sabherwal, 2003, Cross-border listings and price discovery: Evidence from U.S.-listed Canadian stocks, *Journal of Finance* 58, 549-575.

Foucault, T., 1999, Order flow composition and trading costs in a dynamic limit order market, *Journal of Financial Markets* 2, 99-134.

Foucault, T., Kadan, O., and E. Kandel, 2003, Limit order book as a market for liquidity, Working paper, HEC

Glosten, L., 1994, Is the electronic open limit order book inevitable?, *Journal of Finance* 49, 1127-1161.

Gonzalo, J., and C. Granger, 1995, Estimation of common long-memory components in cointegrated systems, *Journal of Business & Economic Statistics* 13, 27-36.

Griffiths M., B. Smith, A. Turnbull and R. White, 2000, The costs and determinants of order aggressiveness, *Journal of Financial Economics* 56, 65-88.

Harris, L., 1998, Optimal dynamic order submission strategies in some stylized trading problems, *Financial Markets, Institutions and Instruments* 7, 1-75.

Harris, L., and V. Panchapagesan, 2003, The information-content of the limit order book: evidence from NYSE specialist trading decisions, Working paper, University of Southern California.

Harris F., T. McInish, G. Shoesmith and R. Wood, 1995, Cointegration, error Correction, and price discovery on informationally-linked security markets, *Journal of Financial and Quantitative Analysis* 30, 563-579.

Harris F., T. McInish, and R. Wood, 2002, Security price adjustment across exchanges: an investigation of common factor components for Dow Stocks, *Journal of Financial Markets* 5, 277-308.

Hasbrouck, J., 1995, One security, many markets: Determining the contributions to price discovery, *Journal of Finance 50*, 1175-1199.

Hasbrouck, J., 2002, Stalking the "efficient price" in market microstructure specifications: an overview, *Journal of Financial Markets* 5, 329-339.

Hollifield, B., R. Miller, and P. Sandås, 2003, Empirical analyis of limit order markets, Working paper, Carnegie Mellon University.

Huang, R., 2002, The quality of ECN and Nasdaq market maker quotes, *Journal of Finance* 57, 1285-1319

Irvine, P., G. Benston, and E. Kandel, 2000, Liquidity beyond the inside spread: measuring and using information in the limit order book, Working paper, Emory University.

Jain, P., 2002, Institutional design and liquidity on stock exchanges around the world, Working paper, Indiana University.

Johansen, S., 1988, Statistical analysis of cointegration vectors, *Journal of Economic Dynamics and Control* 12, 231-254.

Kavajecz, K., and E. Odders-White, 2002, Technical analysis and liquidity provision, Working paper, University of Pennsylvania.

Lehmann, B., 2002, Some desiderata for the measurement of price discovery across markets, *Journal of Financial Markets* 5, 259-276.

Parlour, C., 1998, Price dynamics in limit order markets, *Review of Financial Studies* 11, 789-816.

Ranaldo, A., 2003, Order aggressiveness in limit order book markets, *Journal of Financial Markets*, forthcoming

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.

Table 1: Trading Characteristics of the Australian Stock Exchange ASX-20 Index Constituent Stocks

This table presents summary statistics by firm for the share price, inside spread, market value, number of trade, trade size, number of orders and order size for the 21 constituent stocks of the ASX-20 index. $Avg\,Price$, $Inside\,Spread$, $Market\,Value$, $Daily\,Trades$, $Trade\,Size$, $Daily\,Orders$ and $Order\,Size$ are daily averages of share price, bid-ask spread calculated using the best bid-ask quotes, number of trades, share volume per trade, number of orders and the size of order in shares, respectively. The sample period is from March 1 through March 31, 2000. $MarketValue$ is the market capitalization as of December 31, 1999. All prices are in Australian dollars.

Ticker	Company Name	Industry	Avg. Price ($)	Inside spread (cents)	Market Value ($Mil.)	Daily Trades	Trade Size (shares)	Daily Orders	Order Size (shares)
AMP	AMP Limited	Insurance	15.83	2.9	18,336	965	1,703	1,464	2,991
ANZ	AUS and NZ Banking Group Ltd.	Banks	10.57	1.7	17,374	575	3,931	930	6,685
BHP	BHP Billiton Ltd.	Materials	17.21	2.7	34,781	770	2,948	1,244	4,937
BIL	Brambles Industries Ltd.	Comm. Svc/Supplies	42.99	10.1	9,608	344	1,137	614	1,870
CBA	Commonwealth Bank of Australia	Banks	23.48	2.9	23,665	674	1,997	1,008	3,569
CCL	Coca-Cola Amatil Ltd.	Food Beverage/Tobacco	3.55	2.1	4,275	172	3,089	281	5,391
CGH	Colonial Group	Banks	7.41	1.4	6,355	639	6,230	957	10,189
CML	Coles Myer Ltd.	Food/Drug Retail	6.55	1.3	9,124	1,324	1,590	1,792	3,001
CWO	Cable & Wireless Optus	Telecom. Services	6.74	1.7	19,210	865	5,123	1,293	8,857
FBG	Foster's Group Ltd.	Food Beverage/Tobacco	4.24	1.1	7,542	401	6,109	592	10,644
LLC	Lend Lease Corp. Ltd.	Real Estate	21.12	4.8	10,804	307	1,455	520	2,335
NAB	National Australia Bank Ltd.	Banks	21.20	2.7	34,587	1,039	1,999	1,602	3,619
NCP	News Corp. Ltd.	Media	25.13	4.7	28,627	775	3,698	1,291	6,072
NCPDP	News Corp. Ltd. (Preferred)	Media	21.36	6.6	27,479	395	5,308	708	8,318
PBL	Publishing & Broadcasting Ltd.	Media	15.11	4.3	7,548	278	2,394	461	3,822
RIO	Rio Tinto Ltd.	Materials	24.01	4.5	19,227	425	2,024	710	3,344
TLS	Telstra Corp. Ltd.	Telecom. Services	7.82	1.1	53,175	1,242	6,883	1,804	12,255
WBC	Westpac Banking Corp.	Banks	10.60	1.9	19,315	511	4,949	836	8,231
WMC	Alumina Ltd.	Materials	6.55	1.7	9,600	586	5,010	912	8,555
WOW	Woolworths Ltd.	Food/Drug Retail	5.04	1.3	6,016	431	4,912	640	8,316
WPL	Woodside Petroleum Ltd.	Energy	10.14	3.1	7,500	280	3,031	486	4,779
Average			14.60	3.0	17,817	619	3,596	959	6,085

Table 2: Summary Statistics of the Shape of the Limit Order Book

Summary statistics of the shape of the limit order book for the 21 constituent stocks of the ASX-20 index. The reported numbers are respectively the cross-section averages of the length, relative length and height of the order book. The length is the number of shares on step k ($k = 1, 2, ...,$ 10); the relative length is the number of shares sitting on step k as a fraction of the total number of shares on the first 10 steps; and the height is the absolute difference between prices on steps j and $j - 1$ (in cents). For $j = 1$, the height is set to be the inside spread. Results are reported for both demand and supply sides. For each variable, the ratio is the demand divided by the supply. The sample spans from March 1 to March 31, 2000. Standard errors are reported in parentheses.

	Order Book Length (Shares)			**Relative Length (%)**			**Order Book Height (Cents)**		
Step	Demand	Supply	Ratio	Demand	Supply	Ratio	Demand (-)	Supply (+)	Ratio
Inside Spread							1.5	1.5	1.00
1	14,007	14,972	0.93	13.0	12.5	1.04			
2	14,486	15,879	0.91	13.6	13.3	1.02	2.6	2.7	0.96
3	12,336	14,007	0.88	12.0	11.9	1.00	3.4	3.2	1.06
4	10,910	12,481	0.87	10.6	10.5	1.00	4.2	3.7	1.13
5	10,450	11,740	0.89	10.0	9.7	1.03	5.1	4.2	1.21
6	9,787	11,768	0.83	9.2	9.4	0.97	6.0	4.6	1.30
7	9,076	10,759	0.84	8.4	8.7	0.96	6.8	4.9	1.38
8	8,991	10,232	0.87	8.0	8.4	0.95	7.6	5.0	1.52
9	8,461	9,901	0.85	7.7	8.0	0.96	8.4	5.4	1.55
10	8,104	9,427	0.85	7.4	7.6	0.97	9.1	5.5	1.65
Total	106,609	121,236							
Mean (s.e.)			0.87 (0.01)			0.99 (0.01)			1.27 (0.07)

Table 3: Dynamic Changes in Order Book Shape and Quote and Trading Activities

This table presents the cross sectional distribution of the dynamic changes in order book shape, quote and trading activities. We measure the change in the shape of the book in step 1 and from step n_1 to n_2 by the change in the respective weighted prices, WP^1 and $WP^{n_1-n_2}$, defined as:

$$\begin{aligned} WP^1 &= \frac{Q_1^d P_1^d + Q_1^s P_1^s}{Q_1^d + Q_1^s}, \\ WP^{n_1-n_2} &= \frac{\sum_{j=n_1}^{n_2}(Q_j^d P_j^d + Q_j^s P_j^s)}{\sum_{j=n_1}^{n_2}(Q_j^d + Q_j^s)}, \end{aligned}$$

where Q_j^d (Q_j^s) denotes the aggregate number of shares bid (offered) at price P_j^d (P_j^s) and j denotes the jth best price away from the inside market. The number shown under the WP^1 (or, WP^{2-10}) column is the percentage of time that WP^1 (or, WP^{2-10}) changes during the interval. The results are reported for four sampling intervals: 1 minute, 5 minutes, 10 minutes and 15 minutes.

Sampling Interval		WP^1	WP^{2-10}	**No. of Orders**	**No. of Trades**
1-minute	Mean	70%	52%	3	3
	Min	38%	28%	2	2
	Q1	62%	42%	2	2
	Median	70%	53%	3	3
	Q3	79%	64%	4	3
	Max	92%	72%	6	4
5-minute	Mean	97%	89%	14	9
	Min	84%	71%	4	3
	Q1	94%	84%	9	6
	Median	98%	90%	13	9
	Q3	99%	95%	19	12
	Max	100%	98%	26	19
10-minute	Mean	99%	96%	26	18
	Min	94%	87%	7	5
	Q1	98%	95%	17	11
	Median	99%	97%	25	17
	Q3	100%	99%	34	22
	Max	100%	100%	50	38
15-minute	Mean	99%	98%	40	27
	Min	98%	93%	12	8
	Q1	99%	97%	26	17
	Median	100%	99%	38	25
	Q3	100%	100%	54	34
	Max	100%	100%	75	58

Table 4: Estimates of the Error Correlation Model

This table reports coefficient estimates of the Error Correction model of WP^1 and WP^{1-10} for each of the 21 constituent stocks of the ASX-20 index. The model is specified as:

$$\begin{aligned}\Delta WP_t^1 &= \alpha_{1,0} - \alpha_1(WP_{t-1}^1 - \beta\, WP_{t-1}^{1-10}) + \sum_{i=1}^{p}(\gamma_{1,i}\Delta WP_{t-i}^1 + \delta_{1,i}\Delta WP_{t-i}^{1-10}) + \eta_{1,t},\\ \Delta WP_t^{1-10} &= \alpha_{2,0} + \alpha_2(WP_{t-1}^1 - \beta\, WP_{t-1}^{1-10}) + \sum_{i=1}^{p}(\gamma_{2,i}\Delta WP_{t-i}^1 + \delta_{2,i}\Delta WP_{t-i}^{1-10}) + \eta_{2,t},\end{aligned}$$

where WP^1 and WP^{1-10} measure the change in the shape of the order book in step 1, and from step 2 to 10, respectively, a_1 and a_2 are error correction coefficients, and β reflects the long-run equilibrium relation between WP^1 and WP^{1-10}. The results are based on the 5-minute interval data. A significance level of 5% (indicated by "*") is used to categorize each stock to one of the four cases:

- Case 1: α_1 is insignificant, but α_2 significant
- Case 2: α_1 is significant, but α_2 insignificant
- Case 3: Both α_1 and α_2 are significant
- Case 4: Neither α_1 nor α_2 is significant

Ticker	β	a_1	a_2	Case
AMP	1.001	0.012	0.126*	1
ANZ	1.000	0.008	0.052*	1
BHP	0.999	0.065*	0.061*	3
BIL	1.000	0.028*	0.122*	3
CBA	0.999	0.014	0.068*	1
CCL	0.998	0.055*	0.026	2
CGH	1.001	0.005	0.098*	1
CML	1.001	0.003	0.074*	1
CWO	0.998	0.020	0.087*	1
FBG	1.000	0.008	0.047*	1
LLC	0.999	0.018	0.079*	1
NAB	1.000	0.067*	0.197*	3
NCP	1.000	0.001	0.161*	1
NCPDP	1.004	0.002	0.130*	1
PBL	0.999	0.057*	0.050*	3
RIO	0.998	0.024*	0.037*	3
TLS	1.000	0.005	0.123*	1
WBC	0.998	0.019	0.049*	1
WMC	0.999	0.031	0.063*	1
WOW	0.998	0.027	0.069*	1
WPL	0.998	0.035*	0.092*	3
Average	0.999	0.024	0.086	

Table 5: Hasbrouck Information Share and Gonzalo-Granger Common Factor Weights

Reported below are the Hasbrouck (1995) information share and Gonzalo-Granger (1995) common factor weights for WP^1 (versus WP^{2-10}) for each of the 21 constituent stocks of the ASX-20 index. The maximum (minimum) contribution of WP^1 to the variance of the common factor's innovation is when WP^1 is the first (second) variable in the Cholesky factorization. The contribution of WP^{2-10} is similarly defined. The sum of maximum (minimum) contribution of WP_1 and the minimum (maximum) of WP^{2-10} is 1. The cross-sectional standard errors are in parentheses.

	WP^1					WP^{2-10}				
	Hasbrouck			G-G	$H_0: s_1 = 1$	Hasbrouck			G-G	$H_0: s_1 = 0$
Ticker	Max	Min	Ave.		$s_2 = 0$	Max	Min	Ave.		$s_2 = 1$
					(p-value)					(p-value)
AMP	0.99	0.50	0.75	0.90	(0.44)	0.50	0.01	0.25	0.10	(0.00)
ANZ	0.99	0.47	0.73	0.87	(0.61)	0.53	0.01	0.27	0.13	(0.00)
BHP	0.75	0.18	0.47	0.52	(0.00)	0.82	0.25	0.53	0.48	(0.01)
BIL	0.94	0.57	0.75	0.84	(0.02)	0.43	0.06	0.25	0.16	(0.00)
CBA	0.97	0.39	0.68	0.83	(0.35)	0.61	0.03	0.32	0.17	(0.00)
CCL	0.73	0.10	0.41	0.36	(0.00)	0.90	0.27	0.59	0.64	(0.04)
CGH	0.98	0.78	0.88	0.95	(0.44)	0.22	0.02	0.12	0.05	(0.00)
CML	0.99	0.48	0.74	0.97	(0.90)	0.52	0.01	0.26	0.03	(0.00)
CWO	0.98	0.41	0.70	0.83	(0.52)	0.59	0.02	0.30	0.17	(0.00)
FBG	0.96	0.80	0.88	0.83	(0.48)	0.20	0.04	0.12	0.17	(0.00)
LLC	0.94	0.46	0.70	0.81	(0.10)	0.54	0.06	0.30	0.19	(0.00)
NAB	0.96	0.38	0.67	0.78	(0.04)	0.62	0.04	0.33	0.22	(0.00)
NCP	0.99	0.77	0.88	0.97	(0.70)	0.23	0.01	0.12	0.03	(0.00)
NCPDP	0.99	0.93	0.96	0.98	(0.72)	0.07	0.01	0.04	0.02	(0.00)
PBL	0.78	0.33	0.56	0.58	(0.00)	0.67	0.22	0.44	0.42	(0.00)
RIO	0.81	0.30	0.56	0.67	(0.02)	0.70	0.19	0.44	0.33	(0.00)
TLS	0.99	0.56	0.78	0.96	(0.84)	0.44	0.00	0.22	0.04	(0.00)
WBC	0.97	0.51	0.74	0.83	(0.26)	0.49	0.03	0.26	0.17	(0.00)
WMC	0.92	0.27	0.59	0.68	(0.10)	0.73	0.08	0.41	0.32	(0.00)
WOW	0.93	0.35	0.64	0.74	(0.09)	0.64	0.07	0.36	0.26	(0.00)
WPL	0.85	0.37	0.61	0.71	(0.00)	0.63	0.15	0.39	0.29	(0.00)
Average	0.93	0.47	0.70	0.79		0.52	0.07	0.30	0.21	
(s.e.)	(0.02)	(0.05)	(0.03)	(0.04)		(0.04)	(0.02)	(0.03)	(0.04)	

Table 6: Regression Analysis of Returns and Order Book Length/Height

The results below are based on the following regression equation:

$$\epsilon_t = \alpha_0 + \alpha_1 Spread_{t-1} + \beta_1 QR_{1,t-1} + \sum_{j=2}^{n} \beta_i \, QR_{j,t-1} + \sum_{j=2}^{n} \gamma_j \, HR_{j,t-1} + \eta_t, \quad n = 2, \ldots, 10$$

where ϵ is the standardized innovation in 5-minute return obtained from an AR(5) model, and *Spread* denotes the inside relative spread. For each step j, the imbalance in the length (QR) and height (HR) of the order book are defined as:

$$QR_j = \frac{Q_j^s - Q_j^d}{Q_j^s + Q_j^d}, \qquad j = 1, \ldots, 10$$

$$HR_j = \frac{(P_j^s - P_{j-1}^s) - (P_j^d - P_{j-1}^d)}{(P_j^s - P_{j-1}^s) + (P_j^d - P_{j-1}^d)}, \qquad j = 2, \ldots, 10$$

where P_j^d (P_j^s) is the jth best price on the demand (supply) side of the book and Q_j^d (Q_j^s) is the depth available at that price. The 5-minute interval data from each firm are pooled together. Reported in the table are adjusted R^2 for each regression using order book information up to step n, n=1, 2,...,10. We present results for all observations and for the top and bottom 5% observations with the largest liquidity imbalance up to $5\overline{Q}$, where $\overline{Q}$ is the average (stock-specific) trade size.

Up to Step n	All Obs. Adj. R^2	Top/bottom 5% Imbalanced Obs. Adj. R^2
1	10.78	18.33
2	12.02	20.30
3	12.41	21.16
4	12.52	21.18
5	12.53	21.15
6	12.55	21.31
7	12.56	21.38
8	12.58	21.48
9	12.58	21.53
10	12.60	21.62

Table 7: Regression Analysis of Returns and Liquidity

The regression results below are based on the following equation:

$$\epsilon_t = \alpha_0 + \alpha_1 Spread_{t-1} + \beta LR(q)_{t-1} + \eta_t, \quad q = 1.0\overline{Q}, \ldots, 5.0\overline{Q}$$

where ϵ is the standardized innovation in 5-minute return obtained from an AR(5) model, and *Spread* the inside relative spread. Let $\overline{Q}$ denote the average (stock-specific) trade size. For each stock and a given trade size of q ($q = 1.0\overline{Q}, \ldots, 5.0\overline{Q}$), $LD(q)$ is a measure of liquidity on the demand side, defined as the discount per share below the midpoint of the best bid and ask:

$$LD(q) = 0.5(P_1^d + P_1^s) - \frac{\sum_{j=1}^{m_1-1} P_j^d Q_j^d + P_{m_1}^d Q_{m_1}^{\prime d}}{q}, \qquad \sum_{i=1}^{m_1-1} Q_i^b + Q_{m_1}^{\prime b} = q,$$

where P_j^d is the jth best price on the demand side of the book and Q_j^d the depth available at that price. The step m_1 is determined according to $\sum_{j=1}^{m_1-1} Q_i^d < q \leq \sum_{j=1}^{m_1} Q_i^d$. The liquidity on the supply size, $LS(q)$, is the premium per share above the midpoint of the best bid and ask, and defined as

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

The step m_2 is determined according to $\sum_{j=1}^{m_2-1} Q_i^s < q \leq \sum_{j=1}^{m_2} Q_i^s$. The independent variable is $LR(q)$, the imbalance in liquidity:

$$LR(q) = \frac{LS(q) - LD(q)}{LS(q) + LD(q)}, \quad q = 1.0\overline{Q}, \ldots, 5.0\overline{Q}.$$

The 5-minute interval data from each firm are pooled together. Reported in the table are adjusted R^2 for each regression with a trade size of q ($q = 1.0\overline{Q}, \ldots, 5.0\overline{Q}$). We present results for all observations and for the top and bottom 5% observations with the largest liquidity imbalance up to $5\overline{Q}$, where $\overline{Q}$ is the average (stock-specific) trade size.

Trade Size $q =$	All Obs. Adj. R^2	Top/bottom 5% Imbalance Obs. Adj. R^2
1.0 $\overline{Q}$	6.68	17.75
1.5 $\overline{Q}$	7.70	18.46
2.0 $\overline{Q}$	8.28	18.99
2.5 $\overline{Q}$	8.77	19.04
3.0 $\overline{Q}$	9.24	19.25
3.5 $\overline{Q}$	9.42	19.42
4.0 $\overline{Q}$	9.72	19.54
4.5 $\overline{Q}$	9.86	19.59
5.0 $\overline{Q}$	9.98	19.61

Table 8: Ordered Probit Model of Order Choice

We estimate the following ordered probit model of actions by limit order traders. Possible actions and assigned rank R are: submission of a (marketable) limit order ($R = 6$), strict improvement of the current best bid/offer ($R = 5$), matching of the current best bid/offer ($R = 4$), submission of a limit order at the second or third steps ($R = 3$), submission of a limit order beyond the third step ($R = 2$), and cancellation of a limit order ($R = 1$). The rank of order j is related to a latent variable Z_j

$$R_j = \begin{cases} 6, & \text{if } Z_j \in [\mu_5, \infty), \\ 5, & \text{if } Z_j \in [\mu_4, \mu_5), \\ 4, & \text{if } Z_j \in [\mu_3, \mu_4), \\ 3, & \text{if } Z_j \in [\mu_2, \mu_3), \\ 2, & \text{if } Z_j \in [\mu_1, \mu_2), \\ 1, & \text{if } Z_j \in (-\infty, \mu_1), \end{cases}$$

where the μ_k are intercept parameters to be estimated. Z_j, in turn, is modeled as a function of the observed explanatory variables

$$\begin{aligned} Z_j \quad = \quad & \beta_1 Spread + \beta_2 Q_1^{own} + \beta_3 \sum_{i=2}^{3} Q_i^{own} + \beta_4 \sum_{i=4}^{10} Q_i^{own} + \beta_5 Q_1^{other} + \beta_6 \sum_{i=2}^{3} Q_i^{other} + \beta_7 \sum_{i=4}^{10} Q_i^{other} \\ & + \beta_8 |P_1^{own} - P_{10}^{own}| + \beta_9 |P_1^{other} - P_{10}^{other}| + \epsilon_i. \end{aligned}$$

The superscripts *own* and *other* denote the demand or supply side of the book, depending on the direction of the order. All quantity variables are scaled by a stock's average trade size, and all price variables are scaled by the contemporaneous midprice between the best bid and offer. Finally, all explanatory variables are standardized.

Variable	Overall	Buys	Sells
$Spread$	0.0710 *	0.0808 *	0.0605 *
Q_1^{own}	-0.0912*	-0.0967*	-0.0909*
$\sum_{i=2}^{3} Q_i^{own}$	-0.0001	-0.0014	-0.0016
$\sum_{i=4}^{10} Q_i^{own}$	0.0112 *	0.0081 *	0.0142 *
Q_1^{other}	0.0118 *	0.0209 *	0.0101 *
$\sum_{i=2}^{3} Q_i^{other}$	0.0073 *	0.0155 *	0.0033
$\sum_{i=4}^{10} Q_i^{other}$	-0.0118*	-0.0068*	-0.0190*
$\|P_1^{own} - P_{10}^{own}\|$	-0.0134*	-0.0125*	-0.0170*
$\|P_1^{other} - P_{10}^{other}\|$	0.0074 *	0.0131 *	0.0138 *

* indicates significance at the 5% level.