Carry Forward Modeling for High-Frequency Limit-Order Executions: An Emerging Market Perspective

American Business Review
May 2022, Vol.25(1) 92 - 119

ISSN: 2689-8810 (Online)
ISSN: 0743-2348 (Print)

Aritra Pan[a] and Arun Kumar Misra[b]

https://doi.org/10.37625/abr.25.1.92-119

ABSTRACT

In this study, we estimate the order execution probability of a limit-order book (LOB) and analyze its determinants using high-frequency LOB data from the National Stock Exchange (NSE) of India. For this purpose, we propose an algorithm that estimates the LOB execution time. Using a survival function with log-normal distribution, this study analyzes the significant determinants of the limit-order execution times. The average execution probability is found to be higher for stocks belonging to the information technology and telecom sectors. The limit-order execution probability increases with a larger bid–ask spread, lower limit-order size, and deeper opposite order book. On the other hand, multiple factors, including price aggressiveness, inferior price, limit-order size, and spread, have a direct impact on execution times. The findings could help traders understand various factors influencing the probability of execution and execution time of LOBs. This study is unique in that it models limit-order execution using high-frequency tick-by-tick trading data for emerging markets, such as the NSE of India.

KEYWORDS

High-Frequency Trading, Limit-Order Book, Execution Probability of Limit Orders, Survival Analysis

INTRODUCTION

In recent years, the availability of high-frequency trading data pertaining to orders, quotes, and transactions from stock markets has stimulated research on high-frequency order books. Traders place buy or sell orders of stocks at particular prices below (for buy orders) or above (for sell orders). These orders are electronically executed by the exchange on time- and price-based priorities. A limit order transacts a specific number of shares at a predetermined price. There is always uncertainty in the execution of a limit order as it depends on factors such as trading frequency, bid-ask spread, and market quality (Battalio et al., 1997; Chung et al., 1999). Thus, limit-order execution probability, execution times, and their respective determinants can have direct implications on stock markets. Studies have mentioned limit-order execution as a function of factors like limit order size, quoted bid-ask spread, price aggressiveness, order book depth, price volatility, and inferior price (Al-Suhaibani & Kryzanowski, 2000; Chatterjee & Mukhopadhyay, 2013; Cho & Nelling, 2000; Gava, 2005; Lo et al., 2002; Omura et al., 2000). However, studies have been carried out (using low-frequency as well as hypothetical data) to estimate the probability of limit-order execution. The estimation of execution probability and execution time using high-frequency order book data is an evolving area of research.

We propose econometric models to estimate limit-order execution probability and execution times using high-frequency trading data. Such models will apply logistic regression to estimate limit-order execution probability and survival analysis to estimate limit-order execution times. We further propose

[a] Indian Institute of Management, Bodh Gaya, India (aritra.p@iimbg.ac.in ; aritra.pan360@gmail.com)
[b] Indian Institute of Technology, Kharagpur, India (arunmisra@vgsom.iitkgp.ac.in)

Corresponding Author:
Pan (aritra.p@iimbg.ac.in ; aritra.pan360@gmail.com)

to analyze the impact of limit-order size, trading day, spread, depth, and volatility on execution time and probability from a high-frequency trading context.

This study presents an experiment with a limit-order book (LOB) pertaining to the order-driven market of the National Stock Exchange (NSE) of India. Prior to 1992, trading on the Indian stock market was done in written (paper) form. In 1992, the NSE was established; it commenced operation in an electronic form of trading. Traders are asked to place orders in an electronic trading format, and orders are placed in a format that includes price, volume, and trading-time. To maintain efficiency, liquidity, and transparency in securities trading, fully automated screen-based trading systems were introduced in stock exchanges of India. In online trading, orders are electronically matched based on price-time priority. Traders are allowed to carry over their expired orders to the next trading cycle, which is known as a “carry forward” order (Nath & Dalvi, 2004).

In this study, we consider approximately 4.1 million buy and sell limit orders over a 22 trading-day period. In a controlled experimental setup, the orders are processed for matching and creating order books. High-frequency trading data for 30 frequently traded stocks belonging to NSE-CNX 500 Index are processed using large-scale databases. Stocks are selected from the dominating sectors of the Indian market, such as consumer goods, financial services, information technology and telecom services (IT-Telecom), services and healthcare, pharmaceuticals, and automobile and industrial manufacturing. The period consists of both bullish and bearish market conditions along with regular conditions. The study proposes the design of an order-matching algorithm to construct a limit-order book. This algorithm will estimate the execution probability and execution time for bids and asks as well as their determinants. It will explore the survival functions and examine the determinants of the survival of limit orders.

The findings of this research could help traders to design their limit order execution strategies to improve the execution probability of LOBs. Specifically, the findings could help traders 1) to understand the factors influencing the execution of LOBs and 2) to identify the right time to place their orders. Consequently, traders might not only enhance the execution probability but also reduce the opportunity cost of waiting. Moreover, the findings are applicable to order-driven markets of both developed and emerging countries.

LITERATURE REVIEW

This study undertook an in-depth review of previous literature pertaining to order-matching, order book design, and order execution strategies. A survey of various models for the estimation of execution probability is also presented. The order-matching algorithm and survival methodology have been discussed in-depth.

ORDER BOOK RECONSTRUCTION

The literature on execution probability and execution times of limit orders is primarily modeled on historical data. Using high-frequency order book data of different order execution states, such as submission, modification, partial execution, full execution, cancellation, and expiry, various researchers have reconstructed order books (Al-Suhaibani & Kryzanowski, 2000; Cho & Nelling, 2000; Gava, 2005; Lo et al., 2002; Omura et al., 2000; Wen, 2008). Handa and Schwartz (1996) and Chatterjee and Mukhopadhyay (2013) used “hypothetical” orders and estimated the execution probability. The reconstruction of the order book from historical data and the estimation of execution probabilities using the Weibull and Gamma distributions were undertaken by Omura et al. (2000), Lo et al. (2002), Gava (2005), and Cho and Nelling (2000). Al-Suhaibani and Kryzanowski (2000) censored observations for canceled and expired orders and used the Weibull distribution for the survival analysis with a set

of regressors. They concluded that the aggressive pricing of limit-orders would lower the execution time and improve the execution probability which is consistent with the price priority rule. Order imbalance is found to have a positive effect on limit-order waiting time, which is consistent with the findings of Handa and Schwartz (1996) and Huang et al. (2015).

EXECUTION PROBABILITY OF LIMIT ORDER AND ITS DETERMINANTS

A limit order makes the execution conditional on a limit price. In a limit-order market, prices of new orders are compared with the previously placed orders in the system to determine whether there is a match. In a price-time priority system, a buy (or sell) order with the highest (or lowest) limit price will have the highest trading priority (Bouchaud et al., 2018). A model for determining execution probability is key to determining the order execution strategy in an order-driven market.

Eisler et al. (2009), Raudys and Matkėnaitė (2016), and Yingsaeree (2012) discussed different types of limit orders differentiated by their complete execution, partial execution, and non-execution status. A limit order will be executed if enough market orders arrive during the trading horizon to execute all preceding orders in the order book. Thus, the execution probability depends on the state of the book (book depths and spread), on when traders submit their orders, and on the future incoming flow of market and limit orders. Execution probability depends on a trader's order-submission strategy. Parlour (1998), Foucault (1999), and Foucault et al. (2005) showed that execution probability could be calculated by estimating the probability of traders' submission of orders to trigger the desired limit order. Smith et al. (2003), Hollifield et al. (2001), and Yingsaeree (2012) estimated execution probability using historical data on trades and quotes. Omura et al. (2000) and Hollifield et al. (2001) defined execution probability for a specific time limit as the ratio of the number of limit orders executed within the time limit to the total number of limit orders present.

Bid-ask spread, order size, limit price, and volatility are major determinants of execution probability. Cho and Nelling (2000) used duration models to analyze the relationship between execution probability and variables that included the order size, limit price, market volatility, bid-ask spread, and time of day when an order is submitted. Yingsaeree (2012) specified that the execution probability of limit buy order is positively correlated with the bid-ask spread, the number of ask orders in an order book, market order arrival rate, and order cancellation rate. However, execution probability is negatively correlated with the distance from the opposite best price, the number of buy orders in the order book, and the limit-order arrival rate. Gava (2005) proposed several hypotheses to outline the relationship of limit-order execution time with variables like relative inside spread, price aggressiveness, market volatility, number of shares, trading activity, volume traded, order placing time, and type of last order placed. Palguna and Pollak (2016) studied multiple non-parametric methods to predict mid-price changes in LOBs.

EXECUTION TIME AND ITS DETERMINANTS

Yingsaeree (2012) suggested two types of execution time models: the first-passage time (FPT) model and the empirical execution time model. The first-passage time model is discussed in detail in Appendix A. In the empirical execution time model, limit-order execution times are modeled using actual execution time, which is the difference between the order placement and its complete execution for "time-to-fill."

Survival analysis is a statistical technique used for modeling the probability of occurrence and the timing of events (as a function of elapsed time), which is usually referred to as *survival time* (i.e., the *waiting time* until limit orders are executed) (Al-Suhaibani & Kryzanowski, 2000; Chatterjee & Mukhopadhyay, 2013; Cho & Nelling, 2000; Yingsaeree, 2012). In survival analysis, let T be a non-

negative random variable that represents survival time or time to failure in an order book data set (in our case, "failure" represents order execution). Let $F(.)$ be the *cumulative distribution function* (*c.d.f.*) of execution time T with corresponding *probability density function* (*p.d.f.*) $f(.)$. As $T \geq 0$, then

$$F(t) = Pr\{T \leq t\} = \int_0^t f(x)\, dx \qquad \text{Eq (1)}$$

The *survival function* represents the unconditional probability of surviving (i.e., "order not being executed") longer than t and is represented as:

$$S(t) = Pr(T \geq t) = 1 - F(t) = \int_t^\infty f(x)\, dx \qquad \text{Eq (2)}$$

The *survival function* $S(t)$ is a monotonic decreasing function expressed as follows:

$$S(0) = 1 \text{ and } S(\infty) = \lim_{t\to\infty} S(t) = 0$$

Therefore, the *p.d.f.* can be expressed as:

$$f(t) = \lim_{\Delta t\to 0^+} \frac{Pr\{t \leq T < t + \Delta t\}}{\Delta t} = \frac{dF(t)}{dt} = -\frac{dS(t)}{dt} \qquad \text{Eq (3)}$$

Survival distribution is estimated generally through parametric family, for the distribution of failure times, such as exponential distribution, Weibull distribution, or gamma distribution (Appendix B). In order to examine the effect of the explanatory variables on survival time, Accelerated Failure Time (AFT) and the Cox Proportional Hazard (CPH) models are used (Appendix C discusses the functions used in these two models). Al-Suhaibani and Kryzanowski (2000), Gava (2005), and Chatterjee and Mukhopadhyay (2013), using the AFT model, examined the relationship between the execution time and variables including the relative inside spread, price aggressiveness, market volatility, number of shares, trading activity, percentage of limit order executions on one side of the book, volume traded, order placing time, day of the week of the order placement, and type of last order placed.

LITERATURE GAP AND STUDY OBJECTIVES

To date, no study in the literature has discussed the distribution pattern of order book data or econometric models for estimation of execution probability and execution time. Nor has any research explored the reasons for the differences in the probabilities of bid and ask executions and their determinants. Moreover, too few studies have been carried out to provide information to traders on how to design their order books before entering the market so as to improve the probability of orders of executions. In this respect, the existing literature has hardly discussed the important role played by limit-order size, trading day, spread, depth, and volatility as useful variables for improving the limit-order execution. Furthermore, various determinants of execution probability and its degrees of explanatory power in different sectors of the economy remain insufficiently explored. At the same time, the distribution of execution times, execution probability, and the survival of limit orders in the market need to be examined using high-frequency order-driven data. Therefore, we propose the following objectives for this study:

a) to design an order-matching algorithm to construct a limit-order book (LOB),
b) to estimate the execution probability and execution time for bids and asks,

c) to analyze the explanatory power of various determinants of the execution probability of a limit order, and
d) to explore the survival functions and examine the determinants of the survival of limit orders.

EMPIRICAL DESIGN

ORDER-MATCHING ALGORITHM

The study has carried out order matching in each LOB for both the bid and ask sides. Exchanges work on carry forward LOB principle, and orders are carried forward if these are not executed. However, traders can modify unexecuted orders after a given duration. Exchanges follow batch processing of orders and generally establish fixed intervals for order matching. In NSE India, batch processing for order matching takes place over uniform time intervals; thus, in the present study, we have constructed LOBs in the intervals of 15-minutes.

We consider a time-uniform sample of limit orders in which all the limit orders in the LOB are present in every 15-minute interval. Within this 15-minute interval, fully executed orders are marked as terminated, whereas we continue to track partially executed or non-executed ones. After one hour, all partially filled orders are considered as reinserted, and at the end of the trading day, we cease tracking non-executed orders. We consider one LOB every hour, looking at a total of four intervals of 15 minutes within each LOB. This time-uniform sampling is designed as a per-batch processing criterion, and it does not introduce any specific bias in the sampling as new information is accumulated with every batch of new orders.

In this paper, the total LOB creation and limit-order execution are divided into six stages:

1. **Time structure:** Each trading day is divided into six separate LOBs. There is no carry forward of orders among the six LOBs. Each LOB is divided into four intervals. Non-executed orders in one interval are carried forward to the next interval but within the same LOB only.
2. **Interval structure:** Within each LOB, non-executed orders of one interval are carried forward to the next interval.
3. **Ordering structure:** As per the market concept of *best bid* (highest bid) and *best ask* (lowest ask), on both the bid and ask sides, ordering is done before the matching process.
4. **Searching:** Whenever one match is triggered either on the bid side or ask side, the search process loop stops. After the matching occurs and the flagging of the bid/ask order is completed, the process restarts with a new bid/ask order. However, if no match is found, then the loop continues.
5. **Matching:**
    a) When the bid price ≥ ask price and the bid volume ≤ ask volume, the bid is flagged as FTK ("fill to kill").
    b) When the bid price ≥ ask price and the bid volume > ask volume, the bid is flagged as PARTF ("partial fill").
    c) When the bid price ≥ ask price and the bid volume ≥ ask volume, the ask is flagged as FTK ("fill to kill").
    d) When the bid price ≥ ask price and the bid volume < ask volume, the ask is flagged as PARTF ("partial fill").
    e) When the bid price < ask price, the bid and ask are both flagged NON (non-executed).

On the ask side, matching is done based on the below conditions; subsequently, flagging is performed:

a) When the bid price ≥ ask price and the bid volume ≥ ask volume, the ask is flagged FTK ("fill to kill").
b) When the bid price ≥ ask price and the bid volume < ask volume, the ask is flagged PARTF ("partial fill").
c) When the bid price ≥ ask price and the bid volume ≤ ask volume, the bid is flagged FTK ("fill to kill").
d) When the bid price ≥ ask price and the bid volume > ask volume, the bid is flagged PARTF ("partial fill").
e) When the bid price < ask price, both the bid and ask are flagged NON (non-executed).

6. **Carry forward of non-executed orders:** Non-executed orders of one interval are ready to be carried forward to the next interval within the same LOB.
Probabilities are calculated in two ways:
i) In a specific interval, the probabilities are calculated consecutively for "fill to kill," "partial fill," and non-executed records for the ask and bid sides, respectively.
ii) Throughout the day, probabilities are estimated for each trade and marked as executed or non-executed. For the analysis, both "fill to kill" and "partial fill" flagged records are considered as executed (and assigned a value of 1), and non-executed records are assigned a value of 0.

This paper follows the carry-forward approach for order-book matching, as per stock exchange guidelines, and the algorithm designed for the same. The details of the algorithm are provided in Appendix D.

DETERMINANTS OF THE EXECUTION PROBABILITY OF LIMIT ORDER

As mentioned above, "fill to kill" and "partial fill" are considered as executed orders and assigned value of "1"; non-executed orders are assigned the value of "0". Following Omura et al. (2000), the execution probability is modeled through logistic regression. The article considers ask-side, $AExec_t$, and bid-side, $BExec_t$, execution probabilities separately for logit regression.

$$AExec_{it} = a_0 + a_1 ABook_{it} + a_2 BOpen_{it} + a_3 Tz_{it} + a_4 Spread_{it} + a_5 TickVolume_{it} + \varepsilon_{it}^{A}, \quad i = 1, \dots, n \text{ and } t = 1, \dots, N_A \qquad \text{Eq (4)}$$

$$BExec_{it} = b_0 + b_1 BBook_{it} + b_2 AOpen_{it} + b_3 Tz_{it} + b_4 Spread_{it} + b_5 TickVolume_{it} + \varepsilon_{it}^{B}, \quad i = 1, \dots, n \text{ and } t = 1, \dots, N_B \qquad \text{Eq (5)}$$

where N_A and N_B are the number of total submissions for the ask and bid limit orders for each stock i. The covariates are $ABook_{it}$ (the depth on the ask side of the book at time of execution t), $BBook_{it}$ (the depth on the bid side of the book at time of execution t), $AOpen_{it}$ (the number of open or available orders, depending on the number of existing limit orders in the order book on the ask side during execution t), $BOpen_{it}$ (the number of open or available bid orders during execution t), Tz_{it} (time zone, which indicates the number of intervals remaining before the close of the trading day), $TickVolume_{it}$ (the absolute limit-order size for quotes), and $Spread_{it}$ (measured as quoted bid-ask spread). $Book_{it}$ is calculated based on the number of available and open orders on the same side for submission t. $Open_{it}$ is calculated as the number of open orders on the opposite side for submission t. The depths $ABook_{it}$ and $BBook_{it}$ are the quantities of limit orders that are waiting at the best limit prices when an order is submitted at time t for stock i. Tz_{it} is computed as the number of intervals

remaining within each LOB at time t. The values will lie between 0, 1, 2, and 3 as each LOB of one hour is divided into four intervals of 15 minutes.

SURVIVAL FUNCTION FOR THE EXECUTION TIME OF LIMIT ORDER AND ITS DETERMINANTS

In the present study, three distributions are considered for estimating the survival function, and the best one was selected through the Akaike Information Criterion (AIC).

In the case of the Weibull distribution, the hazard rate is either monotonic increasing, monotonic decreasing, or constant with time. The survivor function is expressed as

$$S(t|\lambda, \kappa) = \exp\left(-(\lambda t)^{\kappa}\right) \qquad \text{Eq (6)}$$

where scale is $\lambda > 0$ and shape is $\kappa > 0$.

In accordance with the log-normal hazard function, the hazard increases from 0 at $t = 0$ to a maximum at the mean and then starts to decrease and approaches 0 as t becomes larger, with $\log(t)$ normally distributed with mean μ and variance σ^2. The survivor function is expressed as

$$S(t|\lambda, \alpha) = 1 - \Phi(\alpha \log(\lambda t)) \qquad \text{Eq (7)}$$

where $\mu = -\log(\lambda)$, $\sigma = \alpha^{-1}$, and $\Phi(.)$ is the cumulative distribution function of standard normal distribution.

In the case of log-logistic distribution, $\log(t)$ is in log-logistic distribution with location parameter μ and scale parameter σ. The survivor function is expressed as

$$S(t|\lambda, \alpha) = \frac{1}{1+(\lambda t)^{\alpha}} \qquad \text{Eq (8)}$$

where $\mu = -\log(\lambda)$ and $\sigma = \alpha^{-1}$.

The determinants of execution time are analyzed using the Accelerated Failure Time (AFT) model under parametric assumptions of the distributions (Escobar & Meeker, 2006). The vector of survival (execution) times T is the Weibull distributed with a scale σ and location $exp\ (X\beta)$.

$$log(T) = X\beta + \sigma\varepsilon \qquad \text{Eq (9)}$$

The expected time to failure can be expressed as:

$$E(T|X) = exp(X\beta)\Gamma(1 + \sigma) \qquad \text{Eq (10)}$$

where X represents the covariates.
The survivor function can be expressed as:

$$S(T|X) = exp\left(-exp(z)\right) \qquad \text{Eq (11)}$$

where $z = \frac{log(T)-X\beta}{\sigma}$.

The AFT Model for Log-Normal, when $\varepsilon \in N(0,1)$, can be expressed as follows:

$$log(T) = X\beta + \sigma\varepsilon \qquad \text{Eq (12)}$$

where T is the log-normal with parameters $X\beta$ and σ. The expected time to failure is expressed as

$$E(T|X) = exp\left(X\beta + \frac{\sigma^2}{2}\right) \qquad \text{Eq (13)}$$

and the survivor function as

$$S(T|X) = 1 - \Phi\left(\frac{log(T) - X\beta}{\sigma}\right) \qquad \text{Eq (14)}$$

where Φ is the standard normal c.d.f.

The AFT Model for Log-Logistic distribution can be expressed as:

$$log(T) = X\beta + \sigma\varepsilon \qquad \text{Eq (15)}$$

where T is log-logistic distributed with location parameter $X\beta$ and scale parameter σ.

Expected time to failure can be represented as

$$E(T|X) = exp\ (X\beta)\Gamma(1 + \sigma)\Gamma(1 - \sigma) \qquad \text{Eq (16)}$$

and the survivor function as

$$S(T|X) = 1 - F_{Logistic}\left(\frac{log(T) - X\beta}{\sigma}\right) \qquad \text{Eq (17)}$$

For empirical estimation, eight covariates and six dummy variables are used. The dummy variables represent sectors to capture the impact of individual sectors on survival time of limit-order execution. The determinants and dummy variables are as follows:

1. the limit-order size on the bid or ask side,
2. the remaining intervals (number of intervals remaining within each LOB; the values range among {3, 2, 1, 0}),
3. the quoted bid-ask spread,
4. the price aggressiveness is defined in terms of how quickly an order is executed:

$$Priceagr_t^{BID} = \frac{LimitPrice_t - BidPrice_{t-1}}{MidQuote_{t-1}} \qquad \text{Eq (18)}$$

$$Priceagr_t^{OFR} = \frac{OfferPrice_{t-1} - LimitPrice_t}{MidQuote_{t-1}} \qquad \text{Eq (19)}$$

5. same-side order book depth and opposite side order book depth,
6. return volatility,
7. inferior price, which is measured as the difference between the mid-quote price and the trade price, and
8. the sector k dummy variable [k = 1 … 6], representing sector-specific identification for each sector.

STUDY PERIOD AND DATA

Tick-by-tick data for the 30 selected stocks listed on the NSE-CNX-500 for the month of June 2016 (22 trading days) were collected from Bloomberg servers and processed as large-scale databases. The dataset consists of stocks from dominating sectors of the Indian economy, such as consumer goods, financial services, information technology (IT) and telecom services, services and healthcare, pharmaceuticals, and automobile and industrial manufacturing.

RESULTS AND DISCUSSION

As shown in Table 1, the findings reveal that the sample data satisfied all the seven *V*s of big data – volume, velocity (tick-by-tick orders), variety, variability, veracity, visualization, and value – because of the inclusion of multiple orders within a very short period of time for each stock. Bid- and ask- side analyses have been conducted to capture the behavior and information absorption capacity of different sides of trades.

Table 1. Tick Counts on Buyer and Seller Sides

Dataset	Dataset	Buyer or Bid Side	Seller or Ask Side
Total	*Original No. of Records*	*4,566,944*	*4,612,919*
	No. of Records after Cleaning	*4,149,703*	*4,187,685*
Consumer Goods Sector	*Original No. of Records*	*636,250*	*679,420*
	No. of Records after Cleaning	*572,582*	*613,037*
Financial Services Sector	*Original No. of Records*	*1,071,275*	*1,060,147*
	No. of Records after Cleaning	*981,526*	*970,232*
IT-Telecom Sector	*Original No. of Records*	*1,074,517*	*1,085,061*
	No. of Records after Cleaning	*979,362*	*989,777*
Non-financial Services Sector	*Original No. of Records*	*272,158*	*257,901*
	No. of Records after Cleaning	*246,197*	*232,679*
Pharmaceuticals Sector	*Original No. of Records*	*695,854*	*718,810*
	No. of Records after Cleaning	*630,865*	*650,802*
Automobile & Industrial Manufacturing Sector	*Original No. of Records*	*816,890*	*811,580*
	No. of Records after Cleaning	*739,171*	*731,158*

Note: Representation of the total order counts on both buyer and seller sides, on which the algorithm was applied, and modeling was performed. The table displays the original record counts and counts after cleaning.

The data period covered both bullish and bearish market conditions (see Figure 1 below). Table 2 presents the Chow test results to confirm the existence of bullish and bearish market conditions in the data set. The study conducted the Chow test on both sell orders and buy orders with the null hypothesis of the non-existence of structural breaks.

Table 2. Results from Chow Test—Structural Break

Order Type	Chow Test
Buy Orders	*Chow test (273,100):* F = 554.11, *p-value* < 2.2e-16
Sell Orders	*Chow test (270,161):* F = 581.19, *p-value* < 2.2e-16

Note. Chow test results on both buy and sell orders to confirm the existence of bullish and bearish market conditions in the dataset. The null hypothesis states that there are no structural breaks.

Since the p-value is less than 0.05 for both the cases, we can reject the null hypothesis and conclude that the sample data for both buy and sell orders have structural breaks and, thus, have concurrent bullish and bearish conditions.

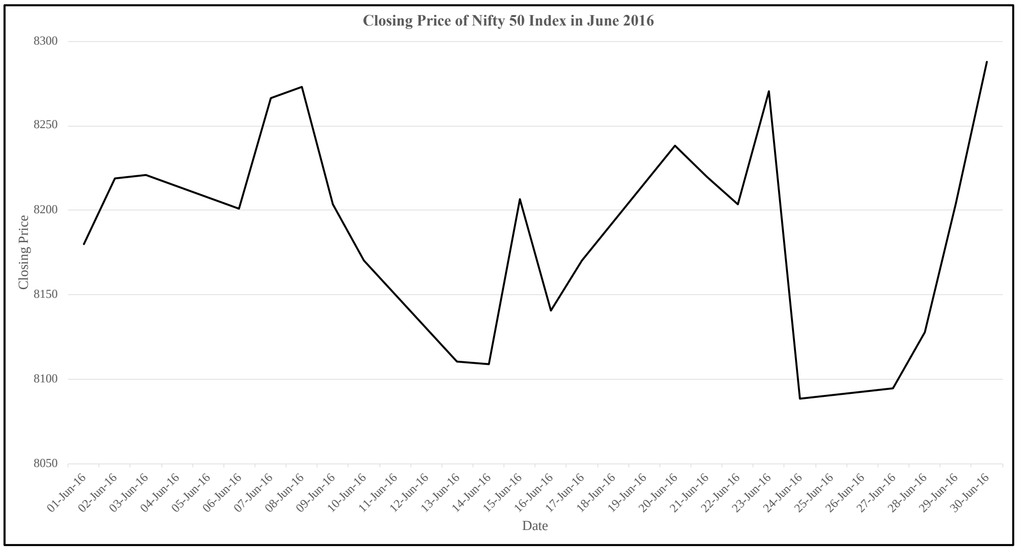

Figure 1. Closing Price Movement of Nifty 50 Index in June 2016

EXECUTION PROBABILITY OF LIMIT ORDERS – BID AND ASK

Execution probability is estimated as the ratio of the total number of executed orders (fill-to-kill and partial fill) to the total number of outstanding orders. We compute the execution probability for both buy and sell limit orders for 30 stocks in every 15-minute interval. The sectoral average probability of execution is computed to show sector-level dominance. Among the stocks, Axis Bank has the highest average execution probability on both the bid and ask sides. However, among sectors, the IT-Telecom sector surpasses all other sectors in the execution probability on both the bid and ask sides. Table 3 presents the estimated average probability of execution for each stock in the sample for June 2016.

Table 3. Buyer and Seller Execution Probability: Stock and Sector Level Period—June 2016

Sector	Stock	Average Buyer Side Probability of Execution	Average Seller Side Probability of Execution
Consumer Goods	Shree Renuka Sugars Ltd.	0.302 (0.273, 0.344)	0.302 (0.276, 0.333)
	I T C Ltd.	0.439 (0.424, 0.458)	0.439 (0.424, 0.457)
	Britannia Industries Ltd.	0.372 (0.352, 0.398)	0.378 (0.358, 0.401)
	Bajaj Electricals Ltd	0.31 (0.287, 0.336)	0.324 (0.303, 0.345)
	Procter & Gamble Hygiene & Health Care Ltd.	0.131 (0.103, 0.158)	0.132 (0.107, 0.146)
	Sectoral Average	**0.317 (0.304, 0.333)**	**0.321 (0.310, 0.334)**
Financial Services	State Bank of India	0.444 (0.428, 0.464)	0.446 (0.435, 0.459)
	Axis Bank Ltd.	0.455 (0.439, 0.475)	0.459 (0.447, 0.474)
	Kotak Mahindra Bank Ltd.	0.421 (0.404, 0.442)	0.408 (0.392, 0.426)
	SREI Infrastructure Finance Ltd.	0.369 (0.352, 0.389)	0.379 (0.365, 0.394)
	ICRA Ltd.	0.088 (0.063, 0.122)	0.092 (0.063, 0.112)
	Sectoral Average	**0.374 (0.363, 0.387)**	**0.376 (0.368, 0.386)**
IT-Telecom	Reliance Communications Ltd.	0.377 (0.359, 0.4)	0.394 (0.379, 0.411)
	Infosys Ltd.	0.444 (0.431, 0.46)	0.451 (0.434, 0.473)
	Tata Consultancy Services Ltd.	0.447 (0.43, 0.467)	0.439 (0.425, 0.457)
	Tata Communications Ltd.	0.368 (0.347, 0.395)	0.364 (0.346, 0.385)
	Oracle Financial Services Software Ltd.	0.296 (0.272, 0.322)	0.289 (0.267, 0.314)
	Sectoral Average	**0.384 (0.373, 0.396)**	**0.385 (0.374, 0.396)**
Non-financial Services	*Jaypee Infratech Ltd.*	*0.222 (0.187, 0.267)*	*0.229 (0.196, 0.269)*
	Fortis Healthcare Ltd.	*0.334 (0.315, 0.354)*	*0.349 (0.329, 0.372)*
	Hotel Leela Venture Ltd.	*0.212 (0.182, 0.236)*	*0.229 (0.198, 0.249)*
	Apollo Hospitals Enterprises Ltd.	*0.363 (0.349, 0.380)*	*0.37 (0.354, 0.389)*
	3M India Ltd.	*0.087 (0.069, 0.113)*	*0.103 (0.084, 0.129)*
	Sectoral Average	***0.26 (0.246, 0.276)***	***0.275 (0.263, 0.287)***
Pharma	*Sun Pharmaceutical Industries Ltd.*	*0.433 (0.414, 0.457)*	*0.439 (0.424, 0.458)*
	Cipla Ltd.	*0.42 (0.399, 0.449)*	*0.413 (0.398, 0.431)*
	Bliss GVS Pharma Ltd.	*0.32 (0.29, 0.343)*	*0.323 (0.294, 0.346)*
	Strides Shasun Ltd.	*0.354 (0.337, 0.373)*	*0.364 (0.349, 0.381)*
	GlaxoSmithKline Pharmaceuticals Ltd.	*0.186 (0.159, 0.211)*	*0.194 (0.165, 0.22)*
	Sectoral Average	***0.343 (0.332, 0.353)***	***0.348 (0.339, 0.356)***

Sector	Stock	Average Buyer Side Probability of Execution	Average Seller Side Probability of Execution
Automobile & Industrial Manufacturing	*Suzlon Energy Ltd.*	*0.229 (0.201, 0.266)*	*0.23 (0.204, 0.262)*
	Tata Motors Ltd.	*0.45 (0.434, 0.47)*	*0.45 (0.434, 0.47)*
	Bajaj Auto Ltd.	*0.398 (0.383, 0.416)*	*0.38 (0.367, 0.394)*
	Siemens Ltd.	*0.369 (0.356, 0.384)*	*0.359 (0.348, 0.373)*
	Honeywell Automation India Ltd.	*0.087 (0.066, 0.102)*	*0.09 (0.07, 0.107)*
	Sectoral Average	***0.331 (0.319, 0.345)***	***0.328 (0.317, 0.341)***

Note: Estimated average probability of execution for each stock considered in the study for the period of June 2016. Notably, sector-wise averages are also computed to show sector-level dominance among different sectors. Execution probabilities were calculated as the ratio between the total number of executed (fill-to-kill and partial fill) orders and the total number of orders. In parentheses, separated by a comma, are the daily lower and upper CI of the daily execution probability using t distributions as the sample sizes of the daily execution probability for each stock and each sector are small in size (estimated figures).

DETERMINANTS OF EXECUTION PROBABILITY OF LIMIT-ORDER BOOK

As discussed in the empirical design section, determinants for the execution probability of LOB were analyzed for both the bid and ask sides. Since the data possess big data properties, the big-GLM model (Lumley, 2011) is deployed to estimate the logistic regression. The logistic regression on execution probability (Table 4), along with its determinants, are found to be significant and consistent with previous literature (Al-Suhaibani & Kryzanowski, 2000; Cho & Nelling, 2000; Omura et al., 2000; Lo et al., 2002; Gava, 2005; Yingsaeree, 2012).

Table 4. Determinants of Probability of Limit Order Execution for Both Buyer and Seller Sides

Coefficients	Buyer	Seller
Intercept	*-0.1471**** *(-10.02)*	*-0.0214* *(-1.440)*
Same-Side Book Depth	*-0.0287**** *(-89.35)*	*-0.0266**** *(-82.133)*
Opposite-Side Open or Available Orders	*0.015**** *(52.68)*	*0.0096**** *(34.058)*
Remaining Trading Time (in Intervals)	*-0.0151**** *(-3.60)*	*-0.0402**** *(-9.554)*
Bid-Ask Spread	*0.0145**** *(25.77)*	*0.0213**** *(31.609)*
Limit Order Size	*-0.0001**** *(-26.28)*	*-0.0001**** *(-24.428)*
Number of Cases Correctly Predicted	*66.43%*	*65.75%*

Note: Method: Logistic bigGLM; carry-forward LOB. Period: June 2016. Dependent variable: execution probability. The figures in parentheses represent t-ratios (estimated).
* $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$

The results show that a negative relationship between bid-side execution probability and buying volume which indicates that execution probability can be improved using smaller-sized limit orders (Harris, 1996; Omura et al., 2000). The inverse relation between the depth of the book on the bid side (same side) and the bid-side execution probability is consistent with the findings of Omura et al. (2000) and Yingsaeree (2012). The positive relation between the ask-side (opposite side) open indicator and the bid-side execution probability is consistent with the findings of Omura et al. (2000) and Yingsaeree (2012). As expected, the remaining trading time (in intervals) has a significant negative relationship with the bid-side execution probability. This implies that for the last intervals of each LOB, the bid-side execution probability will be higher than for previous intervals. A direct positive relationship between the bid-ask spread and bid-side execution probability is consistent with the findings of Yingsaeree (2012). The findings for the determinants of execution probability on the ask-side are similar to those for the bid side.

SECTORAL ANALYSIS OF EXECUTION PROBABILITY

A sectoral analysis of the execution probability is carried out to understand the sector-specific impact on the probability of LOB execution (see Tables 5 and 6 below). Except for a few cases, the overall results are similar across the sectors. However, the remaining trading time (in intervals) is directly related to the bid-side execution probability in all sectors except consumer goods. The bid-ask spread exhibits an inverse relationship with bid-side execution probability in the cases of financial services and the services and healthcare sectors, unlike in other sectors. With a larger bid-ask spread, the probability of execution decreases for stocks belonging to financial, non-financial services, and healthcare sectors. Trading volume is found to have a positive relationship with bid-side execution probability in the cases of financial services, IT-telecom, and pharmaceuticals sectors, thus indicating a higher execution probability with a large volume, unlike in other sectors. The results of sectoral analysis on the ask-side are similar to those of the bid-side. In addition, in the case of the pharmaceuticals sector, the remaining trading time (in intervals) is found to be in an inverse relationship with the ask-side execution probability. The results of sector-wise analysis on both sides imply that the explanatory variables do not behave in a similar way in the different sectors; hence, traders should employ different strategies while dealing with different sectors.

Table 5. Sector-wise Impact of Determinants on Buyer-Side Limit Order Execution Probability

	Consumer Goods	Financial Services	IT-Telecom	Non-Financial Services & Healthcare	Pharma	Automobile & Industrial Manufacturing
Intercept	*0.3056**** *(8.544)*	*-0.0884*** *(-2.406)*	*-0.3299**** *(-8.676)*	*-0.8025**** *(-20.566)*	*-0.3381**** *(-9.175)*	*-0.3500**** *(-9.522)*
Same-Side Book Depth	*-0.0328**** *(-41.902)*	*-0.0307**** *(-38.315)*	*-0.0253**** *(-30.336)*	*-0.0236**** *(-28.711)*	*-0.02.78**** *(-35.569)*	*-0.0245**** *(-30.027)*
Opposite-Side Open or Available Orders	*0.0066**** *(9.251)*	*0.0187**** *(25.487)*	*0.0116**** *(16.826)*	*0.0192**** *(26.415)*	*0.0143**** *(20.153)*	*0.0166**** *(23.625)*
Remaining Trading Time (in Intervals)	*-0.1449**** *(-14.214)*	*0.0338**** *(3.292)*	*0.0669**** *(6.436)*	*0.0345**** *(3.055)*	*0.0129* *(1.253)*	*0.0271*** *(2.562)*
Bid-Ask Spread	*0.0467**** *(24.414)*	*-0.0240**** *(-9.284)*	*0.2232**** *(58.443)*	*-0.0057**** *(-6.185)*	*0.1618**** *(44.421)*	*0.0165**** *(11.699)*
Limit Order Size	*-0.0001**** *(-3.302)*	*0.0001** *(2.146)*	*0.0001**** *(5.289)*	*-0.0001**** *(-9.679)*	*0.0001**** *(6.944)*	*-0.0001**** *(-24.450)*
Number of Cases Correctly Predicted	*68.62%*	*62.76%*	*65.89%*	*74.01%*	*66.14%*	*66.72%*

Note: Sector-wise impact of determinants (explanatory variables) on the buyer side probability of limit order execution. Method: Logistic bigGLM; Period: June 2016. Dependent variable: buyer side execution probability. The figures in parentheses represent the t-ratios (estimated).
* p ≤ 0.05; ** p ≤ 0.01; *** p ≤ 0.001.

Table 6. Sector-wise Impact of Determinants on Seller-Side Limit Order Execution Probability

	Consumer Goods	**Financial Services**	**IT-Telecom**	**Non-Financial Services & Healthcare**	**Pharma**	**Automobile & Industrial Manufacturing**
Intercept	*0.5384**** *(14.796)*	*-0.0993*** *(-2.689)*	*-0.2627**** *(-6.810)*	*-0.5648**** *(-14.433)*	*0.0845** *(2.255)*	*-0.5211**** *(-14.049)*
Same-Side Book Depth	*-0.0373**** *(-45.307)*	*-0.0239**** *(-29.377)*	*-0.0332**** *(-40.442)*	*-0.0173**** *(-21.371)*	*-0.0316**** *(-39.019)*	*-0.0199**** *(-24.828)*
Opposite-Side Open or Available Orders	*0.0069**** *(10.306)*	*0.0119**** *(16.783)*	*0.0197**** *(26.735)*	*0.0088**** *(11.684)*	*0.0081**** *(12.039)*	*0.0156**** *(21.527)*
Remaining Trading Time (in Intervals)	*-0.1964**** *(-19.127)*	*0.0422* *(4.087)*	*0.0385**** *(3.678)*	*0.0219** *(1.953)*	*-0.0831**** *(-8.013)*	*0.0595**** *(5.541)*
Bid-Ask Spread	*0.0482**** *(22.691)*	*-0.0285* *(-9.451)*	*0.2264**** *(59.260)*	*-0.0076**** *(-7.146)*	*0.1672**** *(44.365)*	*0.0184**** *(13.025)*
Limit Order Size	*-0.0001**** *(-12.007)*	*0.0001* *(0.296)*	*0.0001** *(2.466)*	*-0.0001**** *(-28.141)*	*0.0001**** *(8.139)*	*-0.0001* *(-23.171)*
Number of Cases Correctly Predicted	*68.34%*	*62.23%*	*65.93%*	*72.51%*	*66.02%*	*66.74%*

Note: Sector-wise impact of determinants (explanatory variables) on the seller side probability of limit order execution. Method: Logistic bigGLM; Period: June 2016. Dependent variable: seller side execution probability. The figures in parentheses represent the t-ratios (estimated).
* $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$.

DETERMINANTS OF EXECUTION TIME AND SURVIVAL ANALYSIS

The behavior of the determinants of execution times is examined using survival analysis. To identify the appropriate distribution for the survival function, three parametric distributions are estimated, and the log-normal distribution is selected as the best fit on the basis of minimum AIC value (see Table 7 and Figures 2 and 3).

In the log-normal distribution on both the bid and ask sides, the regressors have a very similar relationship with failure time, except in cases of price aggressiveness and inferior price (see Table 8). For the bid-side, inferior price is significant and negative, which implies that execution probability decreases as inferior price increases. This is due to the increasing gap between the mid-quote and trade price, which resembles less trading activity because of the minimal presence of ask orders in the market.

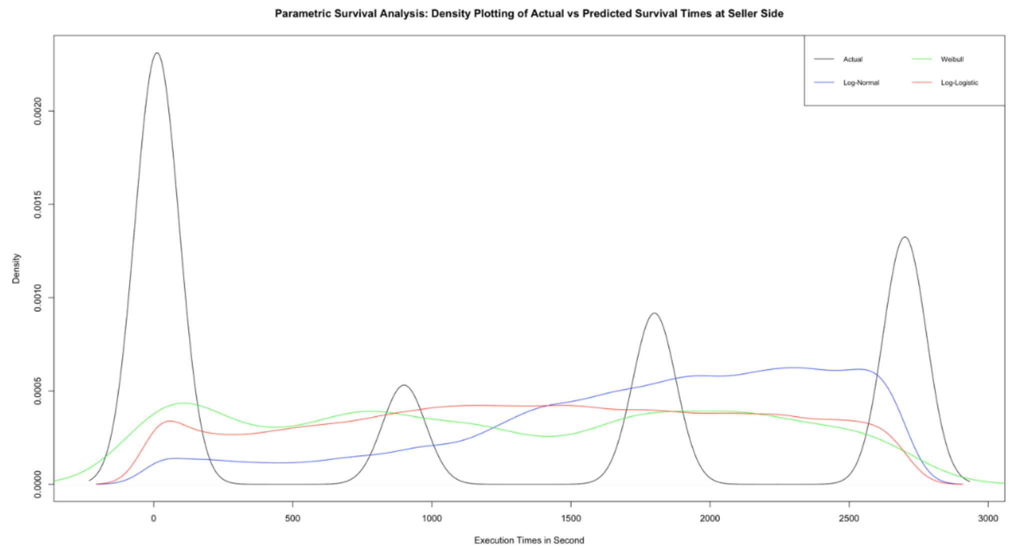

Figure 2. Parametric Survival Analysis: Densities of Fitted Probability Distributions on Seller (Ask) Side

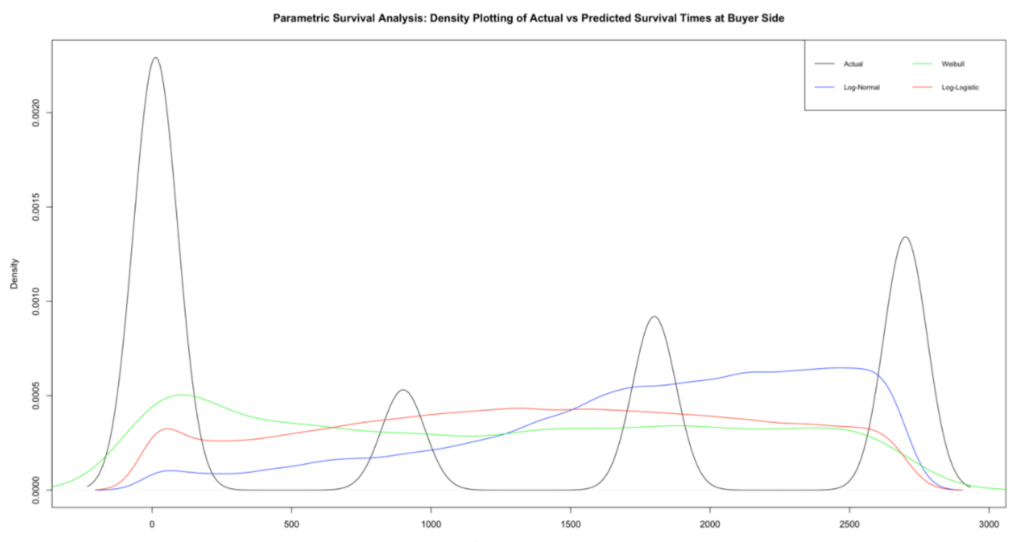

Figure 3. Parametric Survival Analysis: Densities of Fitted Probability Distributions on Buyer (Bid) Side

Table 7. Survival Analysis: Model Selection Criteria Showing Comparative AIC Values* for Selection of Distribution for Parametric Survival Analysis

	Buyer Side			Seller Side		
	Weibull Distribution	Log-Logistic Distribution	Log-Normal Distribution	Weibull Distribution	Log-Logistic Distribution	Log-Normal Distribution
AIC	1,820,933	1,826,705	**1,812,315**	1,553,810	1,561,769	**1,546,193**

* Estimated figures.

Table 8. Analysis of Determinants: Buyer- and Seller-Side Pooled Data Showing the Explanatory Variables Impacting Survival Time (Execution Time) for Both Buyer and Seller Sides

	Buyer Side Log-Normal Distribution	**Seller Side Log-Normal Distribution**
Constant	*−0.3056****	*2.1489****
	(−3.12)	*(18.82)*
Limit Order Size	*0.0001****	*0.0001****
	(14.38)	*(10.29)*
Remaining Trading Time (in Intervals)	*−2.5394****	*−2.8786****
	(−208.2)	*(−202.0)*
Quoted Spread	*−0.0706****	*−0.0923****
	(−40.83)	*(−44.66)*
Price Aggressiveness	*0.0022*	*−0.0179*
	(0.10)	*(−0.56)*
Same-Side Book Depth	*2.5895****	*1.7185****
	(96.81)	*(55.37)*
Opposite-Side Book Depth	*−0.0445****	*−0.0281****
	(−64.56)	*(−34.96)*
Return Volatility	*0.0918*	*0.0348*
	(1.33)	*(0.22)*
Inferior Price	*−0.0249****	*0.0089****
	(−20.55)	*(5.79)*
Sector 1 Dummy Variable	*0.3718****	*0.3989****
	(13.79)	*(12.67)*
Sector 3 Dummy Variable	*−0.0289*	*−0.0164*
	(−1.12)	*(−0.54)*
Sector 4 Dummy Variable	*0.7922****	*0.8890****
	(27.79)	*(26.60)*
Sector 5 Dummy Variable	*0.2263****	*0.2259****
	(8.53)	*(7.27)*
Sector 6 Dummy Variable	*0.0848****	*0.2252****
	(2.99)	*(6.79)*

Note: Figures in parentheses represent t-ratios.
* $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$.

However, limit-order size, same-side order book depth, and return volatility have a positive relationship with limit order execution time. With the increase in the size of the limit orders, waiting time to execution also increases. With large-volume orders, in the same-side book, chances of execution decrease, which, in turn, implies more waiting time in the system before execution.

Regarding carry-forward orders within each LOB in a day, the greater the number of intervals that an order remains active, the higher is its chance of being executed. Therefore, as intervals pass by, the remaining interval count decreases, execution probability increases, and, hence, waiting time in the system decreases. Similarly, for opposite-side order book depth, with a greater number of quotes from the opposite-side order book, execution probability increases, which, in turn, decreases the waiting time in the system. Opposite-side book depth exhibits a positive relationship with execution

probability (on both the bid and ask sides) but in inverse relationship with execution times (on both the bid and ask sides). The analysis revealed that with more depth in the opposite-side book, time to execution decreases, which, in turn, increases execution probability.

In this study, five sector dummy variables were deployed with respect to the baseline dummy, which is the financial services sector. A positive coefficient on the sector dummy in the cases of consumer goods, services and healthcare, pharmaceuticals, and automobile and industrial manufacturing means, with reference to the financial services sector, waiting time to execution is larger. The findings also show that execution time in the financial services sector is the lowest, on both the buy and ask sides, compared to all other sectors (cf. Chatterjee & Mukhopadhyay, 2013).

SUMMARY OF FINDINGS

The analysis confirmed that execution probability and time taken for execution are different for different stocks and sectors of the economy. Execution probability is highest in the case of the IT-telecom sector; execution time is lowest in the case of the financial services sector. The determinants of execution probability and execution time have different degrees of explanatory power in different sectors. The study found that the bid-ask spread is a major explanatory variable in limit-order execution. The differential behavior of execution probability for different sectors calls for sector-specific policy formulation to increase respective execution probabilities.

The results also indicate that the order-book depth, the number of opposite-side open orders, remaining trading time (in intervals), bid-ask spread, and limit-order size are significant explanatory variables for execution probability. The execution probability on the bid side can be improved by placing small, fragmented orders. The execution probability on the bid side will be further enhanced if sufficient ask-side orders come into the market; similarly, ask-side probability can be increased if the market receives more bid-side orders.

The study also found that the log-normal distribution is the best-fitting model for the survival function. The determinant analysis of the survival of limit orders (using the log-normal model) indicates that the limit-order size, remaining trading time (in intervals), quoted spread, price aggressiveness, book depth, return volatility, and inferior price have major explanatory power for limit order to survive.

CONCLUSIONS AND POLICY IMPLICATIONS

The execution of LOBs is an evolving research area of market microstructure. The estimation of execution rate and the probability of limit orders execution have been significantly explored using low-frequency data. Our study conducted similar research using high-frequency data, with 30 stocks selected from the CNX 500 Index across six sectors listed on the NSE India. We computed the execution probability of limit orders (both buy orders and sell orders) with uniformly sampled time intervals of 15 minutes, as per the batch-processing criterion of India's NSE. The execution times of limit orders were estimated using parametric survival analysis and the distribution of execution times were examined for different sectors.

This study has practical implications for traders using LOBs. Such traders should look at the execution probabilities and execution times before entering the market so as to improve their order execution rate. This implies that traders should design strategies using limit-order size, remaining intervals in a trading day, spread, depth, and volatility as variables to decrease the waiting time for limit-order execution. Similarly, traders should initiate fragmented, smaller-sized orders to improve their executions, as smaller-sized LOBs have a higher execution probability. Determinants of execution probability have different degrees of explanatory power in different sectors of the economy. In this

context, exchanges should design sector-specific policies to improve the trading volume and frequency of trading. The findings indicate that execution probability improves with smaller-sized limit orders; therefore, exchanges should regulate higher-volume LOBs during retail trading hours.

Further research could consider the non-linearity aspect of financial time series, which has not been considered in this work. The present study was confined to the cash market and ignored the implications of options and other derivative segments on LOB modeling, which is another potential area of research. Finally, the use of non-parametric distribution for survival analysis and the implementation of machine-learning algorithms in survival analysis could provide more information on modeling limit-order execution times.

REFERENCES

Abergel, F., Anane, M., Chakraborti, A., Jedidi, A., & Toke, I. M. (2016). *Limit order books*. Cambridge, UK: Cambridge University Press.

Al-Suhaibani, M., & Kryzanowski, L. (2000). An exploratory analysis of the order book, and order flow and execution on the Saudi stock market. *Journal of Banking & Finance*, *24*(8), 1323–1357. https://doi.org/10.1016/S0378-4266(99)00075-8

Battalio, R., Greene, J., & Jennings, R. (1997). Do competing specialists and preferencing dealers affect market quality? *The Review of Financial Studies*, *10*(4), 969–993. https://www.jstor.org/stable/2962336

Bouchaud, J. P., Bonart, J., Donier, J., & Gould, M. (2018). *Trades, quotes and prices: Financial Markets Under the Microscope*. Cambridge, UK: Cambridge University Press.

Chatterjee, D., & Mukhopadhyay, C. (2013). Execution times of small limit orders: A simpler modeling approach. *Vikalpa*, *38*(1), 49. https://doi.org/10.1177/0256090920130105

Cho, J. W., & Nelling, E. (2000). The probability of limit-order execution. *Financial Analysts Journal*, 56(5), 28–33. https://doi.org/10.2469/faj.v56.n5.2387

Chung, K. H., Van Ness, B. F., & Van Ness, R. A. (1999). Limit orders and the bid–ask spread. *Journal of Financial Economics*, *53*(2), 255–287. https://doi.org/10.1016/S0304-405X(99)00022-7

Collett, D. (2015). *Modelling survival data in medical research*. Boca Raton, FL: CRC press.

Eisler, Z., Kertesz, J., Lillo, F., & Mantegna, R. N. (2009). Diffusive behavior and the modeling of characteristic times in limit order executions. *Quantitative Finance*, 9(5), 547–563. https://doi.org/10.1080/14697680802512408

Escobar, L. A., & Meeker, W. Q. (2006). A review of accelerated test models. *Statistical Science*, *21*(4), 552–577. https://doi.org/10.1214/088342306000000321

Foucault, T. (1999). Order flow composition and trading costs in a dynamic limit order market. *Journal of Financial Markets*, *2*(2), 99–134. https://doi.org/10.1016/S1386-4181(98)00012-3

Foucault, T., Kadan, O., & Kandel, E. (2005). Limit order book as a market for liquidity. *Review of Financial Studies*, *18*(4), 1171–1217.

Gava, L. (2005). *The speed of limit order execution in the Spanish stock exchange* [Working paper wb057718]. Universidad Carlos III de Madrid. https://ideas.repec.org/p/cte/wbrepe/wb057718.html

Handa, P., & Schwartz, R. A. (1996). Limit order trading. *The Journal of Finance*, 51(5), 1835–1861. https://doi.org/10.2307/2329540

Harris, L. (1996). *Does a large minimum price variation encourage order exposure?* [Working paper]. University of South California.

Hollifield, B., Sandas, P., & Miller, R. A. (2001). *Liquidity supply and demand: Empirical evidence from the Vancouver stock exchange* [Working paper]. Rodney L. White Center for Financial Research.

Hosmer Jr, D. W., Lemeshow, S., & May, S. (2011). *Applied survival analysis: Regression modeling of time-to-event data (vol. 618)*. Medford, MA: John Wiley & Sons.

Huang, W., Lehalle, C. A., & Rosenbaum, M. (2015). Simulating and analyzing order book data: The queue-reactive model. *Journal of the American Statistical Association*, 110(509), 107–122. https://doi.org/10.1080/01621459.2014.982278

Khanal, S. P., Sreenivas, V., & Acharya, S. K. (2014). Accelerated failure time models: An application in the survival of acute liver failure patients in India. *International Journal of Science and Research*, 3, 161–166.

Lo, A. W., MacKinlay, A. C., & Zhang, J. (2002). Econometric models of limit-order executions. *Journal of Financial Economics*, 65(1), 31–71. https://doi.org/10.1016/S0304-405X(02)00134-4

Lumley, T. (2011). Biglm: Bounded memory linear and generalized linear models. R package version 0.8. http://cran.r-project.org/web/packages/biglm

Nath, G. C., & Dalvi, M. (2004). Day of the week effect and market efficiency-Evidence from Indian equity market using high frequency data of national stock exchange. http://dx.doi.org/10.2139/ssrn.1092765

Omura, K., Tanigawa, Y., & Uno, J. (2000). Execution probability of limit orders on the Tokyo Stock Exchange. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=252588

Palguna, D., & Pollak, I. (2016). Mid-price prediction in a limit order book. *IEEE Journal of Selected Topics in Signal Processing*, *10*(6), 1083–1092. https://doi.org/10.1109/JSTSP.2016.2593060

Parlour, C. A. (1998). Price dynamics in limit order markets. *Review of Financial Studies*, 11(4), 789–816. https://doi.org/10.1093/rfs/11.4.789

Raudys, A., & Matkėnaitė, S. (2016). Analysis of execution methods in US and European futures. *The Journal of Trading*, *11*(1), 38–52. https://doi.org/10.3905/jot.2016.11.1.038

Smith, E., Farmer, J. D., Gillemot, L. S., & Krishnamurthy, S. (2003). Statistical theory of the continuous double auction. *Quantitative Finance*, *3*(6), 481–514. https://doi.org/10.1088/1469-7688/3/6/307

Wen, Q. (2008). Econometric models of limit-order completion time. Vast Data Conference, Oxford-Man Institute of Quantitative Finance, September 2008.

Yingsaeree, C. (2012). *Algorithmic trading: Model of execution probability and order placement strategy* [Doctoral dissertation]. University College London.

APPENDICES

APPENDIX A

FIRST-PASSAGE TIME MODELS (FPT)

The execution time of a limit order can be calculated from the first time that the price of an asset reaches or crosses the limit-order price; that is, the FPT. For an asset, with price p_0 at time $\mathrm{t} = 0$, the FPT through a prescribed level $\mathrm{p}_0 + \Delta$ for a fixed distance $\Delta > 0$ is defined as the first time t when $\mathrm{p(t)} \geq \mathrm{p}_0 + \Delta$. When the buy (or sell) order is at the top of the queue then "time to first fill" (TTFF) will actually be the FPT which will work as a lower bound. Yingsaeree (2012) proposed the relation to be:

$$\mathrm{FPT}^{+}(\Delta) \leq \mathrm{TTFF} \leq \mathrm{TTF} \leq \mathrm{FPT}^{+}(\Delta + \varepsilon) \quad \text{Eq (A.1)}$$

where ε is the price difference of the asset and TTF is "$time\ to\ fill.$" The FPT can be modeled by both theoretical and empirical approaches. Lo et al. (2002) explored a theoretical model of FPTs using the Brownian approach and an empirical approach from transaction data before performing the survival analysis. But in both cases, Lo et al. (2002) showed some limitations, such as how the FPT model does not have the ability to distinguish between "time to first fill," "time to completion," and "time to censoring."

APPENDIX B

PARAMETRIC AND NON-PARAMETRIC APPROACHES FOR SURVIVAL ANALYSIS

PARAMETRIC METHODS

Parametric methods assume a specific family of survival distribution and estimate its parameters using a maximum likelihood estimator (Abergel *et al.*, 2016). As discussed by Yingsaeree (2012), any distributions over non-negative values can be used in the modeling of survival distribution, but this paper considers three major models that have been generally used in past research: Weibull distribution, log-normal distribution, and log-logistic distribution.

Weibull Distribution

In the case of Weibull distribution, the hazard rate is either monotonic increasing, monotonic decreasing, or constant with time. In the case of scale $\lambda > 0$ and shape $\kappa > 0$, the probability density function is represented as

$$f(t|\lambda,\kappa) = \lambda\kappa(\lambda t)^{\kappa-1}\exp\left(-(\lambda t)^{\kappa}\right) \quad \text{Eq (B.1)}$$

and the survivor function as

$$S(t|\lambda,\kappa) = \exp\left(-(\lambda t)^{\kappa}\right) \quad \text{Eq (B.2)}$$

The Weibull hazard function is monotonically increasing if $\kappa > 1$ and monotonically decreasing when $\kappa < 1$. This parameter κ is called the shape parameter. On the other hand, λ is called the scale

parameter because it can only change scale over the horizontal axis but does not have any impact on the shape of the curve.

Log-Normal Distribution

In the case of the log-normal hazard function, the hazard increases from 0 at $t = 0$ to a maximum at the mean and then starts to decrease and approaches 0 as t becomes larger. In this distribution, $\log(t)$ is normally distributed with mean μ and variance σ^2, the probability density function is expressed as

$$f(t|\lambda, \alpha) = (2\pi)^{-\frac{1}{2}}\alpha t^{-1}\exp\left(\frac{-\alpha^2(\log(\lambda t))^2}{2}\right) \quad \text{Eq (B.3)}$$

and the survivor function as

$$S(t|\lambda, \alpha) = 1 - \Phi(\alpha \log(\lambda t)) \quad \text{Eq (B.4)}$$

where $\mu = -\log(\lambda)$, $\sigma = \alpha^{-1}$, and $\Phi(.)$ is the cumulative distribution function of the standard normal distribution. In this distribution, the hazard function is non-monotonic. The log-normal hazard function has an inverted U-shape.

Log Logistic Distribution

The log-logistic hazard function also has an inverted U-shape. In this distribution, $\log(t)$ is in log-logistic distribution with the location parameter μ and the scale parameter σ, the probability density function is represented as

$$f(t|\lambda, \alpha) = \frac{\alpha t^{\alpha-1}\lambda^{\alpha}}{(1+(\lambda t)^{\alpha})^2} \quad \text{Eq (B.5)}$$

and the survivor function as

$$S(t|\lambda, \alpha) = \frac{1}{1+(\lambda t)^{\alpha}} \quad \text{Eq (B.6)}$$

where $\mu = -\log(\lambda)$ and $\sigma = \alpha^{-1}$.

NON-PARAMETRIC METHODS

Non-parametric methods estimate survivor functions without any parametric assumptions. This can be achieved by obtaining discrete distribution from non-parametric methods. Generally, a non-parametric estimation of the survivor function is called the Kaplan–Meier estimator, which can be expressed as:

$$\hat{S}(t) = \prod_{(t)}\left(1 - \frac{d_j}{r_j}\right) = \prod_{(t)}\left(\frac{r_j - d_j}{r_j}\right) \quad \text{Eq (B.7)}$$

From the above function, the cumulative hazard function can also be estimated, which is also called the Nelson estimator:

$$\widehat{H}(t) = \sum_{(t)} h_j = \sum_{(t)} \frac{d_j}{r_j} \qquad \text{Eq (B.8)}$$

where d_j is the number of failures among r_j individuals at a point a_j from atoms of a purely discrete distribution.

APPENDIX C

SURVIVAL TIMES MODELLING

ACCELERATED FAILURE TIME (AFT) MODEL

As per the literature, this article follows the parametric AFT models. Let us assume T is the failure time and $V \equiv \{V_1, V_2, V_3, \dots V_n\}$ represents a set of explanatory variables. Using the AFT model, the survival time of each order is

$$T = \frac{T_0}{\psi(V)} \qquad \text{Eq (C.1)}$$

where T_0 is the base survival time when $\psi(V) = 1$, and T is scaled by the explanatory variables. Using these assumptions, the survivor function can be expressed as

$$S(t|V) = S_0\big(t\psi(V)\big) \qquad \text{Eq (C.2)}$$

and the density function can be expressed as:

$$f(t|V) = f_0\big(t\psi(V)\big)\psi(V) \qquad \text{Eq (C.3)}$$

where $S_0(.)$, $f_0(.)$, and $h_0(.)$ are valid for T_0. The natural choice for $\psi(V)$ (provided $\psi(V) > 0$) can be

$$\psi(V|\beta) = \exp(\beta^T V) \qquad \text{Eq (C.4)}$$

where β is a vector of parameters. So T can be written as

$$T = T_0 \exp(-\beta^T V) \qquad \text{Eq (C.5)}$$

$$\log T = \mu_0 - \beta^T V + \epsilon \qquad \text{Eq (C.6)}$$

where $\mu_0 = E(\log T_0)$ and ϵ is a random variable, the distribution of which does not depend on V.

The log-linear form of the AFT model can show the mathematical relation between the log of failure time and covariates (Hosmer et al., 2011; Khanal et al., 2014; Collett, 2015). Assume, if a set of covariates are $x_1, x_2, x_3, \dots, x_n$, then,

$$\log T = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_n x_{ni} + \sigma \varepsilon_i \qquad \text{Eq (C.7)}$$

where β_0 is the intercept, $\beta_1, \beta_2, \dots \beta_n$ are the coefficient values of n, the number of explanatory variables for the i-th order, $\sigma$is the scale value, and ε_i is a random variable depicting deviation of values of $\log T$ from the linear part of the model. Now, with a specific parametric form of baseline distribution, the maximum likelihood function of this parameter vector can be estimated.

COX PROPORTIONAL HAZARD (CPH) MODEL

Proportional hazard models are another way of modeling the relationship between the survival time and explanatory variables. Each observation's hazard can be written as

$$h(t|V) = \psi(V)h_0(t) \quad \text{Eq (C.8)}$$

where $h_0(t)$ is the baseline hazard function. In this case, the explanatory variables behave multiplicatively on the hazard rate. For two different points V_1 and V_2,

$$\frac{h(t|V_1)}{h(t|V_2)} = \frac{\psi(V_1)}{\psi(V_2)} \quad \text{Eq (C.9)}$$

This ratio is called the hazard ratio, which is constant at time t. The survivor function is expressed as

$$S(t|V) = S_0(t)^{\psi(V)} \quad \text{Eq (C.10)}$$

and the density function as

$$f(t|V) = \psi(V)f_0(t)S_0(t)^{\psi(V)-1} \quad \text{Eq (C.11)}$$

and

$$\psi(V|\beta) = \exp(\beta^T V) \quad \text{Eq (C.12).}$$

APPENDIX D

METHODOLOGY FOR LIMIT ORDER BOOK EXECUTION ON BUYER AND SELLER SIDES

This appendix presents the step-by-step formation of the LOB by structuring a trading day into order-book periods and time intervals. The matching of buyer and seller orders is carried out based on the different order-execution criteria. After matching, non-executed orders are carry forwarded to the next interval.

Step	Buyer Side (B)	Seller Side (A)
Time Structure	$PL_T^B = \sum_{l=1}^{6}\sum_{t=1}^{4} E_{tl}^B$ $PL_T^A = \sum_{l=1}^{6}\sum_{t=1}^{4} E_{tl}^A$ *Where* *Trading Day* $T = \{1,2,\dots,k\}$; *Limit Order Book (LOB) in a single trading day* $l = \{1,2,3,4,5,6\}$; *Interval within a LOB* $t = \{1,2,3,4\}$	$PL_T^A = \sum_{l=1}^{6}\sum_{t=1}^{4} E_{tl}^A$ $PL_T^B = \sum_{l=1}^{6}\sum_{t=1}^{4} E_{tl}^B$ *Where* *Trading Day* $T = \{1,2,\dots,k\}$; *Limit Order Book (LOB) in a single trading day* $l = \{1,2,3,4,5,6\}$; *Interval within a LOB* $t = \{1,2,3,4\}$

Step	***Buyer Side (B)***	***Seller Side (A)***
Interval Structure	$E_{tl}^{B} = D_{tl}^{B} + N_{(t-1)l}^{B}$ $E_{tl}^{A} = D_{tl}^{A} + N_{(t-1)l}^{A}$ $PL_{T}^{B} = \sum_{l=1}^{6}\sum_{t=1}^{4} D_{tl}^{B} + N_{(t-1)l}^{B}$ $PL_{T}^{A} = \sum_{l=1}^{6}\sum_{t=1}^{4} D_{tl}^{A} + N_{(t-1)l}^{A}$ *Where, D_{tl}^{B} is Database Entity for buyer side records; $N_{(t-1)l}^{B}$ is non-executed records at buyer side which will be carry forwarded to next interval within same LOB; D_{tl}^{A} is Database Entity for sellerrecords; $N_{(t-1)l}^{A}$ is non-executed records at seller side which will be carry forwarded to next interval within same LOB*	$E_{tl}^{A} = D_{tl}^{A} + N_{(t-1)l}^{A}$ $E_{tl}^{B} = D_{tl}^{B} + N_{(t-1)l}^{B}$ $PL_{T}^{A} = \sum_{l=1}^{6}\sum_{t=1}^{4} D_{tl}^{A} + N_{(t-1)l}^{A}$ $PL_{T}^{B} = \sum_{l=1}^{6}\sum_{t=1}^{4} D_{tl}^{B} + N_{(t-1)l}^{B}$ *Where, D_{tl}^{B} is Database Entity for buyer side records; $N_{(t-1)l}^{B}$ is non-executed records at buyer side which will be carry forwarded to next interval within same LOB; D_{tl}^{A} is Database Entity for seller side records; $N_{(t-1)l}^{A}$ is non-executed records at seller side which will be carry forwarded to next interval within same LOB*
Step	***Buyer Side (B)***	***Seller Side (A)***
Ordering Structure	$D^{B} = \{P^{B}, V^{B}, TS^{B}, Flag^{B}\}$ $D^{A} = \{P^{A}, V^{A}, TS^{A}, Flag^{A}\}$ *Where* P^{B} *is Bid Price tick wise* V^{B} *is Bid Volume tick wise* TS^{B} *Timestamp tick wise at buyer side* $Flag^{B}$ *Flag of each tick at buyer side* P^{A} *is Seller Price tick wise* V^{A} *is Seller Volume tick wise* TS^{A} *Timestamp tick wise at seller side* $Flag^{A}$ *Flag of each tick at seller side* *Ordering:* $min(TS_{ji1}^{B}, TS_{ji2}^{B}, \ldots, TS_{jim}^{B}) \in \{TS_{ji1}^{B}, TS_{ji2}^{B}, \ldots, TS_{jim}^{B}\}$ $min(P_{ji1}^{A}, P_{ji2}^{A}, \ldots, P_{jin}^{A}) \in \{P_{ji1}^{A}, P_{ji2}^{A}, \ldots, P_{jin}^{A}\}$	$D^{A} = \{P^{A}, V^{A}, TS^{A}, Flag^{A}\}$ $D^{B} = \{P^{B}, V^{B}, TS^{B}, Flag^{B}\}$ *Where* P^{A} *is Seller Price tick wise* V^{A} *is Seller Volume tick wise* TS^{A} *Timestamp tick wise at seller side* $Flag^{A}$ *Flag of each tick at seller side* P^{B} *is Bid Price tick wise* V^{B} *is Bid Volume tick wise* TS^{B} *Timestamp tick wise at buyer side* $Flag^{B}$ *Flag of each tick at buyer side* *Ordering:* $min(TS_{ji1}^{A}, TS_{ji2}^{A}, \ldots, TS_{jin}^{A}) \in \{TS_{ji1}^{A}, TS_{ji2}^{A}, \ldots, TS_{jin}^{A}\}$ $max(P_{ji1}^{B}, P_{ji2}^{B}, \ldots, P_{jim}^{B}) \in \{P_{ji1}^{B}, P_{ji2}^{B}, \ldots, P_{jim}^{B}\}$
Step	***Buyer Side (B)***	***Seller Side (A)***
Searching	P_{ji1}^{B} *is to match with* $\{P_{ji1}^{A}, P_{ji2}^{A}, \ldots, P_{jin}^{A}\}$ *1 tick of bid will match with bucket of ask ticks* $\{1,2, \ldots, n\}$ *in each interval* *Matching Algorithm:* *for each bid in {1,2,….,m} do* *for each seller in {1,2,…,n} do* *lookup the value bid.1 in the index of seller;* *if matching occurs* *then break;* *else* *continue;* *end for;* *end for;*	P_{ji1}^{A} *is to match with* $\{P_{ji1}^{B}, P_{ji2}^{B}, \ldots, P_{jim}^{B}\}$ *1 tick of ask will match with bucket of bid ticks* $\{1,2, \ldots, m\}$ *in each interval* *Matching Algorithm:* *for each seller in {1,2,….,n} do* *for each bid in {1,2,…,m} do* *lookup the value seller.1 in the index of bid;* *if matching occurs* *then break;* *else* *continue;* *end for;* *end for;*

Step	***Buyer Side (B)***	***Seller Side (A)***
Matching	$Flag_{ijw}^{B} = \begin{cases} FTK, when\ P_{ijw}^{B} \geq P_{ijv}^{A}\ and\ V_{ijw}^{B} \leq V_{ijv}^{A} \\ PARTF, when\ P_{ijw}^{B} \geq P_{ijv}^{A}\ and\ V_{ijw}^{B} > V_{ijv}^{A} \\ NON, when\ P_{ijw}^{B} < P_{ijv}^{A} \end{cases}$ $Flag_{ijv}^{A} = \begin{cases} FTK, when\ P_{ijw}^{B} \geq P_{ijv}^{A}\ and\ V_{ijw}^{B} \geq V_{ijv}^{A} \\ PARTF, when\ P_{ijw}^{B} \geq P_{ijv}^{A}\ and\ V_{ijw}^{B} < V_{ijv}^{A} \\ NON, when\ P_{ijw}^{B} < P_{ijv}^{A} \end{cases}$ *Where* FTK *represents Fill-to-Kill Orders,* $PARTF$ *represents Partial Filled Orders and* NON *represents Non Executed Orders*	$Flag_{ijv}^{A} = \begin{cases} FTK, when\ P_{ijw}^{B} \geq P_{ijv}^{A}\ and\ V_{ijw}^{B} \geq V_{ijv}^{A} \\ PARTF, when\ P_{ijw}^{B} \geq P_{ijv}^{A}\ and\ V_{ijw}^{B} < V_{ijv}^{A} \\ NON, when\ P_{ijw}^{B} < P_{ijv}^{A} \end{cases}$ $Flag_{ijw}^{B} = \begin{cases} FTK, when\ P_{ijw}^{B} \geq P_{ijv}^{A}\ and\ V_{ijw}^{B} \leq V_{ijv}^{A} \\ PARTF, when\ P_{ijw}^{B} \geq P_{ijv}^{A}\ and\ V_{ijw}^{B} > V_{ijv}^{A} \\ NON, when\ P_{ijw}^{B} < P_{ijv}^{A} \end{cases}$ *Where* FTK *represents Fill-to-Kill Orders,* $PARTF$ *represents Partial Filled Orders and* NON *represents Non Executed Orders*
Step	***Buyer Side (B)***	***Seller Side (A)***
Carry Forward	*Non-executed bid orders are carry forwarded by being added into* $N_{tl}^{B} \leftarrow Bid\ Order\ Ticks\ with\ Flag_{ijw}^{B} = NON$ *Non-executed seller orders are carry forwarded by being added into* $N_{tl}^{A} \leftarrow Ask\ Order\ Ticks\ with\ Flag_{ijv}^{A} = NON$ *All the non-executed orders at either buyer or seller side will be carry forwarded to next interval within same LOB;* *The process continues;*	*Non-executed seller orders are carry forwarded by being added into* $N_{tl}^{A} \leftarrow Ask\ Order\ Ticks\ with\ Flag_{ijv}^{A} = NON$ *Non-executed bid orders are carry forwarded by being added into* $N_{tl}^{B} \leftarrow Bid\ Order\ Ticks\ with\ Flag_{ijw}^{B} = NON$ *All the non-executed orders at either buyer or seller side will be carry forwarded to next interval within same LOB;* *The process continues;*

APPENDIX E

SUMMARY STATISTICS: CARRY FORWARD LOB OF BUYER AND SELLER SIDES (JUNE 2016)

This appendix presents the summary statistics for the explanatory variables impacting the probability of limit order execution on both the buyer and seller sides.

Explanatory Variable	Summary Statistics	Buyer	Seller
Limit Order Size	*Mean*	*25409.7483*	*23898.8833*
	Median	*67*	*65*
	Standard Deviation	*121410.1289*	*109249.7174*
	Quantile	*0 (0%), 9 (25%), 67 (50%), 835 (75%), 4031956 (100%)*	*0 (0%), 9 (25%), 65 (50%), 800 (75%), 2308732 (100%)*
Remaining Trading Time (in Intervals)	*Mean*	*1.0935*	*1.0985*
	Median	*1*	*1*
	Standard Deviation	*1.0456*	*1.0461*
	Quantile	*0 (0%), 0 (25%), 1 (50%), 2 (75%), 3 (100%)*	*0 (0%), 0 (25%), 1 (50%), 2 (75%), 3 (100%)*
Bid-Ask Spread (INR)	*Mean*	*1.3943*	*1.3305*
	Median	*0.15*	*0.15*
	Standard Deviation	*5.0993*	*4.4595*
	Quantile	*0.00 (0%), 0.00 (25%), 0.15 (50%), 0.85 (75%), 188.05 (100%).*	*0.00 (0%), 0.00 (25%), 0.15 (50%), 0.85 (75%), 195.00 (100%).*
Same-Side Book Depth	*Mean*	*35.3658*	*35.0651*
	Median	*34*	*34*
	Standard Deviation	*15.6208*	*15.3974*
	Quantile	*1 (0%), 22 (25%), 34 (50%), 46 (75%), 80 (100%)*	*1 (0%), 22 (25%), 34 (50%), 46 (75%), 80 (100%)*
Opposite-Side Open or Available Orders in the Book	*Mean*	*31.6067*	*32.1227*
	Median	*29*	*30*
	Standard Deviation	*14.7763*	*14.9700*
	Quantile	*0 (0%), 20 (25%), 29 (50%), 40 (75%), 80 (100%)*	*1 (0%), 20 (25%), 30 (50%), 41 (75%), 80 (100%)*