Behind Stock Price Movement: Supply and Demand in Market Microstructure and Market Influence

Jingle Liu and Sanghyun Park
Behind Stock Price Movement: Supply and Demand in Market Microstructure and Market Influence

JINGLE LIU AND SANGHYUN PARK

The determination of a security’s price is mainly driven by the balance between supply of and demand for that security’s liquidity in the marketplace. A limit order book where buy/sell orders are displayed reflects the interests of market participants and, consequently, contains information about market consensus. Therefore, the study of how market participants trade and quote is important to the understanding of how a security’s price is decided. Evans and Lyons [1999] used signed transaction data on the foreign exchange dealer market and showed that the imbalance explains the Forex returns quite well. Bouchard [2002] empirically studied statistical properties of limit order books of three liquid stocks on the Paris Bourse. Chordia et al. [2002] explored the relationship between the aggregated daily order imbalance and returns. Farmer et al. [2004] showed that large price fluctuations are driven by liquidity fluctuations—variations in the market’s ability to absorb new orders. Hopman [2007] found that imbalance between uninformed buy/sell pressure explains most of the stock price changes on the Paris Bourse, looking at a much longer time horizon than that of our study. Cont et al. [2014] studied price change caused by order flow imbalance, which is defined as the imbalance between supply and demand at the best bid and ask prices.

In this article, we approach this problem through the study of explanatory factors for short-term stock price movement in the U.S. equity market by exploiting the relationship between liquidity provisions and taking activities at the market-microstructure level by quantitatively measuring central limit order book imbalance and trade imbalance. Furthermore, we investigate the macro-level effect of whole market movement on individual stock prices.

Most previous studies were based on trade and quote (TAQ) data of a single exchange, such as NYSE (Chordia et al. [2002]), Paris Bourse (Bouchard et al. [2002]; Hopman [2007]), or the Tokyo Stock Exchange (Cont et al. [2010]). In the past decade, global equity markets have seen a proliferation of electronic trading venues and a consequent fragmentation of order flow. Modeling the consolidated central limit order book dynamics has become necessary to capture private and public information of quoting and trading events in order to explain the stock price return. We investigate the correlation between stock price changes and the central limit order book by combing level 1 data from all 12 lit exchanges in the U.S. equity market.

One aspect that previous market impact models rarely addressed is dark pool trading versus lit exchange trading. Trading in dark pools has less direct impact on the displayed
liquidity in the order book than does trading in lit exchanges. Because of the different natures of those two trading venue types and the different roles dark pools play for various stocks, we separated trades in lit exchanges from trades in dark pools in the model and found that this separation can make the model better explain the price movement.

Furthermore, we develop an intuitive and empirical framework to model the supply–demand behavior on a micro level by computing order book quote imbalance and trade imbalance. This model also makes a linkage between price movement of individual stocks and that of the whole market on the macro level. Only a few studies have been done on this micro-to-macro perspective of stock price dynamics.

We systematically studied the influences of these driving forces on stock price change by investigating how micro and macro driving forces affect stock price change differently for various types of stocks and various time intervals. On the aspect of stock spread, our study indicates that the quote imbalance has higher explanatory power for stocks with tighter spread measured by number of ticks compared with stocks with a wider spread. On the aspect of time interval, we also show that as the time interval increases from 30 seconds to 1 hour, the influence of the trade imbalance and market movement on stock price grows higher while the quote imbalance becomes less. These results reveal how the information values of those elements vary within the dimensions of price and time.

This model not only sheds light on fundamental elements that drive price movement for a broad range of stocks and time intervals but also suggests practical solutions for traders or brokers using trading algorithms to optimize order execution performance by intelligently controlling their footprints in the public marketplace and avoiding being picked up by predatory traders. In portfolio management, a more accurate measure of expected trading cost is a vital component of effective portfolio strategy implementation and trading performance monitoring (Almgren and Chriss [2000]; Engle and Ferstenberg [2007]; Kritzman et al. [2007]; Obizhaeva and Wang [2013]; Liu and Phadnis [2013]; Alfonsi et al. [2012]).

The article is organized as follows: First, we describe a stylized multivariate linear model for the relationship between stock price return and stock limit order book dynamics along with the market movement. Second, we show the summary statistics of stock universe and model variables. Third, we show the estimation of model parameters using high-frequency time series of stock order book and market return; we illustrate the model using data from the U.S. equity market. Last, we explore the explanatory power of each factor across different tickers and time intervals and analyze the dependence of the model’s explanatory power on the bid–ask spread.

**MODEL**

The dynamics of stock price movement are complex in nature. Multiple buy and sell orders are posted on central limit order books with double auction processes going on at both the bid and ask sides. The actual price of a stock, typically measured by the mid quote, varies based on contemporaneous liquidity supply and demand for the stock itself as well as the influence of the entire market movement.

Higher demand for and lower supply of sell (buy) limit orders will drive up (down) the price, whereas lower demand for and higher supply of buy (sell) limit orders will depress (move up) the price. Quantifying these effects has been the key to explaining stock price changes. For a micro-level perspective of an order book, the supply of liquidity could be measured by the number of buy/sell limit orders posted in an order book; the demand of liquidity could be measured by buyer-initiated market orders and seller-initiated market orders. Limit orders can only be posted in lit exchanges; dark pools do not have displayed liquidity. But trades can occur in both lit and dark venues. Trades in lit exchanges typically involve the removal of liquidity on the quote, while trades in dark pools do not have such direct effect. Because of the different natures of those two types of trading venues, it is reasonable to treat trades in lit exchanges differently from those done in dark pools. To formulate a mathematically precise description of supply–demand imbalance, we break down the imbalance into quotes imbalance $Q$, lit trade imbalance $LT$, and dark trade imbalance $DT$.

The quote imbalance $Q$ is defined as net limit posting order flow—which is the difference between net added buy limit orders $Q_B$ and net added sell limit orders $Q_S$. For quote imbalance calculation, only limit orders on best bid–ask prices are being taken into consideration. $Q_B$ is newly added limit orders minus cancelled
limit orders on best bid, while $Q_s$ is newly added limit orders on the ask $Q_{S^+}$ minus cancelled limit orders on best ask $Q_{S^-}$.

$$Q = Q_B - Q_S = (Q_{B^+} - Q_{B^-}) - (Q_{S^+} - Q_{S^-})$$  \(1\)

The lit trade imbalance $LT$ is defined as

$$LT = LT_B - LT_S$$  \(2\)

$$DT = DT_B - DT_S$$  \(3\)

where $LT_B$, $DT_B$ are buyer-initiated trades in lit exchanges and dark pools, respectively, and $LT_S$, $DT_S$ are seller-initiated trades in lit exchanges and dark pools, respectively. The rule to classify the sign of the trade is based on whether the trade is above or below the mid quote price at the time of trade. If trade price is above mid quote, it is classified as a buyer-initiated trade. If trade price is below mid quote, it is classified as a seller-initiated trade. If trade price is the same as mid quote, it is not counted when calculating trade imbalance.

In addition to micro-level supply–demand imbalance in the order book, we also include the market movement in the model as a macro-level factor that reflects the influence of the industry-wide movement caused by central bank decision(s), government policy, geopolitical events and so on. Some part of the stock price movement that is not explainable by quote and trade imbalance can be attributed to this exogenous market effect. In our study, the S&P 500 Index is used to gauge the U.S. equity market movement.

To explain the stock price change $r_i(t)$ (in $\$\$) for given time interval $i$, we developed a multivariate linear model as follows:

$$r_i(t) = \beta_0 + \beta_1 Q(t) + \beta_2 LT(t) + \beta_3 DT(t) + \beta_4 r_{m}(t)$$  \(4\)

where $Q(t)$, $LT(t)$, $DT(t)$ and $r_{m}(t)$ are stock order book quote imbalance, lit trade imbalance, dark trade imbalance and market return in time interval $i$, respectively, and are calculated using Equations (1)–(3). Market movement $r_m$ is calculated as log return of the S&P 500 Index during time interval $i$. $\beta_i$, $i = 0, 1 \ldots 4$ are coefficients for the linear model. $\beta_0$, $\beta_4$ have units of $\$$ and $\beta_1$, $\beta_2$, $\beta_3$ have units of $\$$/shares. The reason to use a linear model is that it is simple, intuitive, easily understood and robust. The empirical results of applying this model will be discussed in the following sections.

**STATISTICS OF MODEL VARIABLES**

The multivariate linear model described earlier was applied to high-frequency order-level data for 42 U.S. tickers with diverse profiles of price, spread, liquidity and volatility. The summary statistics of average daily volumes, average bid–ask spreads, average prices and volatilities of the stock universe here are shown in Exhibit 1. As a measure of liquidity, average daily volumes range from 18,367 shares to 49,434,976 shares. Volume-weighted average prices range from $2.647 to $1,142.558. As a major component of trading cost, bid–ask spreads of the sample universe range from 1 cent, which is the smallest tick size in the U.S. market, to $1.49—more than 100 times the smallest tick size. Most-volatile stocks and least-volatile stocks have volatilities of 116.48 and 9.13, respectively. The diversity of the stock universe allows us to explore model dynamics in various aspects and extract insights on price movement.

We used Bloomberg-managed B-PIPE market data, which contain top-level quote updates from all U.S. exchanges and trades in lit/dark venues, to compute all independent and dependent variables. We investigated the model fitting for time intervals from 30 seconds to 1 hour. The time period of the data set used in this study was from 4/21/2014 to 5/21/2014. Only continuous trading data between 9:30 and 16:00 ET were used to fit the model; auction prints and off-hour trading were excluded to remove other effects, such as company earnings releases, that are not in the scope of this study.

The independent and dependent variables were computed for each time interval separately before the regression test was performed. For example, all the

EXHIBIT 1

Summary Statistics of Average Daily Volumes, Average Bid–Ask Spreads, Average Prices and Volatilities of 42 U.S. Securities

<table>
<thead>
<tr>
<th></th>
<th>Avg Daily Volume</th>
<th>Avg Bid–Ask Spread ($)</th>
<th>Avg Price ($)</th>
<th>Volatility (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max</td>
<td>49,434,976</td>
<td>1.490</td>
<td>112,558</td>
<td>116.48</td>
</tr>
<tr>
<td>75th Percentile</td>
<td>5,501,776</td>
<td>0.595</td>
<td>206,673</td>
<td>31.61</td>
</tr>
<tr>
<td>Median</td>
<td>1,037,268</td>
<td>0.884</td>
<td>78,720</td>
<td>25.71</td>
</tr>
<tr>
<td>25th Percentile</td>
<td>523,687</td>
<td>0.019</td>
<td>36,322</td>
<td>18.00</td>
</tr>
<tr>
<td>Min</td>
<td>18,367</td>
<td>0.010</td>
<td>2,647</td>
<td>9.13</td>
</tr>
<tr>
<td>Mean</td>
<td>4,013,778</td>
<td>0.333</td>
<td>166,617</td>
<td>28.87</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>8,123,805</td>
<td>0.431</td>
<td>223,670</td>
<td>19.51</td>
</tr>
</tbody>
</table>
BEHIND STOCK PRICE MOVEMENT: SUPPLY AND DEMAND IN MARKET MICROSTRUCTURE AND MARKET INFLUENCE

SUMMER 2015

quote and trade data between 4/21/2014 and 5/21/2014 were grouped in five-minute intervals, such as 9:30–9:35, 9:35–9:49 and 15:55–16:00. Each full trading day has 78 intervals, with one dependent variable and four independent variables for each interval. To remove outliers, the intervals containing the top 1% and bottom 1% of price changes, quote imbalances and trade imbalances were removed from the data sample. As an illustration of data, summary statistics of model variables for six stocks are shown in Exhibit 2.

The skewness levels of most variable distributions are between –1 and +1, which means that variables are quite symmetrically distributed. The fact that means of variable distributions are small compared with the standard deviation suggests that the center of variable distribution is close to zero, as would be expected in a balanced market.

ESTIMATION OF MODEL COEFFICIENTS

The fitted model coefficients for time intervals of five minutes for 42 tickers are shown in Exhibit 3. Results show that the percentage of tickers whose absolute t-statistics of \( \beta_0, \beta_1, \beta_2, \beta_3 \) and \( \beta_4 \) are larger than 1.97—which corresponds to the critical two-tailed \( t \) value for 95% confidence level with degree of freedom of around 1,700—are as follows: 26.2%, 100.0%, 100.0%, 61.9% and 95.2%, respectively. Therefore, quote imbalance, lit trade balance and market movement are quite significant for most tickers tested.

For time intervals of five minutes, the coefficient of lit trade imbalance \( \beta_2 \) is typically larger than the coefficient of quote imbalance \( \beta_1 \). The explanation: market participants consider that trades contain more information than quote updates, that is, the same amount of traded shares would create more market impact than the impact created by the same amount of quote change. For example, for AMZN in time intervals of five minutes, 1,000 shares traded on the offer side would,
on average, move the mid-price by $1.8 \times 10^{-5} \times 1000 = 0.018$, while a 1,000-share limit order cancellation on the offer side would, on average, move the mid-price by $9.6 \times 10^{-6} \times 1000 = 0.0096$—which is almost half of the trade impact.

Dark trade imbalance coefficients $\beta_3$ are much smaller than $\beta_1$ and $\beta_2$. This suggests that trades in dark pools have less market impact than lit trades (which remove the posted liquidity in exchanges) and agrees with the conventional notion that information leakage...
EXHIBIT 4
Coefficients of Quote Imbalance (QTE\_IM), Lit Trade Imbalance (L\_TRD\_IM) and Dark Trade Imbalance (D\_TRD\_IM) on Stock Price Movements in Various Time Intervals for DNR, CB, IBM, PPG, NFLX and ICPT
on the order can be reduced by trading in dark pools. Market movement coefficients $\beta_4$ show how stock prices are correlated with market prices. All the tickers but one show positive coefficients, which means their prices move together with the market to some degree.

The model was also fitted at time intervals from 30 seconds to 1 hour. Results of six tickers are shown in Exhibit 4. As the time interval gets longer, $\beta_1$ and $\beta_2$ generally decrease; for example, on ticker CB, $\beta_1$ and $\beta_2$ decrease from $5.6e^{-6}$ $/shares and $8.9E{-6}$ $/shares in time intervals of 30 seconds to $3.5e^{-6}$ $/shares and $3.9e{-6}$ $/shares in time intervals of 1 hour, respectively. This suggests that the short-term price movement is more sensitive to the liquidity supply–demand on (quote and trade) imbalance in lit order books than the long-term price movement. For most time intervals and tickers, $\beta_1$ is larger than $\beta_2$. The difference between them is larger for shorter time intervals and for tickers with wider spreads. On the other side, $\beta_3$ tends to be larger for time intervals of 2 minutes to 30 minutes than for very short or very long time intervals. $\beta_3$ is smaller than $\beta_1$ and $\beta_2$ for DNR, CB and IBM—whose spreads are less than $0.5$—while $\beta_3$ becomes larger than $\beta_2$ for PPG, NFLX and ICPT—whose spreads are wider than $0.5$. The different behavior between narrow-spread tickers and wide-spread tickers will be discussed in detail later.

EXPLANATORY POWER

The explanatory power of the model, that is, percentage of stock price movement variance that can be explained by the model, is measured by the adjusted R-squared. The attribution of explanatory power to each factor can be readily deduced by performing variance analysis on the linear model. The measured model’s explanatory power for 42 U.S. tickers in five-minute time intervals is summarized in Exhibit 5. The total R-squared ranges from 3.3% for NATH to 78.2% for VOD. On average, quote imbalance explains most of the stock price movement, or 38.5% of the price change. Next are two significant factors: lit trade imbalance and market return, which explain 8.5% and 5.9% of price movement, respectively. The dark trade imbalance has
Exhibit 6
Explanatory Powers of Quote Imbalance (QTE_IM), Lit Trade Imbalance (L_TRD_IM), Dark Trade Imbalance (D_TRD_IM) and Market Return (RET_MKT) on Stock Price Movements in Various Time Intervals for DNR, CB, IBM, PPG, NFLX and ICPT
the lowest explanatory power, with average $R^2$ of 0.3%, which is consistent with the observed smaller fitted values of $\beta_3$.

In different time intervals, the explanatory power of each factor as well as the whole mode change drastically. Exhibit 6 plots the explanatory powers of factors on stock price movements in various time intervals for DNR, CB, IBM, PPG, NFLX and ICPT. The results show that as the time interval gets longer, the explanatory power of quote imbalance decreases for all tickers except for DNR, and the explanatory power of lit trade imbalance, dark trade imbalance and market return becomes higher. This suggests that 1) information about the market sentiment or news is digested and reflected in the individual stock price at a slower rate than quote update and trade information. Market return exerts greater influence on stock price on a long time horizon. 2) In very short intervals, price movement is dominantly influenced by quote add or quote cancel. And, 3) trades in lit/dark venues have more explanatory power at longer time intervals. For example for NFLX, quote imbalance, lit trade imbalance and market move account for 39.1%, 15.9% and 4.1% of stock price move in 30-second intervals, respectively, while they account for 12.6%, 23.7% and 11.4% in 1-hour intervals, respectively.

**BID–ASK SPREAD**

Bid–ask spread, one of most important fundamental properties of a stock, is the dominant part of immediate transaction cost. Taking liquidity on the opposite side will cost a spread versus passive order posting on the same side. Exhibit 4 and Exhibit 6 show that coefficients and explanatory powers vary with different tickers. Average spreads of DNR, CB, IBM, PPG, NFLX and ICPT are $0.010, $0.022, $0.048, $0.102, $0.315 and $1.045, respectively. The results show that as the bid–ask spread widens, both the ratio of trade imbalance coefficient to quote imbalance coefficient and the ratio of trade imbalance $R^2$ to quote imbalance $R^2$ become larger. This means that narrow-spread stocks are more quote-driven in nature, while wide-spread stocks are more trade-driven. When the spread is 1 tick or $0.01, an update on best bid and ask quantities has a direct impact on price pressure with a similar magnitude of trade. Whereas, when spread is more than one tick wide, trades offer more value in the process of price discovery.

The explanatory powers of the four factors are measured for 42 U.S. tickers with spreads ranging from $0.01 to $1.49; results are plotted against bid–ask spread and are shown in Exhibit 7. As the bid–ask spread widens, the explanatory power of quote imbalance decreases dramatically—while the explanatory power of trade imbalance and market movement increases moderately. A possible explanation is that on wider-spread stocks liquidity takers pay higher transaction costs, which must then be compensated for by the greater amount of information traders have at the time of trade. Consequently, the trade would be interpreted by the public market as containing more private information. Another possible reason is that to avoid the higher spread cost, more trading activities occur within the spread through hidden order type in lit exchanges or dark pools, and those activities are captured by trade imbalance instead of quote imbalance, thus quote imbalance explains less of price variation. For example, ICPT has an average spread of $1.04 and 49.5% of trades occur in dark pools, while CB has an average spread of $0.02 and 22.1% of trades occur in dark pools.

**SUMMARY**

In this article, we develop a multivariate linear model to explain short-term stock price movement from the perspectives of both micro-level order book supply–demand dynamics (quote imbalance, lit trade imbalance, dark trade imbalance) and macro-level market influence (market movement). This empirical and intuitive model has explanatory power of up to around 80% for time intervals of 30 seconds to 1 hour in the U.S. equity market. The influence of each factor varies with the type of stock as well as the length of the time interval. The model also reveals the various and interesting mechanisms of moving stock price by classifying stocks into quote–driven stocks and trade-driven stocks. The bid–ask spread has been shown to be the determining factor for classification. Further work can be done to exploit this model for a wider range of applications such as price explanation or prediction for other asset classes.
EXHIBIT 7
Relationships between Bid–Ask Spread and R² Values of Quote Imbalance (QTE_IM), Lit Trade Imbalance (L_TRD_IM), Dark Trade Imbalance (D_TRD_IM) and Market Return (RET_MKT) in Time Intervals of 300 Seconds
ENDNOTE

The authors thank Kapil Phadnis and Amber Anand for their insightful discussions and comments on this work.

REFERENCES


To order reprints of this article, please contact Dewey Palmieri at dpalmieri@iijournals.com or 212-224-3675.