

A Model for Queue Position Valuation in a Limit Order Book*

Ciamac C. Moallemi
Graduate School of Business
Columbia University
email: ciamac@gsb.columbia.edu

Kai Yuan
Graduate School of Business
Columbia University
email: kyuan17@gsb.columbia.edu

Initial Version: December 2016

Current Revision: June 2017

Abstract

Many financial markets operate as electronic limit order books under a price-time priority rule. In this setting, among all resting orders awaiting trade at a given price, earlier orders are prioritized for matching with contra-side liquidity takers. This creates a technological arms race among high-frequency traders and other automated market participants to establish early (and hence advantageous) positions in the resulting first-in-first-out (FIFO) queue. We develop a model for valuing orders based on their relative queue position that incorporates both economic (informational) and stochastic modeling (queueing) aspects. Our model identifies two important components of positional value: (i) a static component that relates to the trade-off at an instant of trade execution between earning a spread and incurring adverse selection costs, and incorporates the fact that adverse selection costs are increasing with queue position; (ii) a dynamic component, that captures the optionality associated with the future value that accrues by locking in a given queue position. Our model offers predictions of order value at different positions in the queue as a function of market primitives, and can be empirically calibrated. We validate our model by comparing it with estimates of queue value realized in backtesting simulations and find the predictions to be accurate. Moreover, for some large tick-size stocks, we find that queue value can be of the same order of magnitude as the bid-ask spread. This suggests that accurate valuation of queue position is a necessary and important ingredient in considering optimal execution or market-making strategies for such assets.

1. Introduction

Modern financial markets are predominantly electronic. In modern exchanges, market participants interact with each other through computer algorithms and electronic orders. The image of traders frantically gesturing and yelling to each other on the trading floor has largely given way to impersonal computer terminals. In terms of market structure, the electronic limit order book (LOB) has become dominant for certain asset classes such as equities and futures in the United States. Figure 1 illustrates how a limit order book works. It is presented as a collection of resting limit orders, each of which specifies a quantity to be traded and the worst acceptable price. The limit

*The authors wish to thank Robert Almgren, Matt Cushman, Costis Maglaras, and Giuseppe Nuti for helpful discussions. Moallemi was supported by NSF grant CMMI-1235023.

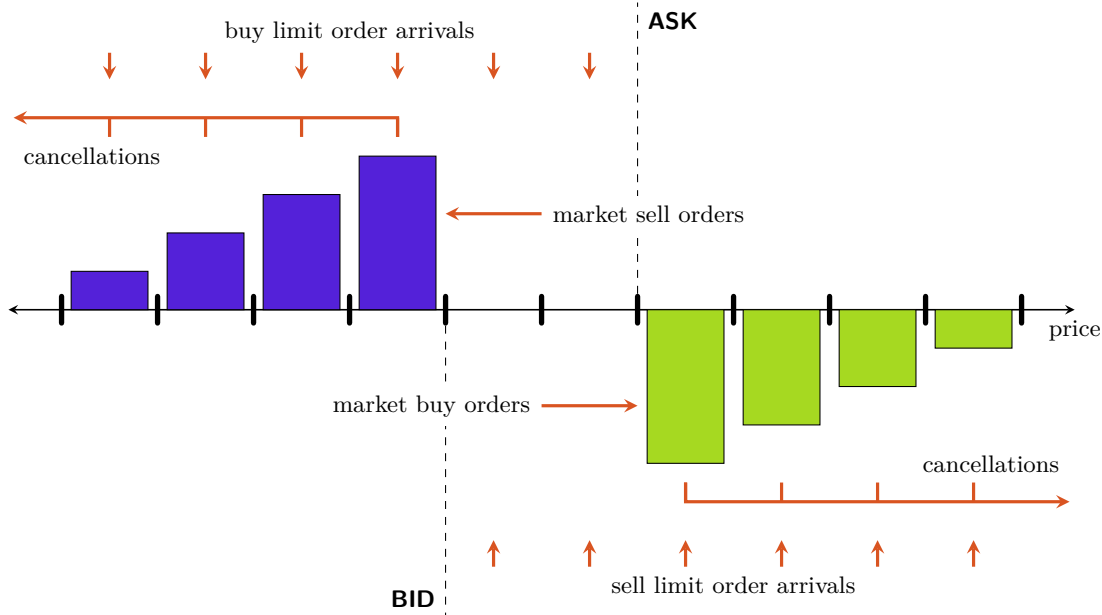


Figure 1: An illustration of a limit order book.

orders will be matched for execution with market orders¹ that demand immediate liquidity. Traders can therefore either provide liquidity to the market by placing these limit orders or take liquidity from it by submitting market orders to buy or sell a specified quantity.

Most limit order books are operated under the rule of *price-time priority*, that is used to determine how limit orders are prioritized for execution. First of all, limit orders are sorted by the price and higher priority is given to the orders at the best prices, i.e., the order to buy at the highest price or the order to sell at the lowest price. Orders at the same price are ranked depending on when they entered the queue according to a *first-in-first-out* (FIFO) rule. Therefore, as soon as a new market order enters the trading system, it searches the order book and automatically executes against limit orders with the highest priority. More than one transaction can be generated as the market order may run through multiple subsequent limit orders.² In fact, the FIFO discipline suggests that the dynamics of a limit order book resembles a queueing system in the sense that limit orders wait in the queue to be filled by market orders (or canceled). Prices are typically discrete in limit order books and there is a minimum increment of price which is referred to as tick size. If the tick size is small relative to the asset price, traders can obtain priority by slightly improving the order price. But it becomes difficult when the tick size is economically significant. As a result, queueing position becomes important as traders prefer to stay in the queue and wait for their turn of execution.

High-level decision problems such as market making and optimal execution are of great interest in both academia and industry. A central question in such problems is understanding when to use

¹We do not make a distinction between market orders and marketable limit orders.

²There is an alternative rule called *pro-rata*, which works by allocating trades proportionally across orders at the same price. In a pro-rata setting, queue position is not relevant to order value, and hence we will not consider pro-rata markets in this paper.

limit orders as opposed to market orders and how to place limit orders if they are preferred. The key ingredient necessary to answer this question is the estimation of the value of a limit order. In this paper, we try to relate the value of a limit order to its queue position. We claim that, from a valuation perspective, queue positions are relevant and that positions at the front of the queue are very valuable for the following reasons:

1. Good queue positions guarantee early execution and less waiting time. This is particularly important for algorithmic traders who potentially have a large number of trades scheduled to be executed in a limited time horizon. Additionally, less waiting time can translate to a higher fill rate, because there is less chance that the market price will move away while the limit orders are resting in the queue.
2. Good queue positions also mean lower adverse selection costs. Orders at the end of a large queue will be executed in the next instance only against large trades. On the other hand, orders at the very front of the queue will be executed against the next trade no matter what its size will be. Large trades often originate from informed traders who are confident about the trades' profitability. In this way, a good queue position acts as a filter on the population of contra-side market orders so that the liquidity provider is less likely to be disadvantaged by trading against informed traders. This relationship between queue positions and adverse selection is first observed by Glosten (1994), who considers a single-period setting.

In practice, certain classes of market participants expend significant effort trying to take obtain better queue positions in the limit order book. For example, there has been controversy in recent years over exotic order types on certain exchanges that allow traders to attain priority in the limit order book. These exotic order types “allow high-speed trading firms to trade ahead of less-sophisticated investors, potentially disadvantaging them and violating regulatory rules.”³ Another example is that there has been the technological arms race between high-frequency traders who invest in technologies for low-latency trading. Note that a central driver for low-latency trading is attaining good queue positions — all things being equal, orders that are emitted faster will wind up in a better queue position, and this is especially important when traders are reacting to common signals. Indeed, one situation where it is important to trade quickly is the instant right after a price change. For example, when a trade wipes out the current ask and the price is about to tick up, there will be a race to establish queue positions at the new price, and small differences in speed among competing high-frequency traders may translate to large differences in queue position.

In the literature, some earlier work, such as that of Glosten (1994), has implications about the value of queue positions. Although these models point out the importance of adverse selection, they are fundamentally *static* models in which the value of the order is assumed to be determined by whether it will be executed by the next trade or not. In the presence of a large queue, the life cycle of the order will not end with the next trade and traders will not cancel and resubmit their

³Patterson, S. and Strasburg, J., “For Superfast Stock Traders, a Way to Jump Ahead in Line.” *The Wall Street Journal*, September 19, 2012.

limit orders after every single trade. What is more likely to happen is that the order will move up in the queue if not executed by the next trade. This implies that there is value in moving up in the queue, and that this value may accrue over a number of trades and cancellations. As a result, aside from adverse selection, there should be an additional, fundamentally *dynamic* component that can capture the optionality associated with future value that accrues by locking in a given queue position. In order to account for this dynamic component, a multi-period model is needed.

1.1. Contributions

In this paper, we provide a dynamic model for valuing limit orders in large-tick stocks based on their relative queue positions. We appear to be one of the first to study the limit order book queue position value through the lens of dynamic multi-period model. Our model identifies two important components of positional value. First, there is a static component that relates to the adverse-selection costs originating from the possibility of information-motivated trades. We capture the fact that adverse selection costs are increasing with queue position. Second, there is also a dynamic component that captures the value of positional improvement that accrues after order book events such as trades and cancellations.

By making reasonable simplifications, we provide a tractable way to predict order value at different positions in the queue as a function of market primitives. We then empirically calibrate our model in a subset of U.S. equities and find that queue values can be very significant in large-tick assets. Additionally, we validate our model by checking the model-free estimates of queue values using a backtesting technique.

There are many higher-level decision problems that have an ingredient of valuing limit orders. One such example is that market makers need to constantly value limit orders in order to come up with the optimal order-placing strategy. Another example is that in the optimal execution of a large block, algorithmic traders often have to decide between market orders and limit orders. In both cases, we need to value the limit orders and use them as building blocks for the higher-level control problem. What we observe empirically in our model is that queue positions do matter and that positional value is roughly of the same magnitude for large-tick assets. As a result, queue positional value should be an important ingredient downstream of solving optimal control problems with large-tick assets.

1.2. Literature Review

Our paper builds on the classical financial economics literature on market microstructure that studies the informational motives of trading. Kyle (1985) and Glosten and Milgrom (1985) were among the first to recognize the importance of adverse selection in analyzing the price impact of trades and the spread, by assuming competitive suppliers of liquidity. Both of their models highlight the fact that the possibility of trading against an informed trader creates incentives for liquidity providers to charge additional premiums. However, these models do not consider queueing effects. Glosten (1994) further extended this type of model, with implications for valuing orders in

the limit order book. One implication of that paper is that it states that in cases where the prices are discrete, the queue length should be determined by the fact that the value of the last order in the queue is zero. Basically, the investor putting in the marginal order should be indifferent between joining the queue or not. In this way, while the paper does not explicitly model the value of queue positions,⁴ it does manage to relate queue length to order values. Moreover, the model in Glosten (1994) is a single-period static model in which the order values are calculated toward the next trade. However, what's more likely is that an order will move up in the queue if it is not executed. Our model incorporates the dynamic values embedded in the queue position improvement. Additionally, by considering a dynamic model, we are also able to consider order book events such as cancellations. As a result, queue position actually matters in our model, and is clearly correlated with the order values. For example, if the queue position is decreasing, then either there is a trade or people are canceling, and either event conveys information about asset value.

Recently, there has been a growing literature from the stochastic modeling and financial engineering communities on the development of queueing models that solve various kinds of problems regarding limit order books while recognizing that the price-time priority structure in the limit order books can be modeled as a multi-class queueing system. Cont et al. (2010) was the first to model the limit order book as a continuous-time Markov model that tracks the limit orders at each price level. By assuming that order flows can be described as Poisson processes, the authors provided a parametric way to calculate the conditional probability of various order book events such as the probability of executing an order before a change in price. Cont and De Larrard (2013) further modeled the order book events in a Markovian queueing system, and studied the endogenous price dynamics resulting from executions. Lakner et al. (2013) studied a similar setup but focused on the high-frequency regime where the arrival rate of both limit orders and market orders is large. Blanchet and Chen (2013) derived a continuous-time model for the joint evolution of the mid price and the bid-ask spread. Several papers such as Guo et al. (2013), Cont and Kukanov (2013), and Maglaras et al. (2015) have been working on optimizing trading decisions in the context of a queueing model for the limit order book. More specifically, Guo et al. (2013) proposed a model to optimally place orders, given price impact. Cont and Kukanov (2013) derived the optimal split between limit and market orders across multiple exchanges. Maglaras et al. (2015) studied optimal decision making in the placement of limit orders as well as in trying to execute a large trade over a fixed time horizon. Avellaneda et al. (2011) tried to forecast the price change based on order book imbalance, while in our settings price changes are exogenous. However, the limitation of the queueing literature is that it lacks the informational component of adverse selection. And yet an important ingredient in modeling the positional value of limit orders is the concept of adverse selection, i.e., of a correlation between trades and prices. Our model tries to bridge this gap by considering the economics of adverse selection in a queueing framework.

⁴In fact, Glosten (1994) assumes that competing limit orders in the same queue are executed in a *pro-rata* fashion, where queue position is not directly relevant.

From the empirical front, there is a significant body of literature conducting empirical analyses of the dynamics of limit order books in major exchanges. Bouchaud et al. (2006) showed that the random-walk nature of traded prices is nontrivial. Biais et al. (1995) and Griffiths et al. (2000) studied the limit-order submission under different market conditions. Hollifield et al. (2004) further stated that optimal order submission depends not only on the valuation of the assets but also on the trade-offs between order prices, execution probabilities, and picking-off risks.

There are several successful examples of modeling the optionality embedded in limit orders. Copeland and Galai (1983) argued that informed traders are willing to pay a “fee” to obtain immediacy in trading with liquidity providers. Chacko et al. (2008) further modeled limit orders as American options that require delivery of the underlying shares upon execution. However, these models are fundamentally static in that they do not explicitly model the queue positions.

The rest of this chapter is organized as follows: Section 2 provides an overview of our approach and describe the dynamic of the order book. In Section 3, we provide closed-form solution for the value function. In Section 4, we consider empirical calibration of the model to trading data from NASDAQ. Section 5 describes a procedure of backtesting and analyzes the quality of the predictions of our model with backtesting results. Section 6 concludes and discusses practical implications of our analysis. Proofs are deferred until the technical appendix.

2. Model

In many modern exchange-traded financial markets, while the price per share differs substantially across assets, the tick size is artificially fixed. For example, all stocks traded at NYSE have a minimum increment of \$0.01. Large-tick assets are those which, according to Eisler et al. (2012), are such that “the bid-ask spread rarely exceeds the minimum tick size”. These are the assets where the tick size is economically significant, and therefore they are typically traded with the bid-spread equal to the tick size. Another important characteristic of large-tick assets is that they tend to have large queues in the limit order book. The reason is that the cost of moving the price by one tick will be very economically significant. For example, adjusting the price by a single tick on a low priced asset in order to obtain order book priority can translate into a large negative return. Hence, instead of competing through price, market participants tend to form queues. Figure 2 shows the relationship between bid-ask spread and displayed liquidity for various future contracts, and we can see a clear pattern that queueing effect is more prominent for large-tick assets. In this paper, we will restrict our attention to the large-tick assets where queueing is important. In contrast, for small-tick assets, where the typical big-ask spread is much larger than the minimum price increment, investors can obtain priority in the order book by competing on price and the important of queueing (and hence the importance of valuing queue position) is diminished.

For simplicity, we will assume that over the time scale of our model, the bid and ask prices do not change as the tick size is large. Additionally, we will assume that the bid-ask spread is constant and equal to the tick size (which is almost always true for large-tick assets). Without loss

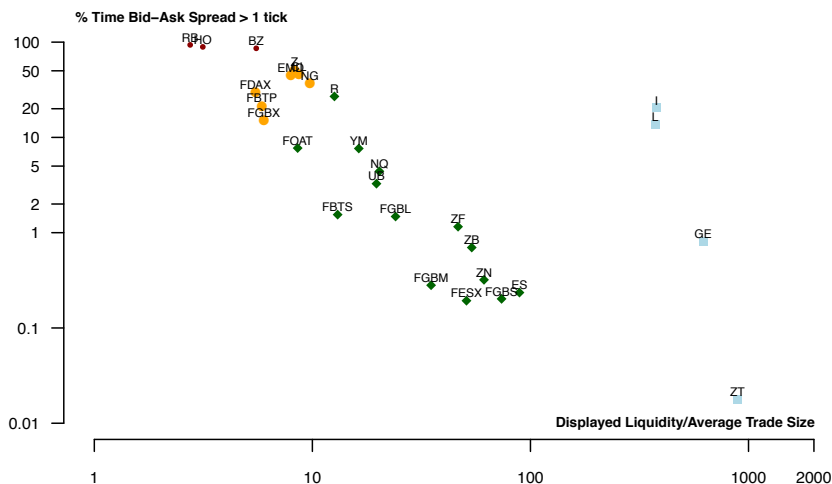


Figure 2: A comparison of liquidity and bid-ask spread across various Futures Contracts, July–August 2013 (courtesy Robert Almgren).

of generality, we will normalize prices so that the tick size (and hence, the bid-ask spread) is 1. We will focus only on the ask side of the market, where limit orders are posted to sell the asset and wait to be executed against market orders from buyers. The case for the bid side can be derived similarly. We will also consider a single-exchange setup to avoid the complications of merging limit order books from different exchanges.

As we are interested in situations where the queue length is large, we ignore the integrality issues related to the fact that assets must be traded in discrete quantities. Instead, we assume that the queue position is continuous, and we are interested in modeling the positional value in placing an incremental order of infinitesimal size. We are concerned with short intraday time horizons over which an order might get executed. Over this short time period, we assume that the risk-free rate is zero since there is typically no interest associated with intraday borrowing or lending. Further, we assume that the agent is risk neutral. Risk neutrality is appropriate for several reasons. First of all, we are looking at a single infinitesimal order here, which is relatively small compared to the agent’s wealth. Therefore we can assume that the agent’s utility function is linear for this particular order. Second, we expect the agent to submit many such orders over non-overlapping time intervals to accumulate a large position. Then the law of large numbers will apply, making the agent effectively risk neutral.

2.1. Order Valuation

Our goal is to estimate the value of a limit order, especially as it relates to the queue position of the order, in a dynamic multi-period setting. To this end, we consider a stylized problem where an agent arrives seeking to provide liquidity by selling⁵ an infinitesimal quantity of an asset via a limit order. The order is placed at time $t = 0$, at the best ask price P^A , and remains in the order book

⁵This is without loss of generality, since the buying case is symmetric.

until either it transacts (i.e., is filled) or until its price changes and (by assumption) the order is canceled.

To understand the value of the order, it is necessary to develop a model for the value of the underlying asset. To this end, we assume that the asset will be liquidated at a random future time T , and at that time will realize a (random) cash flow P . The cash flow P can be viewed as the *fundamental value* of the asset. T should be viewed as the time when all information regarding the price of the underlying asset has been made public. Denote by $\{\mathcal{F}_t\}$ the filtration that represents the information possessed by the agent, at each time $t \geq 0$, and define *latent efficient price process* P_t , for $t \geq 0$, according to

$$P_t \triangleq \mathbf{E}[P|\mathcal{F}_t].$$

We will further assume that the filtration is right-continuous in the sense that $\mathcal{F}_t = \mathcal{F}_{t+}$ for all $t \geq 0$. By construction, P_t is a right-continuous Doob martingale.

Now, define $\tau^* \in [0, T)$ to be the \mathcal{F}_t -measurable stopping time when the order is either filled or canceled. If the order is filled, the agent is paid P_A in exchange for a short position with (eventual) fundamental value P . If the order is canceled, the agent receives nothing. Therefore (assuming risk-neutrality and a zero risk-free rate), the value of this order to the agent is given by

$$V_t \triangleq \mathbf{E} \left[(P_A - P) \mathbb{I}_{\{\text{FILL}\}} \mid \mathcal{F}_t \right],$$

for all $t \geq 0$. For $t \in [0, \tau^*)$, since P_t is a right-continuous martingale, we can apply the optional stopping theorem (e.g., Theorem 3.22, Karatzas and Shreve, 2012) to see that

$$\begin{aligned} V_t &= \mathbf{E} \left[(P_A - P_t) \mathbb{I}_{\{\text{FILL}\}} - (P - P_t) \mathbb{I}_{\{\text{FILL}\}} \mid \mathcal{F}_t \right] \\ &= \alpha_t (\delta_t - \text{AS}_t), \end{aligned} \tag{1}$$

where

$$\begin{aligned} \alpha_t &\triangleq \mathbf{P}(\text{FILL} \mid \mathcal{F}_t), \\ \delta_t &\triangleq P_A - P_t, \\ \text{AS}_t &\triangleq \mathbf{E}[(P_{\tau^*} - P_t) \mid \mathcal{F}_t, \text{FILL}]. \end{aligned}$$

These stochastic processes have natural interpretations at each time $t \in [0, \tau^*)$:

- α_t is the *fill probability* of the order.
- δ_t captures the difference between the order's posted price P_A and the latent efficient price P_t ; we call this the *liquidity premium* or *liquidity spread* earned by the order.⁶
- AS_t measures the revision of the agent's estimate of the asset's fundamental value from the

⁶For example, if P_t happened to coincide with the mid-market price, δ_t would equal a half-spread. In this way, the quantity δ_t generalizes the intuition that a limit order "earns a half-spread" to situations where the fundamental value of the asset differs from the mid-market price.

present time (P_t) to the time of a fill (P_{τ^*}), conditional on a fill. Note that $AS_t = 0$ if fills are independent of the efficient price process. However, in realistic settings with asymmetrically informed traders, one typically expects that $AS_t > 0$. This is because of the possibility that the contra-side trader, who is demanding liquidity and paying associated spread costs by buying at the ask price, is motivated by private information about the fundamental value of the asset. Hence, trades and innovations of the efficient price process are dependent in a way that is to the detriment of the liquidity provider and, accordingly AS_t is known as *adverse selection*. Adverse selection is an important issue in evaluating the value of limit orders, and has been noted in many studies, such as those by Glosten and Milgrom (1985) and Kyle (1985).

The decomposition in (1) can be interpreted informally as an accounting identity that breaks down the expected profitability of liquidity provision at the level of an individual order as follows:

$$\text{order value} = \text{fill probability} \times (\text{liquidity spread premium} - \text{adverse selection cost}).$$

Hollifield et al. (2004) used a similar decomposition to (1) to describe the agent's expected pay-off in placing the order. Their approach is slightly general as they included an error term to represent the trader's private value for the assets. In our model, we are looking from the perspective of competitive market makers who have no private information. As a result, the private values are assumed to be zero. Hollifield et al. (2004) do not explicitly consider queue positions, and the fill probabilities are estimated in a non-parametric way for different price levels. In fact, their approach finding the trader's optimal submission strategy across different price levels is fundamentally static, whereas our approach estimating values of orders at different queue positions uses a dynamic model.

2.2. Price Dynamics

We assume that innovations in the latent efficient price process are driven by two types of discrete exogenous events, *trades* and *price jumps*. Trades correspond to the arrival of an impatient buyer (resp., seller), who demands immediate liquidity and is matched with a seller (resp., buyer) at the best ask (resp., bid) price. For the i th trade, denote its arrival time by $\tau_i^u > 0$ and its signed⁷ trade size by $u_i \in \mathbb{R}$. Price jumps, on the other hand, represent an instant in time at which price levels across the board shift up (resp., down) due to the arrival of new information. In an upward (resp., downward) jump, we assume that all orders at the best ask (resp., best bid) price are filled. We denote the arrival time of the k th jump by τ_k^J and its size by J_k .

We posit the following dynamics for the latent efficient price,

$$P_t = P_0 + \lambda \sum_{i: \tau_i^u \leq t} u_i + \sum_{k: \tau_k^J \leq t} J_k, \quad (2)$$

⁷The case where $u_i < 0$ represents a market order to sell, while $u_i > 0$ represents a market order to buy.

or, equivalently, for liquidity premium,

$$\delta_t = \delta_0 - \lambda \sum_{i: \tau_i^u \leq t} u_i - \sum_{k: \tau_k^J \leq t} J_k, \quad (3)$$

for $t \in [0, \tau^*)$. Accordingly, we make the following assumptions:

- *Linear price impact.* The i th trade impacts the latent efficient price by λu_i ; i.e., there is a permanent linear price impact. The quantity $\lambda > 0$ captures the sensitivity of prices to trade size. This is consistent with the strategic model of Kyle (1985), where such price impact results from asymmetrically informed traders. Although our model is reduced form in that the price impact is specified exogenously, the spirit of it is that large trades are more likely to be due to informed traders, and hence have a greater impact on the posterior beliefs of the trader.
- *Poisson trade arrivals.* We will assume that the trade times $\{\tau_i^u\}$ are Poisson arrivals with rate $\mu > 0$.
- *I.i.d. trade sizes.* We will assume that the trade sizes $\{u_i\}$ are independent and identically distributed with probability density function $f(\cdot)$ over \mathbb{R} . In order to ensure that P_t is a martingale, we will require that $\mathbb{E}[u_i] = 0$. To avoid technicalities, we further assume that $f(\cdot)$ is continuous and $f(u) > 0$ for all $u \in \mathbb{R}$; i.e., the support of the distribution is all of \mathbb{R} .
- *Poisson jump arrivals.* We will assume that the jump times $\{\tau_k^J\}$ are Poisson arrivals with rate $\gamma > 0$.
- *I.i.d. jump sizes.* We will assume that the jump sizes $\{J_k\}$ are independent and identically distributed. In order to ensure that P_t is a martingale, we require that $\mathbb{E}[J_k] = 0$.

We require that arrival times, trade sizes, and jump sizes be \mathcal{F}_t -measurable, so that P_t is an \mathcal{F}_t -adapted process with sample paths that are right continuous with left limits (RCLL) — in fact, P_t is a piecewise constant pure jump process.

Note that the dynamics of P_t are determined by the arrival rate parameters $(\lambda, \mu, \gamma) \in \mathbb{R}_+^3$ and the distributions of trade sizes and jump sizes. An application of the law of total variance yields, for $t \in [0, T)$,

$$\text{Var}(P_t) = (\mu\lambda^2\sigma_u^2 + \gamma\sigma_J^2) t,$$

where $\sigma_u^2 \triangleq \text{Var}(u)$ is the variance of trade sizes and $\sigma_J^2 \triangleq \text{Var}(J)$ is the variance of jump sizes. Expressing this as a per-unit time price volatility of the asset σ_P , we have

$$\sigma_P \triangleq \sqrt{\text{Var}(P_t)/t} = \sqrt{\mu\lambda^2\sigma_u^2 + \gamma\sigma_J^2}.$$

2.3. Limit Order Book Dynamics

The limit order is placed at the best ask price P_A , and remains in the order book either until it is filled, or until the price changes and (by assumption) the order is canceled. Moreover, during the time that is active, the order moves toward the front of its position, as orders with greater queue priority are filled or canceled, according to price-time priority rules.

Specifically, subsequent to its placement, denote the queue position of the limit order by $q_t \in \mathcal{Q} \triangleq \mathbb{R}_+ \cup \{\text{FILL}, \text{CANCEL}\}$. Specifically, at each time $t \in [0, \tau^*)$ at which the order has not been filled or canceled, $q_t \in \mathbb{R}_+$ and the quantity q_t of asset shares, available for sale at the best ask price, is of greater priority than the limit order. If the order has been filled (resp., canceled) prior to time t , then $q_t = \text{FILL}$ (resp., $q_t = \text{CANCEL}$). Until the order is filled or canceled, the queue position q_t evolves according to a sequence of arrivals of one of the following types of events at each event time $\tau > 0$:

1. A trade occurs with size $u_i \in \mathbb{R}$. As per equation (2), the liquidity spread evolves according to

$$\delta_\tau = \delta_{\tau-} - \lambda u_i.$$

For the evolution of the queue position, there are three cases:

- (a) $u_i \in [q_{\tau-}, \infty)$. In this case, the quantity of shares is purchased at the best ask price that exceeds the limit order queue position; hence, the order is filled and realizes a final expected value of

$$V_\tau = \mathbf{E}[P_A - P | \mathcal{F}_\tau] = \delta_\tau = \delta_{\tau-} - \lambda u_i,$$

where, for the last inequality, we apply the price dynamics of equation (2).

- (b) $u_i \in [0, q_{\tau-})$. In this case, the quantity of shares is purchased at the best ask price but it is insufficient to result in a fill; however, the order position improves according to

$$q_\tau = q_{\tau-} - u_i > 0.$$

- (c) $u_i \in (-\infty, 0)$. In this case, the quantity of shares is purchase; hence the queue position q_τ remains fixed.

2. A price jump occurs with size $J_k \in \mathbb{R}$. As per equation (2), the liquidity spread evolves according to

$$\delta_\tau = \delta_{\tau-} - J_k.$$

For the evolution of the queue position, there are two cases:

- (a) $J_k > 0$. Under a positive price jump, the order is assumed to be filled and realizes a final expected value of

$$V_\tau = \mathbf{E}[P_A - P | \mathcal{F}_\tau] = \delta_\tau = \delta_{\tau-} - J_k.$$

(b) $J_k > 0$. Under a negative price jump, the price levels shift down and the order is assumed to be canceled, realizing a final value of $V_T = 0$.

3. The next event is the cancellation of a quantity of higher priority at the best ask price level. We will describe the underlying assumptions of cancellation model shortly, but for now it suffices to note that the i th cancellation event is associated with a proportion $\ell_i \in [0, 1]$, and therefore a fraction $1 - \ell_i$ of the shares with higher priority at the best ask price level are canceled. Hence,

$$q_\tau = \ell_i q_{\tau-}.$$

While the impact of trades is easy to model with the FIFO rule, cancellations can happen at any position in the queue. Moreover, we are interested only in the cancellations that happened in front of the current position. In order to model cancellations, we introduce the two assumptions as follows:

- *Proportional and Uniform Cancellations.* We will assume that after each cancellation on the ask side, the ask queue is homogeneously contracted by a certain proportion ℓ , where $\{\ell_i\}$ are i.i.d. with continuous p.d.f. $g(\cdot)$ over $[0, 1]$. Further, cancellations occur on the ask side at times associated with a Poisson process of rate η^+ . Additionally, we assume that the cancellation happens uniformly across different queue positions. Under this assumption, the queue position of a limit order will be updated from q to $\ell_i q$ after the i th cancellation.
- *Uninformed Cancellations.* We assume that cancellations happen randomly and possess no extra information. Some empirical work, such as that of Cont et al. (2014), has argued that there is a correlation between price moves and cancellations; however, the market impact of cancellations should be much smaller than that of market orders and hence we will neglect this effect due to its technicality.

All things being equal, we expect cancellations to be larger when the queue is larger. Therefore, instead of modeling both the size and the position of cancellations, we assume proportional cancellations with a specific distribution fitted from the data. The order dynamics with cancellations is then presented as follows:

1. If the cancellation happens on the ask side with cancellation fraction ℓ , then the queue position of the order (currently q) is assumed to shrink to ℓq .
2. If the cancellation happens on the bid side, then the referenced order is not affected at all.

3. Analysis

Now we consider the value of the queue position from the perspective of the agent. In Section 2.1, we defined the value of a limit order. In this section, under the dynamics described in Section 2.2

and Section 2.3, we will characterize this value. In what follows, we assume that the agent places his order at time 0.

Naturally, the value of a limit order is determined by the price at which the order is placed (P_A), the latent efficient price (P_t) at the time it is executed (resp., canceled), and the probability of execution. Because our price dynamics do not depend on price levels, we can consider prices relative to the ask price of time zero (P_A), which is denoted by δ_t . In addition, the probability of execution is a function of queue position according to the order dynamics in our model. Hence the value of a limit order can be uniquely determined by the state variable (δ, q) . Given that all the events in our model (trades, price jumps, and cancellations) are assumed to have Poisson arrival times, the evolution of state variable (δ, q) over time can be viewed as a continuous-time Markov chain. By setting the uniformization parameter as $\zeta = \mu + \gamma + \eta^+$, we can transfer the continuous-time Markov chain to a discrete-time Markov chain (see, e.g., Chapter 5.8, Ross, 1996). Following our discussion in Section 2.3, the transitions of states are as follows:

- With probability $\frac{\mu}{\zeta}$, the next event will be a trade. Suppose that the trade size is u .
 1. If $u < 0$, the state will be updated to $(q, \delta - \lambda u)$.
 2. If $0 \leq u < q$, the state will be updated to $(q - u, \delta - \lambda u)$.
 3. If $u \geq q$, the order value is realized at $\delta - \lambda u$.
- With probability $\frac{\gamma}{\zeta}$, the next event will be a price jump, with jump size J .
 1. If $J > 0$, the order value is realized at $\delta - J$.
 2. If $J \leq 0$, the order value is realized at 0.
- With probability $\frac{\eta^+}{\zeta}$, the next event will be a cancellation, with cancellation fraction ℓ . The state will be updated to $(\ell q, \delta)$, where ℓ is the proportion that remains after the cancellation.

Putting together all of the above, we have the following lemma.

Lemma 1. *The order value process V_t takes the form*

$$V_t = V(q_t, \delta_t),$$

for $t \in [0, \tau^*)$, where $V(\cdot)$ is the unique solution of the equation

$$\begin{aligned} V(q, \delta) &= \frac{\mu}{\zeta} \mathbf{E}_u \left[\mathbb{I}_{\{0 \leq u < q\}} V(q - u, \delta - \lambda u) + \mathbb{I}_{\{u \geq q\}} (\delta - \lambda u) + \mathbb{I}_{\{u < 0\}} V(q, \delta - \lambda u) \right] \\ &\quad + \frac{\gamma}{\zeta} \mathbf{E}_J \left[\mathbb{I}_{\{J > 0\}} (\delta - J) \right] \\ &\quad + \frac{\eta^+}{\zeta} \mathbf{E}_\ell [V(\ell q, \delta)], \end{aligned} \tag{4}$$

for all $(q, \delta) \in \mathbb{R}_+ \times \mathbb{R}$.

In what follows, define the quantities

$$p_u^+ \triangleq P(u > 0), \quad \bar{u}^+ \triangleq E[u\mathbb{I}_{\{u>0\}}], \quad p_J^+ \triangleq P(J > 0), \quad \bar{J}^+ \triangleq E[J\mathbb{I}_{\{J>0\}}].$$

Theorem 1 (Value Function). *The value function $V(q, \delta)$ is linear in δ ; that is, it takes the form*

$$V(q, \delta) = \alpha(q)\delta - \beta(q), \quad (5)$$

where the functions $\alpha: \mathbb{R}_+ \rightarrow \mathbb{R}$ and $\beta: \mathbb{R}_+ \rightarrow \mathbb{R}$ are uniquely determined by the integral equations

$$\begin{aligned} \alpha(q) = & \frac{\mu}{\mu p_u^+ + \gamma + \eta^+} \left\{ p_u^+ + \int_0^q (\alpha(q-x) - 1) f(x) dx \right\} + \frac{\gamma p_J^+}{\mu p_u^+ + \gamma + \eta^+} \\ & + \frac{\eta^+}{\mu p_u^+ + \gamma + \eta^+} \int_0^1 \alpha(\ell q) g(\ell) d\ell, \end{aligned} \quad (6)$$

$$\begin{aligned} \beta(q) = & \frac{\mu}{\mu p_u^+ + \gamma + \eta^+} \left\{ \int_0^q \beta(q-x) f(x) dx + \lambda \int_0^q (\alpha(q-x) - 1) x f(x) dx \right. \\ & \left. - \lambda \bar{u}^+ (\alpha(q) - 1) \right\} + \frac{\gamma \bar{J}^+}{\mu p_u^+ + \gamma + \eta^+} + \frac{\eta^+}{\mu p_u^+ + \gamma + \eta^+} \int_0^1 \beta(\ell q) g(\ell) d\ell, \end{aligned} \quad (7)$$

for $q > 0$, with boundary conditions

$$\alpha(0) = \frac{\mu p_u^+ + \gamma p_J^+}{\mu p_u^+ + \gamma}, \quad \beta(0) = \frac{\mu[\gamma(1 - p_J^+)]}{(\mu p_u^+ + \gamma)^2} \lambda \bar{u}^+ + \frac{\gamma}{\mu p_u^+ + \gamma} \bar{J}^+. \quad (8)$$

Theorem 1 shows that the value function is quasi-linear on the premium δ while the coefficients are determined by the queue position. Specifically, if the order is executed, the agent will earn the premium δ but incur cost $\beta(q)$; if the order is not executed, the order value is just zero. Note that the Volterra integral equations (6)–(7) can be readily solved numerically.

In order to estimate the value function, the following parameters need to be obtained from data:

1. γ/μ , the ratio of arrival rate of jumps to arrival rate of trades.
2. η^+/μ , the ratio of arrival rate of cancellations to arrival rate of trades.
3. $f(\cdot)$, the distribution of trade sizes.
4. λ , the price impact coefficient.
5. $p_J^+ = P(J_i > 0)$, the probability that a price jump is positive.
6. $\bar{J}^+/p_J^+ = E[J_i | J_i > 0]$, then expected value of a positive jump.

Notice that the value function is determined by the ratio of arrival rates rather than their absolute value. Intuitively, ratios of arrival rates determine whether an order is executed, while

their absolute values determine when that happens. As our model does not incorporate time value, absolute values of arrival rates do not change the value of the order. Additionally, we require only the first moment of price jumps rather than their distribution. This is because the size of a price jump is used only to calculate the expected order value at the time that the price jump happens. The distribution of trade size is important as it helps to determine the optionality of an order that has been executed. The price impact coefficient captures the adverse selection cost due to trading, and hence appears only in the expression of $\beta(\cdot)$.

We can now establish the following properties of $\alpha(\cdot)$ and $\beta(\cdot)$:

Theorem 2. 1. Compared with equation (1), we have

$$\alpha_t = \alpha(q), \quad AS_t = \frac{\beta(q)}{\alpha(q)}.$$

2. The probability of execution $\alpha(q)$ is non-increasing in queue position.

3. The adverse selection is positive, i.e.,

$$\beta(q)/\alpha(q) > 0.$$

4. With no cancellations ($\eta = 0$), we have

$$\lim_{q \rightarrow \infty} \alpha(q) = p_J^+, \quad \lim_{q \rightarrow \infty} \beta(q) = \bar{J}^+.$$

The first statement provides intuition for the two value function components. A by-product of the proof shows that the quasi-linear form of the value function in equation (6) is a general result that does not require a Poisson arrival of events.

The second statement shows that the probability of execution is smaller for orders with a larger queue position. This is expected due to the FIFO rule.

The third statement suggests that the adverse selection cost is always positive, which is in line with intuition. Specifically, adverse selection can be broken down into two parts. The first part originates from price jumps, and the second comes from the asymmetric information between liquidity takers and liquidity providers.

The last statement provides the asymptotic behavior of the value function when there is no cancellation. Intuitively, if the queue position is extremely large, it is unlikely that the order will be executed by trades that deplete the queue of higher priority orders. Hence the probability of execution $\alpha(q)$ is just the probability of a positive price jump. The case with cancellations is technically more complicated as we assume that cancellations cause a *shrinking* of the queue length.

While in general it's difficult to obtain close-form solutions to Volterra integral equations, some special cases can be solved using Laplace transform. Theorem 3 provides such an example.

Theorem 3 (Exponential Trade Sizes). *Suppose there are no cancellations and that the trades sizes*

follow a two-sided exponential distribution with parameter $\theta > 0$, i.e.,

$$f(u) \triangleq \frac{\theta}{2} e^{-\theta|u|},$$

for all $u \in \mathbb{R}$. Then, the value function is given by $V(\delta, q) = \alpha(q)\delta - \beta(q)$, where

$$\alpha(q) = p_J^+ + \frac{\mu(1 - p_J^+)}{\mu + 2\gamma} e^{-bq}, \quad (9)$$

$$\beta(q) = \bar{J}^+ \left(1 - \frac{\mu}{\mu/2 + \gamma} e^{-bq}\right) + \frac{\lambda\mu\gamma(p_J^+ - 1)}{2(\gamma + \mu/2)^2\theta} e^{-bq} + \frac{\lambda(\gamma - \mu)\gamma(p_J^+ - 1)}{2(\gamma + \mu/2)^3} q e^{-bq}, \quad (10)$$

for all $q \geq 0$, with $b \triangleq (\gamma + \zeta)\theta/(\mu/2 + \gamma)$.

4. Empirical Calibration

Having laid the framework, we now test our model using NASDAQ ITCH data for a number large-tick U.S. stocks with high liquidity. NASDAQ ITCH data is a so-called *market-by-order* data feed. As opposed to *market-by-level* data, which displays orders accumulated on price, market-by-order data contains all order book events including limit order postings, trades, and limit order cancellations. Market-by-order data makes it possible to reconstruct the limit order book at any given time and hence can be used to view queue position and size of individual orders at a price while remaining anonymous.

One advantage of our model is that it offers predictions of order value at different positions in the queue as a function of market primitives, and hence can be easily calibrated. In this section, we will take Bank of America (BAC) as an example to illustrate our estimation process and model results. We will first describe the calibration of our model parameters, and then solve for the predicted queue position values using the market primitives obtained.

4.1. Data Overview

Our attention is restricted to *large-tick assets*, where the queueing effects are large. Bank of America (BAC) is one of the most liquid stocks traded, with an average daily volume of 88 million shares in August 2013. The bid-ask spread is almost always equal to one tick and is large (about 7 basis points) relative to its price. Hence BAC qualifies as a large-tick asset.

U.S. equities can be traded on multiple exchanges simultaneously. To avoid the complexity of aggregating multiple limit order books, we consider only the NASDAQ order book by using ITCH data, which provides historical data for full order depth. ITCH enables us to track the status of each order from the time it is placed to the time it is either executed or canceled. We use the database of Yahoo Finance for daily closing prices.

4.2. Calibrating Parameters

The main parameters involved in our model are: distribution of order size, trade arrival rate μ , price jump arrival rate γ , cancellation arrival rate η , market impact λ , and jump size J . These parameters exhibit significant day-to-day heterogeneity as some days are more active than others. In what follows, these parameters will be estimated on a daily basis and we will see how their heterogeneity changes order values.

Price jumps are instances when the ask or the bid price changes. A *trade* happens when a market order (or a marketable limit order) is executed with existing limit orders. Sometimes trades and price jumps can coincide. This happens when an execution is large enough to eliminate the entire queue and cause a price jump. In the following analysis, trades will refer to executions that do not cause price moves, while executions that are large enough to deplete the queue will be counted as price jumps. As a result, a price jump can come in the form of an order being executed with arbitrary size.

Price Jumps. In our settings, the size of price jumps is defined by changes in the latent efficient price. Since the latent efficient price is not observable, we assume that the price Δt later is an unbiased estimate of the latent price after a jump.⁸ The intuition here is that the market will take some time (Δt) to digest and factor in the information. Hence, the size of a price jump is calculated as the price change Δt after the price moves. Δt is expected to differ among stocks due to differences in factors such as liquidity. Here, we take Δt to be proportional to the expected time interval between price jumps. Notice that in this case the jump size can be smaller than one tick when a reversion happens within Δt . The number of price jumps is counted separately for both the ask side and the bid side, and then the average is taken. The arrival rate for price jumps is calculated simply by counting price jumps.

Trades. In our model, trade size is defined as the size of an aggressive market order. In electronic markets, once an aggressive market order comes, it is matched with the very first limit order in the queue. If, however, the aggressive market order is too large to be filled with a single limit order, it may trade with multiple resting limit orders, resulting in multiple individual fills. Notice that what we observe from the ITCH data feeds are individual fills, and therefore it is necessary to combine these fills to reconstruct the size of the original market order. We take a time window of two milliseconds, and calculate the order size by putting together the trades of the same side within that time window. If the price changes during that time, we consider the execution to be a price jump.

Our empirical results show that the shape of order size distribution closely resembles a log-normal distribution, which is consistent with findings in Kyle and Obizhaeva (2016). In particular, we obtained the MLE estimate of the mean and standard deviation under this distributional assumption. We obtained the arrival rate, however, in a much more straightforward manner, we simply counted the number of trades.

Cancellations. With market-by-order data, we can keep close track of the position and size of

⁸In our analysis, Δt is set at one minute.

every canceled order. As we mentioned in Section 3, we view each cancellation as a contraction of the whole queue. In other words, we assume that each cancellation decreases the queue size uniformly by a certain proportion l . We then fit the cancellation proportion l using Beta distribution.

Market Impact. The calibration of market impact has always been of great interest in the market microstructure literature. Kyle (1985) suggested linear market impact in a continuous-time theoretical model. He argued that the price impact of one unit of asset is determined by the fundamental volatility and variance of order-flow imbalance. Other researchers, such as Breen et al. (2002), took a purely empirical approach by regressing the price changes on order-flow imbalances. In this paper, we derive the market impact parameter by following the market invariant approach of Kyle and Obizhaeva (2016). Specifically, Kyle and Obizhaeva (2016) proposed a model in which the market impact parameter λ is given by the following equation:

$$\lambda = C(P\sigma)^{\frac{4}{3}}V^{-\frac{2}{3}}, \quad (11)$$

where C is a constant⁹ calibrated from a portfolio transition data set, P is the asset price, σ is the asset's volatility of daily return, and V is the daily trading volume (in shares).

Liquidity Premium. In reality, the latent price is not observable. We will assume that on average it can be approximated by the mid-price. In other words, we will assume that the liquidity premium is a half-spread. However, we will make an adjustment in order to factor in a liquidity rebate of 0.3 ticks offered by NASDAQ. The rebate is offered by the exchange in order to encourage market participants to provide liquidity. Hence the liquidity premium is given by

$$\delta_0 = (\text{half-spread}) + (\text{rebate}) = 0.8 \text{ (ticks)}.$$

Table 1 provides the estimated parameters for Bank of America over 22 trading days. As we can see, the average jump size is very close to one tick, which means that the price process is driven primarily by single tick jumps. Note that the jump size can be less than one tick as we approximate it as the price change Δt after the price moves. Our empirical findings show that the order size distribution is roughly consistent across trading days. The market impact parameter λ too is subject to very little variation across trading days. The only parameters with much variation from day to day are the ratios between arrival rates $(\gamma/\mu, \eta/\mu)$, which, as will see, are the driving force of interday heterogeneity in the order values predicted by our model.

4.3. Observations

Given the market parameters estimated above, the main output of our model is the value function of queue position, which can be obtained by numerically solving equation (6) and (7) in Section 3. Figure 3 provides the plots of the value function, execution probability, and adverse selection for BAC on two representative trading days (8/9/2013 and 8/20/2013).

⁹ $C = 0.0156$ according to Kyle and Obizhaeva (2016).

Date	μ (/min)	Average Trade Size (shares)	Trade Size STD (shares)	η (/min)	Average Cancellation Size (shares)	γ (/min)	Average Jump Size (ticks)	λ	Average Queue Size (shares)
8/30/13	1.43	2270	4793	148.3	1971	0.85	0.86	3.91	46092
8/29/13	1.25	2635	6535	128.8	2103	0.57	0.86	3.97	48454
8/28/13	1.55	2526	4463	140.1	2140	0.79	1.12	4.41	52845
8/27/13	2.25	2435	5395	178.2	2049	1.05	0.91	4.79	48526
8/26/13	1.27	2481	5441	95.5	2114	0.51	0.82	3.53	52434
8/23/13	1.11	2058	3922	114.5	2340	0.56	0.85	3.68	59337
8/22/13	1.20	2998	6082	95.6	1835	0.53	0.92	3.84	37617
8/21/13	1.74	2091	4329	163.7	2134	1.39	1.14	3.93	42240
8/20/13	1.38	3669	7157	127.8	1970	0.95	1.03	4.19	46665
8/19/13	1.21	1979	3676	126.5	1868	0.72	0.84	4.87	45502
8/16/13	1.26	2223	4512	134.5	1737	0.67	0.94	5.37	45480
8/15/13	1.94	1582	3037	165.0	1698	0.82	0.75	5.23	46627
8/14/13	1.37	2034	4449	122.5	1455	0.90	0.95	4.51	33647
8/13/13	1.54	2327	5008	107.8	1628	0.72	1.09	4.68	37320
8/12/13	1.08	2426	4977	103.5	1887	0.49	0.92	4.39	60888
8/9/13	1.65	2762	5986	122.4	1875	0.71	0.91	4.81	50032
8/8/13	1.29	2303	4351	108.4	1764	0.95	1.25	5.16	38409
8/7/13	1.99	2494	5106	130.5	2303	1.01	0.91	5.99	53673
8/6/13	1.47	2610	5692	85.8	1930	0.55	0.93	5.19	50083
8/5/13	0.81	1598	2941	88.3	1502	0.44	0.78	4.59	38647
8/2/13	1.57	2039	4545	112.4	1511	0.53	0.86	5.05	47777
8/1/13	1.58	2853	7978	120.0	1854	0.71	0.85	5.52	51163

Table 1: Estimated market parameters for BAC in a month. λ is estimated as the price impact in basis points for one percent of daily volume. Note that here we consider only shares traded on NASDAQ.

First, as predicted by Theorem 2, the probability of execution is decreasing with queue length and becomes quite flat when the queue length is large. Intuitively, when the queue length is extremely large, the order on the ask side can be executed only by positive price jumps. Hence, the execution probability should converge toward the probability of a positive price jump (p_j^+ as in Theorem 2). Second, the adverse selection cost remains positive and is increasing with queue length. Intuitively, this is because orders at the end of a large queue are more likely to be executed against a large trade. With our assumption of linear price impact, large trades translate to higher adverse selection costs. Third, the order value curve is decreasing as the queue gets longer. From equation (1), we can see that the decreasing value curve is due to a combined effect of decreasing execution probability and increasing adverse selection cost. Fourth, the value difference between an order placed at the very front of the queue and an order placed in a queue length of average was about 0.26 ticks on 8/9/2013 and 0.21 ticks on 8/20/2013, which is comparable to the bid-ask spread. This shows that the queue’s positional value cannot be neglected in higher-level control problems such as optimal execution and market making. Finally, Figure 3 provides comparisons of model outputs on two different trading days. We can see that orders in the same queue position were worth less on 8/20/2013, and had a lower fill probability. This is because the ratio of arrival rate γ/μ was significantly higher on 8/20/2013 (0.69) than on 8/9/2013 (0.43). Intuitively, large γ/μ means that the order is less likely to be executed against a trade before the price changes, and hence translate to a lower fill probability.

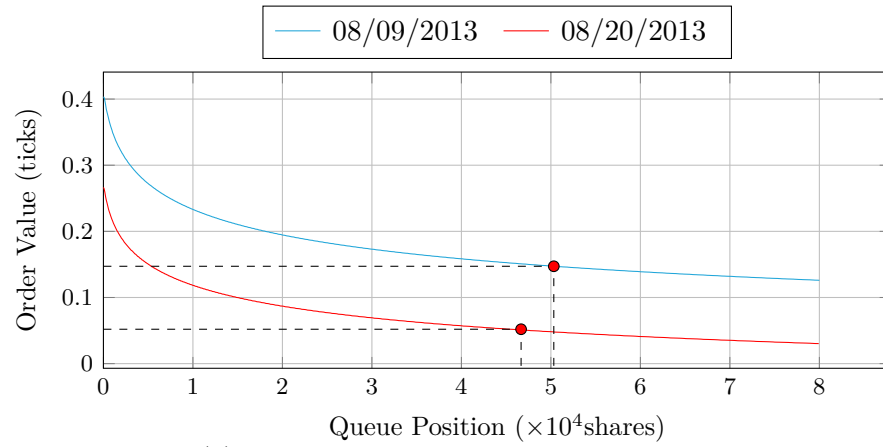
5. Empirical Validation: Backtesting

In the previous section, we calibrated a parametric model to estimate the positional value of limit orders using market data. Now we want to verify these predictions using a non-parametric model based on backtesting. One challenge is that the order value cannot be measured by the profitability of the actual historical orders in the limit order book, since actual orders may have private information. In order to bypass this difficulty, instead of actual orders, we will simulate the outcome of randomly placed artificial orders.

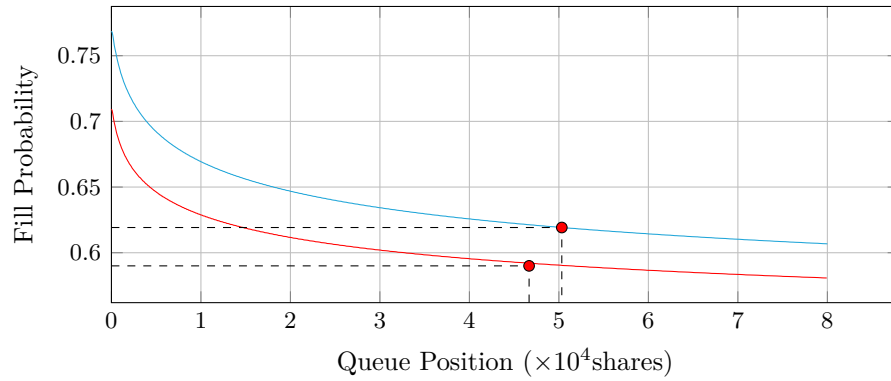
Market-by-order data enables us to simulate the life-span of each artificial order in the limit order book. We can then calculate various statistics such as order value and fill probability for orders at different positions. We then compare the backtesting results with the parametric estimations. More specifically, we restrict our attention to 9 highly liquid U.S. equities or ETFs with a bid/ask spread close to 1 tick. A list of the stocks and their descriptive statistics are given in Table 2.

5.1. Backtesting Simulation

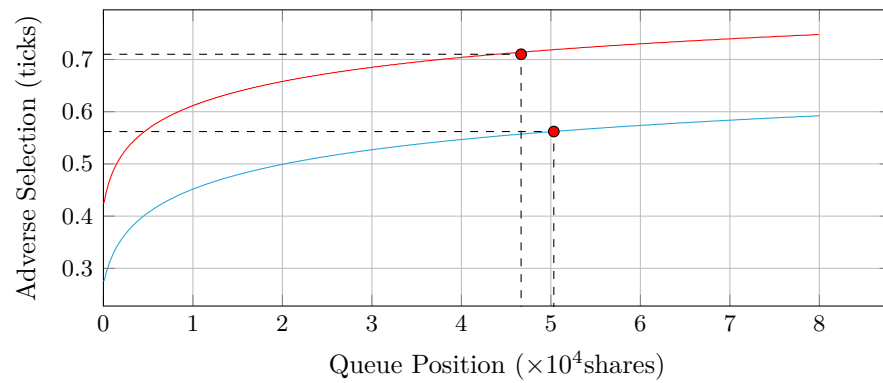
The technique of backtesting is widely used in the financial industry to test a predictive model with existing historical data. Our paper benefited from the advantage of accessing ITCH data, a source of market-by-order data provided by NASDAQ. With full information on historical order/trade data, we were able to construct a simulator to backtest our proposed valuation model. Backtesting



(a) Order value as a function of queue position



(b) Fill probability as a function of queue position



(c) Adverse selection as a function of queue position

Figure 3: Model outputs as functions of queue positions on two different trading days (08/09/2013 and 08/20/2013). The red dots represent the average queue length of that trading day.

	Symbol	Listing Exchange	Price		Average Bid-Ask Spread (\$)	Average Volatility (daily)	Average Daily Volume (shares, $\times 10^6$)
			Low (\$)	High (\$)			
Bank of America	BAC	NYSE	14.11	14.95	1.017	1.2%	87.9
Cisco	CSCO	NASDAQ	23.31	26.38	0.996	1.0%	38.7
General Electric	GE	NYSE	23.11	24.70	1.002	0.9%	29.6
Ford	F	NYSE	15.88	17.50	1.005	1.4%	33.6
Intel	INTC	NASDAQ	21.90	23.22	1.005	1.1%	24.5
Pfizer	PFE	NYSE	28.00	29.37	1.007	0.7%	23.3
Petroleo Brasileiro	PBR	NYSE	13.39	14.98	1.010	2.6%	17.9
iShares MSCI Emerging Markets	EEM	NYSE	37.35	40.10	1.006	1.2%	64.1
iShares MSCI EAFE	EFA	NYSE	59.17	62.10	1.021	0.7%	14.4

Table 2: Descriptive statistics for 9 stocks over the 21 trading days of August 2013. The average bid/ask spread is defined as the time average computed from the ITCH data. The volatility is defined as the standard deviation of percentage daily returns. All other statistics were retrieved from Yahoo Finance.

with artificial orders requires one assumption: real orders may influence other market participants, while counterfactual artificial orders clearly do not. Here we will assume that all of the artificial orders are of infinitesimal size and hence have no market impact. This is in consistent with our model assumptions. First, the historical data will be used to recreate the state and dynamics of order book; then artificial orders will be placed and processed according to market rules; finally, the value of the artificial orders will be calculated. The details of our procedure are as follows:

Placement of Artificial Orders. We start by defining two types of artificial orders based on the position at which they are inserted.

- *Regular orders* are orders that are appended to the end of the queue at the current best price. The name *regular orders* comes from the fact that these orders are placed according to the FIFO rule.
- *Touch orders* are orders that are inserted at the very front of the queue at the current best price. These orders are used to evaluate the value of being placed at the front of the queue. Comparing touch orders with regular orders will help to illustrate the magnitude of the effect of the value of queue position.

In the simulation, we associate each real limit order with an *entry-time stamp* to keep track of the time that the order entered the order book. The side (bid or ask) of each artificial order is randomly picked. Suppose that it is an ask order; then its evolution in the limit order book will be as follows.

- At a random time, the artificial order is generated and inserted at the end of the queue at the then current best ask price.
- The artificial order is processed following the market rule of price/time priority. We start updating the limit order book according to the real data until one of the following events occurs.

1. *New order arrival*: If a new limit order is added to the same side at a better price (lower for the ask side, higher for the bid side) than that of the artificial order, then the artificial order will no longer be at the best price, and we will assume that it is canceled immediately.
2. *Fill*: If a limit order at the same price that has arrived after the artificial order is filled, we will assume that the artificial order is also filled.
3. *Cancellation*: If the price moves because all other orders in the queue are canceled (which is rare), we will assume that the artificial order is canceled as well.

In order to eliminate outliers, we ignore the first and last half hours of the trading day. Accordingly, we pick 1000 time points uniformly at random between 10:00 and 15:30 on each trading day, and insert an artificial order on a random side of the market at each of these times.

Order Valuation. If the artificial order is canceled then it possesses no value. If, however, the artificial order is filled then its value will be the difference between the execution price and the fundamental value of the asset. In order to backtest order values at different positions, we need to determine the fundamental value. Since the fundamental value cannot be observed directly in the historical data, we need to calibrate it through a tractable valuation process. In this paper, we assume that the mid-price one minute after the order’s execution is an unbiased point estimate of the fundamental value at the time of execution. This is a noisy approximation and requires many observations for a reasonably accurate estimate. Hence, we choose to estimate the average order value over all orders placed across 30 trading days instead of using a shorter period.

5.2. Observations

Table 3 shows the comparison of the results from backtesting and model outputs. The order value measures the value of regular orders that are placed at the end of the queue, while the touch value measures the value of touch orders placed at the very front of the queue.

Symbol	Order Value		Fill Probability		Adverse Selection		Touch Value	
	Model (ticks)	Simulation (ticks)	Model	Simulation	Model (ticks)	Simulation (ticks)	Model (ticks)	Simulation (ticks)
BAC	0.14	0.14	0.62	0.60	0.57	0.57	0.36	0.31
CSCO	0.08	0.07	0.63	0.59	0.68	0.68	0.24	0.21
GE	0.08	0.09	0.62	0.60	0.67	0.65	0.19	0.23
F	0.13	0.15	0.65	0.64	0.60	0.53	0.24	0.23
INTC	0.11	0.09	0.64	0.61	0.63	0.56	0.28	0.23
PFE	0.12	0.11	0.63	0.58	0.62	0.61	0.16	0.21
PBR	-0.03	-0.04	0.57	0.53	0.85	0.89	0.03	0.03
EMM	0.07	0.08	0.63	0.63	0.69	0.64	0.21	0.15
EFA	0.03	0.04	0.57	0.53	0.74	0.73	0.06	0.09

Table 3: Estimated model values vs. simulation values. All the values above were calculated as the average across 30 trading days. Touch value refers to the value of orders at the very front of the queue.

We can see that the values estimated from our model are very close to the backtesting results. Further, if we break down the value into fill probability and adverse selection cost, we can see that the values are also close. This shows that our model provides a good approximation of the value of queue positions.

Notice that the difference between the value of touch orders and the value of regular orders provides information about the magnitude of the value of queue position. First of all, the value of orders placed at the front of the queue is always larger than the value of orders placed at the end. This shows that better queue position does carry an advantage. Second, the magnitude of the gap differs between symbols. For some symbols, such as BAC and CSCO, the gap can be very large and comparable to the bid ask spread (> 0.1 ticks). For others, such as PFE and PBR, the gap is less prominent (< 0.1 ticks). These differences are consistent in both the predictions made by our model and by the values estimated using the non-parametric backtest.

5.3. Discussion

In this section, we provide a framework based on backtesting to estimate the value of queue positions. This non-parametric approach enables us to test the accuracy of our model. This leads to a natural question: if a non-parametric, model-free estimation method is available, why do we need a parametric method, such as the one discussed in this paper? The reasons is as follows. In the backtest, the value of artificial orders is estimated through price changes post execution. These estimates are very noisy, and hence many independent observations are needed to obtain accurate estimates. As a result, the non-parametric can be used only to estimate the average value of orders across large intervals of time (e.g., 30 days). However, market parameters, such as arrival rates of order book events, are constantly changing on a daily or even intraday basis. Backtesting cannot capture this variation. On the other hand, the estimates from our model are conditional on market primitives that can be estimated in real time and hence provide more precise predictions in real time.

6. Concluding Remarks

In this paper, we exhibited a dynamic model for valuing queue position in limit order books. We provided analytic evidence for sizable difference in values for orders at different queue positions. We specifically quantified the disadvantage of bad queue positions that originate from decreasing execution probability and increasing adverse selection costs.

The formulation of the model is based entirely on observable quantities so that the parameters can be estimated from market data. This tractability allowed us to calibrate our model empirically. We further validated the model by comparing the outputs with results from backtesting simulations.

This analysis has practical implications for both market participants and regulators:

1. For large tick-size assets, queueing effects can be very significant.

2. Accounting for queue position cannot be ignored when solving market making or algorithmic trading problems in large-tick assets.
3. The value embedded in the queue position rewards the trading speed of high-frequency firms. This creates a disadvantage for individual traders who have less or no access to fast-trading technologies. From a regulatory level, an important question is whether this time-price priority rule is a good mechanism for organizing exchanges for the trading of large-tick assets.

References

- M. Avellaneda, J. Reed, and S. Stoikov. Forecasting prices from level-I quotes in the presence of hidden liquidity. *Algorithmic Finance*, 1(1):35–43, 2011.
- B. Biais, P. Hillion, and C. Spatt. An empirical analysis of the limit order book and the order flow in the Paris bourse. *The Journal of Finance*, 50(5):1655–1689, 1995.
- J. Blanchet and X. Chen. Continuous-time modeling of bid-ask spread and price dynamics in limit order books. *arXiv preprint arXiv:1310.1103*, 2013.
- J. Bouchaud, J. Kockelkoren, and M. Potters. Random walks, liquidity molasses and critical response in financial markets. *Quantitative Finance*, 6(02):115–123, 2006.
- W. J. Breen, L. S. Hodrick, and R. A. Korajczyk. Predicting equity liquidity. *Management Science*, 48(4):470–483, 2002.
- G. C. Chacko, J. W. Jurek, and E. Stafford. The price of immediacy. *The Journal of Finance*, 63(3):1253–1290, 2008.
- R. Cont and A. De Larrard. Price dynamics in a Markovian limit order market. *SIAM Journal on Financial Mathematics*, 4(1):1–25, 2013.
- R. Cont and A. Kukanov. Optimal order placement in limit order markets. *Available at SSRN 2155218*, 2013.
- R. Cont, S. Stoikov, and R. Talreja. A stochastic model for order book dynamics. *Operations Research*, 58(3):549–563, 2010.
- R. Cont, A. Kukanov, and S. Stoikov. The price impact of order book events. *Journal of Financial Econometrics*, 12(1):47–88, 2014.
- T. E. Copeland and D. Galai. Information effects on the bid-ask spread. *The Journal of Finance*, 38(5):1457–1469, 1983.
- Z. Eisler, J. Bouchaud, and J. Kockelkoren. The price impact of order book events: market orders, limit orders and cancellations. *Quantitative Finance*, 12(9):1395–1419, 2012.

- L. R. Glosten. Is the electronic open limit order book inevitable? *The Journal of Finance*, 49(4): 1127–1161, 1994.
- L. R. Glosten and P. R. Milgrom. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*, 14(1):71–100, 1985.
- M. D. Griffiths, B. F. Smith, D. A. S. Turnbull, and R. W. White. The costs and determinants of order aggressiveness. *Journal of Financial Economics*, 56(1):65–88, 2000.
- X. Guo, A. De Larrard, and Z. Ruan. Optimal placement in a limit order book. *Preprint*, 2013.
- W. Hackbusch. *Integral Equations: Theory and Numerical Treatment*, volume 120. Springer Science & Business Media, 1995.
- B. Hollifield, R. A. Miller, and P. Sandås. Empirical analysis of limit order markets. *The Review of Economic Studies*, 71(4):1027–1063, 2004.
- I. Karatzas and S. Shreve. *Brownian Motion and Stochastic Calculus*, volume 113. Springer Science & Business Media, 2012.
- A. S. Kyle. Continuous auctions and insider trading. *Econometrica*, pages 1315–1335, 1985.
- A. S. Kyle and A. Obizhaeva. Market microstructure invariance: Empirical hypotheses. *Econometrica*, 84(4):1345–1404, 2016.
- P. Lakner, J. Reed, and S. Stoikov. High frequency asymptotics for the limit order book. *Preprint*, 2013.
- C. Maglaras, C. C. Moallemi, and H. Zheng. Optimal execution in a limit order book and an associated microstructure market impact model. *Available at SSRN*, 2015.
- S. M. Ross. *Stochastic Processes*, volume 2. John Wiley & Sons, 1996.

Supplementary Material

A. Theorem Proofs

Theorem 1 (Value Function). *The value function $V(q, \delta)$ is linear in δ ; that is, it takes the form*

$$V(q, \delta) = \alpha(q)\delta - \beta(q), \quad (5)$$

where the functions $\alpha: \mathbb{R}_+ \rightarrow \mathbb{R}$ and $\beta: \mathbb{R}_+ \rightarrow \mathbb{R}$ are uniquely determined by the integral equations

$$\begin{aligned} \alpha(q) = & \frac{\mu}{\mu p_u^+ + \gamma + \eta^+} \left\{ p_u^+ + \int_0^q (\alpha(q-x) - 1) f(x) dx \right\} + \frac{\gamma p_J^+}{\mu p_u^+ + \gamma + \eta^+} \\ & + \frac{\eta^+}{\mu p_u^+ + \gamma + \eta^+} \int_0^1 \alpha(\ell q) g(\ell) d\ell, \end{aligned} \quad (6)$$

$$\begin{aligned} \beta(q) = & \frac{\mu}{\mu p_u^+ + \gamma + \eta^+} \left\{ \int_0^q \beta(q-x) f(x) dx + \lambda \int_0^q (\alpha(q-x) - 1) x f(x) dx \right. \\ & \left. - \lambda \bar{u}^+ (\alpha(q) - 1) \right\} + \frac{\gamma \bar{J}^+}{\mu p_u^+ + \gamma + \eta^+} + \frac{\eta^+}{\mu p_u^+ + \gamma + \eta^+} \int_0^1 \beta(\ell q) g(\ell) d\ell, \end{aligned} \quad (7)$$

for $q > 0$, with boundary conditions

$$\alpha(0) = \frac{\mu p_u^+ + \gamma p_J^+}{\mu p_u^+ + \gamma}, \quad \beta(0) = \frac{\mu[\gamma(1 - p_J^+)]}{(\mu p_u^+ + \gamma)^2} \lambda \bar{u}^+ + \frac{\gamma}{\mu p_u^+ + \gamma} \bar{J}^+. \quad (8)$$

Proof. First of all, we solve for the solution to equation (4). The boundary condition can be verified by setting $q = 0$ in (4), which gives

$$\begin{aligned} V(0, \delta) = & \frac{\mu}{\zeta} \mathbf{E} \left[\mathbb{I}_{\{u \geq 0\}} (\delta - \lambda u) + \mathbb{I}_{\{u < 0\}} V(q, \delta - \lambda u) \right] \\ & + \frac{\gamma}{\zeta} \mathbf{E} \left[\mathbb{I}_{\{J > 0\}} (\delta - J) \right] \\ & + \frac{\eta^+}{\zeta} V(0, \delta). \end{aligned}$$

Notice that it's an integral equation with a linear drift on δ . Hence the solution of $V(0, \delta)$ should also be linear on δ . The equation above thus boils down to

$$\begin{aligned} \frac{\mu + \gamma}{\mu} (\alpha(0)\delta - \beta(0)) = & \mathbf{E} \left[\mathbb{I}_{\{u > 0\}} (\delta - \lambda u) \right] + \mathbf{E} \left[\mathbb{I}_{\{u \leq 0\}} (\alpha(0)(\delta - \lambda u) - \beta(0)) \right] + \gamma(\delta - \bar{J}^+)/\mu \\ = & p^+ \delta - \lambda \int_0^{+\infty} u f(u) du + \int_{-\infty}^0 (\alpha(0)(\delta - \lambda u) - \beta(0)) f(u) du + \gamma(\delta - \bar{J}^+)/\mu. \end{aligned}$$

Solving the equation above for $\alpha(0)$ and $\beta(0)$, we obtain the boundary condition:

$$\alpha(0) = \frac{\mu p_u^+ + \gamma p_J^+}{\mu p_u^+ + \gamma}, \quad \beta(0) = \frac{\mu[\gamma(1 - P_J^+)]}{(\mu p_u^+ + \gamma)^2} \lambda \bar{u}^+ + \frac{\gamma}{\mu p_u^+ + \gamma} \bar{J}^+.$$

For $q > 0$, the integral equation still has a linear drift, which provides the linearity of the solution:

$$\begin{aligned} \frac{\mu + \gamma + \eta^+}{\mu} (\alpha(q)\delta - \beta(q)) &= \mathbf{E} [\mathbb{I}_{\{u > q\}} (\delta - \lambda u)] + \mathbf{E} [\mathbb{I}_{\{u \leq 0\}} (\alpha(q)(\delta - \lambda u) - \beta(q))] \\ &\quad + \mathbf{E} [\mathbb{I}_{\{0 < u \leq q\}} (\alpha(q - u)(\delta - \lambda u) - \beta(q - u))] \\ &\quad + \gamma(\delta - \bar{J}^+)/\mu + \eta^+ \mathbf{E} [\alpha(\ell q)\delta - \beta(\ell q)] \\ &= \int_q^{+\infty} (\delta - \lambda u) f(u) du + \int_0^q (\alpha(q - u)(\delta - \lambda u) - \beta(q - u)) f(u) du \\ &\quad + \int_{-\infty}^0 (\alpha(q)(\delta - \lambda u) - \beta(q)) f(u) du \\ &\quad + \gamma(\delta - \bar{J}^+)/\mu + \frac{\eta^+}{\mu} \int_0^1 (\alpha(\ell q)\delta - \beta(\ell q)) d\ell. \end{aligned}$$

Solving the equation for $\alpha(q)$ and $\beta(q)$, we obtain the solution:

$$V(q, \delta) = \alpha(q)\delta - \beta(q).$$

Now we would like to prove the uniqueness and existence of the solution to equations (6) and (7). Notice that $\alpha(\cdot)$ is defined on \mathbb{R}^+ . Then the expression of $\alpha(q)$ is a Volterra integral equation of the second kind. We can rewrite the expression as follows:

$$\alpha(q) = k_1(q) + \int_0^q k_2(x, q, \alpha(x)) dx, \quad \forall q \in \mathbb{R}^+, \quad (12)$$

where

$$\begin{aligned} k_1(q) &= \frac{\mu}{\mu p_u^+ + \gamma + \eta^+} (p_u^+ - \int_0^q f(x) dx) + \frac{\gamma p_J^+}{\mu p_u^+ + \gamma + \eta^+}, \quad \forall q \in \mathbb{R}^+, \\ k_2(x, q, z) &= \left\{ \frac{\mu f(q - x)}{\mu p_u^+ + \gamma + \eta^+} + \frac{\eta^+ g(x/q)/q}{\mu p_u^+ + \gamma + \eta^+} \right\} z, \quad \forall q \in \mathbb{R}^+, x \in \mathbb{R}^+, z \in \mathbb{R}. \end{aligned}$$

Given the continuity of $f(\cdot), g(\cdot)$, we have $k_1 \in C(\mathbb{R}^+), k_2 \in C(\mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R})$. Also, it is trivial that k_2 satisfies the following Lipschitz condition:

$$|k_2(x, q, z) - k_2(x, q, z')| \leq L(x, q) |z - z'|, \text{ for some } L \in C(\mathbb{R}^+ \times \mathbb{R}^+).$$

Hence by Theorem 2.1.1 of Hackbusch (1995), there is exactly one solution of the integral equation (12). Additionally, the solution $\alpha(\cdot)$ is continuous on \mathbb{R}^+ .

The existence and uniqueness of $\beta(\cdot)$ can be established in a similar way. Specifically, we can

write equation (7) in the following form:

$$\beta(q) = k_3(q) + \int_0^q k_4(x, q, \beta(x)) dx, \quad \forall q \in \mathbb{R}^+, \quad (13)$$

where

$$k_3(q) = \frac{\mu}{\mu p_u^+ + \gamma + \eta^+} \left\{ \lambda \int_0^q (\alpha(q-x) - 1) x f(x) dx - \lambda \bar{u}^+ (\alpha(q) - 1) \right\} + \frac{\gamma \bar{J}^+}{\mu p_u^+ + \gamma + \eta^+}, \quad \forall q \in \mathbb{R}^+,$$

$$k_4(x, q, z) = k_2(x, q, z), \quad \forall q \in \mathbb{R}^+, x \in \mathbb{R}^+, z \in \mathbb{R}.$$

Hence, by a similar analysis, there is exactly one solution to integral equation (7), and that solution is continuous. ■

Theorem 2. 1. Compared with equation (1), we have

$$\alpha_t = \alpha(q), \quad AS_t = \frac{\beta(q)}{\alpha(q)}.$$

2. The probability of execution $\alpha(q)$ is non-increasing in queue position.

3. The adverse selection is positive, i.e.,

$$\beta(q)/\alpha(q) > 0.$$

4. With no cancellations ($\eta = 0$), we have

$$\lim_{q \rightarrow \infty} \alpha(q) = p_j^+, \quad \lim_{q \rightarrow \infty} \beta(q) = \bar{J}^+.$$

Proof. 1. Consider an order placed on the ask side at position q at time 0; denote τ^* to be the time that it is filled or canceled.

$$\begin{aligned} V(q, \delta) &= \mathbb{E} \left[(P^A - P) \mathbb{I}_{\{\text{FILL}\}} \right] \\ &= \mathbb{E} \left[\left((P^A - P_0) - (P - P_0) \right) \mathbb{I}_{\{\text{FILL}\}} \right] \\ &= \mathbb{E} \left[(\delta - (P - P_0)) \mathbb{I}_{\{\text{FILL}\}} \right] \\ &= \mathbb{P}(\text{FILL}) \delta - \mathbb{P}(\text{FILL}) \mathbb{E} [P_{\tau^*} - P_0 | \text{FILL}]. \end{aligned} \quad (14)$$

Notice that $\mathbb{E} [P_{\tau^*} - P_0 | \text{FILL}]$ represents the opportunity cost conditional on executing the order, which coincides with the definition of adverse selection.

Compared to the notations in equation (5), it is easy to see that

$$\alpha(q) = \mathbb{P}(\text{FILL}) \quad \beta(q)/\alpha(q) = \mathbb{E}[P_{\tau^*} - P_0 | \text{FILL}].$$

Hence $\alpha(q)$ is exactly the probability of the order being executed, and $\beta(q)/\alpha(q)$ represents the adverse selection cost.

2. It suffices to show that $\forall 0 \leq q_0 < q_1$, we have $\alpha(q_0) \geq \alpha(q_1)$.

Consider an infinitesimal order A_0 with a queue position q_0 , and let \mathcal{E}_0 be the set of events that the order is eventually filled. Then we have

$$\alpha(q_0) = \mathbb{P}(\mathcal{E}_0).$$

Notice that in our model, the value of an order does not depend on the orders that follows it in the queue. Hence it is possible to *couple* the order A_0 with an infinitesimal order A_1 in the exact same queue but with a position q_1 . Similarly, we define \mathcal{E}_1 to be the set of events that A_1 is eventually executed.

Notice that since the size of the orders is infinitesimal, the marginal probabilities $\mathbb{P}(\mathcal{E}_0), \mathbb{P}(\mathcal{E}_1)$ should be intact with *coupling*. There are two scenarios where \mathcal{E}_1 can happen:

- A_1 is executed by a trade. Then, in our setup, there can be no price jump before this event. As A_0 is placed in front of A_1 , it should be executed already.
- A_1 is executed by a positive price jump. In our setup, there can be no negative price jump before this event. Hence A_0 can be executed either by this positive price jump or by an earlier trade.

The above analysis shows that $\{\mathcal{E}_1\} \subseteq \{\mathcal{E}_0\}$; hence

$$\alpha(q_1) = \mathbb{P}(\mathcal{E}_1) \leq \mathbb{P}(\mathcal{E}_0) = \alpha(q_0).$$

3. Since $\forall q > 0, 0 < \alpha(q) < 1$, it suffices to show that $\forall q > 0, \beta(q) > 0$.

We have already proved that $\alpha(q)$ is increasing in q ; hence $\forall 0 \leq x < q, \alpha(q-x) \geq \alpha(q)$. According to equation (6), we have

$$\begin{aligned} \beta(q) \geq & \frac{\mu}{\mu p_u^+ + \gamma + \eta^+} \left\{ \int_0^q \beta(q-x) f(x) dx - \lambda \int_q^\infty (\alpha(q) - 1) x f(x) dx \right\} \\ & + \frac{\gamma \bar{J}^+}{\mu p_u^+ + \gamma + \eta^+} + \frac{\eta^+}{\mu p_u^+ + \gamma + \eta^+} \int_0^1 \beta(\ell q) g(\ell) d\ell. \end{aligned} \quad (15)$$

Notice that we have $\beta(0) > 0$ and that $\beta(\cdot)$ is continuous. Now suppose that $\beta(q)$ is not always positive for $q \geq 0$; then there must exist q_0 such that $\beta(\cdot)$ attains a value of zero for the first time. By continuity, we have $\beta(q) > 0$ for $q \in [0, q_0)$. Notice that at q_0 , the

right-hand side of equation (15), is strictly positive; hence it is impossible that $\beta(q_0) = 0$. As a result, it must be that $\beta(\cdot)$ is positive for all $q \geq 0$.

4. Given that p.d.f. of trade size $f(\cdot)$ is assumed to be continuous over $[0, +\infty)$, we have

$$\int_0^\infty e^{-st} f(t) dt \leq \int_0^\infty f(t) dt = p_u^+, \forall s \geq 0.$$

Hence the Laplace transform of $f(\cdot)$ exists on $[0, +\infty)$. Let $P(s)$ denote the Laplace transform of p.d.f. $f(\cdot)$ of trade size and define $C(q) = \alpha(q) - 1$. We have

$$C(q) = \frac{\mu}{\mu p_u^+ + \gamma} \int_0^q C(q-x) f(x) dx + \frac{\gamma(p_J^+ - 1)}{\mu p_u^+ + \gamma}. \quad (16)$$

Now take the Laplace transform on both sides of equation (16); we have

$$\begin{aligned} \mathcal{L}\{C\}(s) &= \frac{\mu}{\mu p_u^+ + \gamma} \mathcal{L}\{C\}(s) P(s) + \frac{\gamma(p_J^+ - 1)}{s(\mu p_u^+ + \gamma)}, \\ \Rightarrow \frac{\mu p_u^+ + \gamma - \mu P(s)}{\mu p_u^+ + \gamma} \mathcal{L}\{C\}(s) &= \frac{\gamma(p_J^+ - 1)}{s(\mu p_u^+ + \gamma)}. \end{aligned}$$

Given the fact that $\forall s \geq 0, P(s) \leq p_u^+$, we have $\mu p_u^+ + \gamma - \mu P(s) > 0$. As a result, the Laplace transform of $C(q)$ is well defined on $[0, +\infty)$ and takes the form

$$\mathcal{L}\{C\}(s) = \frac{\gamma(p_J^+ - 1)}{s(\mu p_u^+ + \gamma - \mu P(s))}. \quad (17)$$

Hence, the Laplace transform for $\alpha(q)$ is

$$\mathcal{L}\{\alpha\}(s) = \mathcal{L}\{C\}(s) + 1/s = \frac{\gamma(p_J^+ - 1)}{s(\mu p_u^+ + \gamma - \mu P(s))} + 1/s. \quad (18)$$

By the final value theorem of Laplace transform, we have

$$\lim_{q \rightarrow \infty} \alpha(q) = \lim_{s \rightarrow 0} s \mathcal{L}\{\alpha\}(s) = \frac{-\gamma(p_J^+ - 1)}{\mu p + \gamma - \mu P(0)} + 1 = p_J^+.$$

Similarly, it is easy to see that the Laplace transform of $\beta(q)$ is also well defined on $[0, +\infty)$; hence we have

$$\mathcal{L}\{\beta\}(s) = -\frac{\mu}{\mu p_u^+ + \gamma - \mu P(s)} [\lambda \mathcal{L}\{C\}(s) P'(s) + \lambda \bar{u}^+ \mathcal{L}\{C\}(s) - \gamma \bar{J}^+ / (s\mu)]. \quad (19)$$

Then by the finite value theorem of Laplace transform:

$$\begin{aligned}
\lim_{q \rightarrow \infty} \beta(q) &= \lim_{s \rightarrow 0} s\mathcal{L}\{\beta\}(s) \\
&= \lim_{s \rightarrow 0} -\frac{\mu}{\mu p_u^+ + \gamma - \mu P(s)} [\lambda s\mathcal{L}\{C\}(s)P'(s) + \lambda \bar{u}^+ s\mathcal{L}\{C\}(s) - \gamma \bar{J}^+ / \mu] \\
&= \bar{J}^+.
\end{aligned} \tag{20}$$

■

Theorem 3 (Exponential Trade Sizes). *Suppose there are no cancellations and that the trades sizes follow a two-sided exponential distribution with parameter $\theta > 0$, i.e.,*

$$f(u) \triangleq \frac{\theta}{2} e^{-\theta|u|},$$

for all $u \in \mathbb{R}$. Then, the value function is given by $V(\delta, q) = \alpha(q)\delta - \beta(q)$, where

$$\alpha(q) = p_J^+ + \frac{\mu(1 - p_J^+)}{\mu + 2\gamma} e^{-bq}, \tag{9}$$

$$\beta(q) = \bar{J}^+ \left(1 - \frac{\mu}{\mu/2 + \gamma} e^{-bq}\right) + \frac{\lambda\mu\gamma(p_J^+ - 1)}{2(\gamma + \mu/2)^2\theta} e^{-bq} + \frac{\lambda(\gamma - \mu)\gamma(p_J^+ - 1)}{2(\gamma + \mu/2)^3} q e^{-bq}, \tag{10}$$

for all $q \geq 0$, with $b \triangleq (\gamma + \zeta)\theta/(\mu/2 + \gamma)$.

Proof. First denote $P(s)$ as the Laplace transform of the truncated p.d.f. of trade size on the positive domain ($f(u) = \frac{\theta}{2} e^{-\theta u}$). We have

$$P(s) = \frac{\theta}{2(s + \theta)} \quad \bar{u}^+ = \frac{1}{2\theta}. \tag{21}$$

Plugging equation (21) into (18), we obtain the Laplace transform of $\alpha(q)$:

$$\mathcal{L}\{\alpha\}(s) = \frac{\gamma(p_J^+ - 1)}{s(\mu p_u^+ + \gamma - \mu P(s))} + 1/s. \tag{22}$$

Then, by taking the inverse Laplace transform, we get

$$\alpha(q) = p_J^+ + \frac{\mu(1 - p_J^+)}{\mu + 2\gamma} e^{-\frac{\gamma\theta q}{\mu/2 + \gamma}}. \tag{23}$$

Similarly, we can plug equation (21) into (19) to obtain the Laplace transform of $\beta(q)$:

$$\mathcal{L}\{\beta\}(s) = \frac{\bar{J}^+}{s} - \frac{\mu}{\mu + 2\gamma} \frac{\bar{J}^+}{s + b} + \frac{\lambda\mu\gamma(p_J^+ - 1)}{2(\gamma + \mu/2)^2\theta(s + b)} + \frac{\lambda(\gamma - \mu)\gamma(p_J^+ - 1)}{2(\gamma + \mu/2)^3(s + b)^2}. \tag{24}$$

where $b = \frac{\gamma\theta}{\gamma+\mu/2}$. Taking the inverse Laplace Transform, we get

$$\beta(q) = \bar{J}^+ \left(1 - \frac{\mu}{\mu + 2\gamma} e^{-bq} \right) + \frac{\lambda\mu\gamma(p_J^+ - 1)}{2(\gamma + \mu/2)^2\theta} e^{-bq} + \frac{\lambda(\gamma - \mu)\gamma(p_J^+ - 1)}{2(\gamma + \mu/2)^3} q e^{-bq}. \quad (25)$$

■