Operations Research

A Stochastic Model for Order Book Dynamics

Rama Cont, Sasha Stoikov, Rishi Talreja,

A Stochastic Model for Order Book Dynamics

Rama Cont
Department of Industrial Engineering and Operations Research, Columbia University, New York, New York 10027,
rama.cont@columbia.edu

Sasha Stoikov
Cornell Financial Engineering Manhattan, New York, New York 10004,
sashastoikov@gmail.com

Rishi Talreja
Department of Industrial Engineering and Operations Research, Columbia University, New York, New York 10027,
rt2146@columbia.edu

We propose a continuous-time stochastic model for the dynamics of a limit order book. The model strikes a balance between three desirable features: it can be estimated easily from data, it captures key empirical properties of order book dynamics, and its analytical tractability allows for fast computation of various quantities of interest without resorting to simulation. We describe a simple parameter estimation procedure based on high-frequency observations of the order book and illustrate the results on data from the Tokyo Stock Exchange. Using simple matrix computations and Laplace transform methods, we are able to efficiently compute probabilities of various events, conditional on the state of the order book: an increase in the midprice, execution of an order at the bid before the ask quote moves, and execution of both a buy and a sell order at the best quotes before the price moves. Using high-frequency data, we show that our model can effectively capture the short-term dynamics of a limit order book. We also evaluate the performance of a simple trading strategy based on our results.

The evolution of prices in financial markets results from the interaction of buy and sell orders through a rather complex dynamic process. Studies of the mechanisms involved in trading financial assets have traditionally focused on *quote-driven* markets, where a market maker or dealer centralizes buy and sell orders and provides liquidity by setting bid and ask quotes. The NYSE specialist system is an example of this mechanism. In recent years, electronic communications networks (ECNs) such as Archipelago, Instinet, Brut, and Tradebook have captured a large share of the order flow by providing an alternative *order-driven* trading system. These electronic platforms aggregate all outstanding limit orders in a *limit order book* that is available to market participants and market orders are executed against the best available prices. As a result of the ECN's popularity, established exchanges such as the NYSE, NASDAQ, the Tokyo Stock Exchange, and the London Stock Exchange have adopted electronic order-driven platforms, either fully or partially through "hybrid" systems.

The absence of a centralized market maker, the mechanical nature of execution of orders and, last but not least, the availability of data have made order-driven markets interesting candidates for stochastic modelling. At a fundamental level, models of order book dynamics may provide some insight into the interplay between order flow, liquidity, and price dynamics (Bouchaud et al. 2002, Smith et al. 2003, Farmer et al. 2004, Foucault et al. 2005). At the level of applications, such models provide a quantitative framework in which investors and trading desks can optimize trade execution strategies (Alfonsi et al. 2010, Obizhaeva and Wang 2006). An important motivation for modelling high-frequency dynamics of order books, is to use the information on the current state of the order book to predict its short-term behavior. We focus, therefore, on *conditional probabilities* of events, given the state of the order book.

The dynamics of a limit order book resembles in many aspects that of a queuing system. Limit orders wait in a queue to be executed against market orders (or canceled). Drawing inspiration from this analogy, we model a limit order book as a continuous-time Markov process that tracks the number of limit orders at each price level in the book. The model strikes a balance between three desirable features: it can be estimated easily using high-frequency data, it reproduces various empirical features of order books, and it is analytically tractable. In particular, we show that our model is simple enough to allow the use of Laplace transform techniques from the queuing literature to compute various *conditional probabilities*. These include the probability of the midprice increasing in the next move, the

probability of executing an order at the bid before the ask quote moves, and the probability of executing both a buy and a sell order at the best quotes before the price moves, given the state of the order book. Although here we only focus on these events, the methods we introduce allow one to compute conditional probabilities involving much more general events such as those involving latency associated with order processing (see Remark 1). We illustrate our techniques on a model estimated from order book data for a stock on the Tokyo Stock Exchange.

Related literature. Various recent studies have focused on limit order books. Given the complexity of the structure and dynamics of order books, it has been difficult to construct models that are both statistically realistic and amenable to rigorous quantitative analysis. Parlour (1998), Foucault et al. (2005), and Rosu (2009) propose equilibrium models of limit order books. These models provide interesting insights into the price formation process, but contain unobservable parameters that govern agent preferences. Thus, they are difficult to estimate and use in applications. Some empirical studies on properties of limit order books are Bouchaud et al. (2002), Farmer et al. (2004), and Hollifield et al. (2004). These studies provide an extensive list of statistical features of order book dynamics that are challenging to incorporate in a single model. Bouchaud et al. (2008), Smith et al. (2003), Bovier et al. (2006), Luckock (2003), and Maslov and Mills (2001) propose stochastic models of order book dynamics in the spirit of the one proposed here, but focus on unconditional/steady–state distributions of various quantities rather than the conditional quantities we focus on here.

The model proposed here is admittedly simpler in structure than some others existing in the literature: It does not incorporate strategic interaction of traders as in the game-theoretic approaches of Parlour (1998), Foucault et al. (2005), and Rosu (2009), nor does it account for "long memory" features of the order flow as pointed out by Bouchaud et al. (2002, 2008). However, contrarily to these models, it leads to an analytically tractable framework where parameters can be easily estimated from empirical data and various quantities of interest may be computed efficiently.

Outline. The paper is organized as follows. Section 1 describes a stylized model for the dynamics of a limit order book, where the order flow is described by independent Poisson processes. Estimation of model parameters from high-frequency order book time-series data is described in §2 and illustrated using data from the Tokyo Stock Exchange. In §3 we show how this model can be used to compute conditional probabilities of various types of events relevant for trade execution using Laplace transform methods. Section 4 explores steady-state properties of the model using Monte Carlo simulation, compares conditional probabilities computed by simulation to those computed with the Laplace transform methods presented in §3, and analyzes a high-frequency trading strategy based on our results in §4.3. Section 5 concludes.

1. A Continuous-Time Model for a Stylized Limit Order Book

1.1. Limit Order Books

Consider a financial asset traded in an order-driven market. Market participants can post two types of buy/sell orders. A *limit* order is an order to trade a certain amount of a security at a given price. Limit orders are posted to a electronic trading system, and the state of outstanding limit orders can be summarized by stating the quantities posted at each price level: this is known as the *limit order book*. The lowest price for which there is an outstanding limit sell order is called the *ask price* and the highest buy price is called the *bid price*.

A *market* order is an order to buy/sell a certain quantity of the asset at the best available price in the limit order book. When a market order arrives it is matched with the best available price in the limit order book, and a trade occurs. The quantities available in the limit order book are updated accordingly.

A *limit* order sits in the order book until it is either executed against a market order or it is canceled. A limit order may be executed very quickly if it corresponds to a price near the bid and the ask, but may take a long time if the market price moves away from the requested price or if the requested price is too far from the bid/ask. Alternatively, a limit order can be canceled at any time.

We consider a market where limit orders can be placed on a price grid $\{1, \ldots, n\}$ representing multiples of a price tick. The upper boundary n is chosen large enough so that it is highly unlikely that orders for the stock in question are placed at prices higher than n within the time frame of our analysis. Because the model is intended to be used on the time scale of hours or days, this finite boundary assumption is reasonable. We track the state of the order book with a continuous-time process $X(t) \equiv (X_1(t), \ldots, X_n(t))_{t \geqslant 0}$, where $|X_p(t)|$ is the number of outstanding limit orders at price p, $1 \leqslant p \leqslant n$. If $X_p(t) < 0$, then there are $-X_p(t)$ bid orders at price p; if $X_p(t) > 0$, then there are $X_p(t)$ ask orders at price p.

The *ask price* $p_A(t)$ at time t is then defined by

$$p_A(t) = \inf\{p = 1, \ldots, n, X_p(t) > 0\} \wedge (n+1).$$

Similarly, the *bid price* $p_B(t)$ is defined by

$$p_B(t) \equiv \sup\{p = 1, \ldots, n, X_p(t) < 0\} \vee 0.$$

Notice that when there are no ask orders in the book we force an ask price of $n + 1$, and when there are no bid orders in the book we force a bid price of 0. The *midprice* $p_M(t)$ and the *bid-ask spread* $p_S(t)$ are defined by

$$p_M(t) \equiv \frac{p_B(t) + p_A(t)}{2} \quad \text{and} \quad p_S(t) \equiv p_A(t) - p_B(t).$$

Because most of the trading activity takes place in the vicinity of the bid and ask prices, it is useful to keep track

of the number of outstanding orders at a given distance from the bid/ask. To this end, we define

$$Q_i^B(t) = \begin{cases} X_{p_A(t)-i}(t) & 0 < i < p_A(t) \\ 0 & p_A(t) \leqslant i < n, \end{cases} \quad (1)$$

the number of buy orders at a distance i from the ask, and

$$Q_i^A(t) = \begin{cases} X_{p_B(t)+i}(t) & 0 < i < n - p_B(t) \\ 0 & n - p_B(t) \leqslant i < n, \end{cases} \quad (2)$$

the number of sell orders at a distance i from the bid. Although $X(t)$ and $(p_A(t), p_B(t), Q^A(t), Q^B(t))$ contain the same information, the second representation highlights the shape or depth of the book relative to the best quotes.

1.2. Dynamics of the Order Book

Let us now describe how the limit order book is updated by the inflow of new orders. For a state $x \in \mathbb{Z}^n$ and $1 \leqslant p \leqslant n$, define

$$x^{p\pm 1} \equiv x \pm (0, \ldots, 1, \ldots, 0),$$

where the 1 in the vector on the right-hand side is in the pth component. Assuming that all orders are of unit size (in empirical examples we will take this unit to be the average size of limit orders observed for the asset),

• a limit buy order at price level $p < p_A(t)$ increases the quantity at level p: $x \to x^{p-1}$

• a limit sell order at price level $p > p_B(t)$ increases the quantity at level p: $x \to x^{p+1}$

• a market buy order decreases the quantity at the ask price: $x \to x^{p_A(t)-1}$

• a market sell order decreases the quantity at the bid price: $x \to x^{p_B(t)+1}$

• a cancellation of an oustanding limit buy order at price level $p < p_A(t)$ decreases the quantity at level p: $x \to x^{p+1}$

• a cancellation of an oustanding limit sell order at price level $p > p_B(t)$ decreases the quantity at level p: $x \to x^{p-1}$

The evolution of the order book is thus driven by the incoming flow of market orders, limit orders, and cancellations at each price level, each of which can be represented as a counting process. It is empirically observed (Bouchaud et al. 2002) that incoming orders arrive more frequently in the vicinity of the current bid/ask price and the rate of arrival of these orders depends on the distance to the bid/ask.

To capture these empirical features in a model that is analytically tractable and allows computation of quantities of interest in applications, most notably conditional probabilities of various events, we propose a stochastic model where the events outlined above are modelled using independent Poisson processes. More precisely, we assume that, for $i \geqslant 1$,

• Limit buy (respectively sell) orders arrive at a distance of i ticks from the opposite best quote at independent, exponential times with rate $\lambda(i)$,

• Market buy (respectively sell) orders arrive at independent, exponential times with rate μ,

• Cancellations of limit orders at a distance of i ticks from the opposite best quote occur at a rate proportional to the number of outstanding orders: If the number of outstanding orders at that level is x, then the cancellation rate is $\theta(i)x$. This assumption can be understood as follows: if we have a batch of x outstanding orders, each of which can be canceled at an exponential time with parameter $\theta(i)$, then the overall cancellation rate for the batch is $\theta(i)x$.

• The above events are mutually independent.

Order arrival rates depend on the distance to the bid/ask with most orders being placed close to the current price. We model the arrival rate as a function $\lambda: \{1, \ldots, n\} \to [0, \infty)$ of the distance to the bid/ask. Empirical studies (Zovko and Farmer 2002 or Bouchaud et al. 2002) suggest a power law,

$$\lambda(i) = \frac{k}{i^\alpha},$$

as a plausible specification.

Given the above assumptions, X is a continuous-time Markov chain with state space $\mathbb{Z}^n$ and transition rates given by:

$x \to x^{p-1}$ with rate $\lambda(p_A(t) - p)$ for $p < p_A(t)$,

$x \to x^{p+1}$ with rate $\lambda(p - p_B(t))$ for $p > p_B(t)$,

$x \to x^{p_B(t)+1}$ with rate μ,

$x \to x^{p_A(t)-1}$ with rate μ,

$x \to x^{p+1}$ with rate $\theta(p_A(t) - p)|x_p|$ for $p < p_A(t)$,

$x \to x^{p-1}$ with rate $\theta(p - p_B(t))|x_p|$ for $p > p_B(t)$.

In practice, the ask price is always greater than the bid price. We say a state is *admissible* if it fulfills this requirement:

$$\mathscr{A} \equiv \{x \in \mathbb{Z}^n \mid \exists k, l \in \mathbb{Z} \text{ s.t. } 1 \leqslant k \leqslant l \leqslant n, x_p \geqslant 0 \text{ for } p \geqslant l,$$
$$x_p = 0 \text{ for } k \leqslant p \leqslant l, x_p \leqslant 0 \text{ for } p \leqslant k\}. \quad (3)$$

If the initial state of the book is admissible, it remains admissible with probability one:

PROPOSITION 1. *If* $X(0) \in \mathscr{A}$, *then* $\mathbb{P}[X(t) \in \mathscr{A}, \forall t \geqslant 0] = 1$.

PROOF. It is easily verified that $\mathscr{A}$ is stable under each of the six transitions defined above, which leads to our assertion. $\square$

PROPOSITION 2. *If* $\theta \equiv \min_{1 \leqslant i \leqslant n} \theta(i) > 0$, *then* X *is an ergodic Markov process. In particular, X has a proper stationary distribution.*

PROOF. Let $N \equiv (N(t), t \geqslant 0)$, where $N(t) \equiv \sum_{p=1}^{n} |X_p(t)|$, and let $\tilde{N}$ be a birth-death process with birth rate given by $\lambda \equiv 2 \sum_{p=1}^{n} \lambda(p)$ and death rate in state i, $\mu_i \equiv 2\mu + i\theta$. Notice that N increases by one at a rate bounded from above by λ and decreases by one at a rate bounded from below by $\mu_i \equiv 2\mu + i\theta$ when in state i. Thus, for all $t \geqslant 0$, N is stochastically bounded by $\tilde{N}$. For $k \geqslant 1$, let T_0^k and T_{-0}^k denote the duration of the kth visit to 0 and the duration between the $(k-1)$th and kth visit to 0 of process N, respectively. Define random variables $\tilde{T}_0^k$ and $\tilde{T}_{-0}^k$, $k \geqslant 1$, for process $\tilde{N}$ similarly. Then the point process with interarrival times $T_{-0}^1, T_0^1, T_{-0}^2, T_0^2, \ldots$ and the point process with interarrival times $\tilde{T}_{-0}^1, \tilde{T}_0^1, \tilde{T}_{-0}^2, \tilde{T}_0^2, \ldots$ are alternating renewal processes. By Theorem VI.1.2 of Asmussen (2003) and the fact that N is stochastically dominated by $\tilde{N}$, we then have for each $k \geqslant 1$,

$$\frac{\mathbb{E}[T_0^k]}{\mathbb{E}[T_0^k] + \mathbb{E}[T_{-0}^k]} = \lim_{t \to \infty} \mathbb{P}[N(t) = 0]$$

$$\geqslant \lim_{t \to \infty} \mathbb{P}[\tilde{N}(t) = 0] = \frac{\mathbb{E}[\tilde{T}_0^k]}{\mathbb{E}[\tilde{T}_0^k] + \mathbb{E}[\tilde{T}_{-0}^k]}. \quad (4)$$

Notice that in state 0 both N and $\tilde{N}$ have birth rate λ. Thus,

$$\mathbb{E}[T_0^k] = \mathbb{E}[\tilde{T}_0^k] = \frac{1}{\lambda}. \quad (5)$$

Combining (4) and (5) gives us

$$\mathbb{E}[T_{-0}^k] \leqslant \mathbb{E}[\tilde{T}_{-0}^k]. \quad (6)$$

To show $\tilde{N}$ is ergodic, notice the inequalities

$$\sum_{i=1}^{\infty} \frac{\lambda^i}{\mu_1 \cdots \mu_i} < \sum_{i=1}^{\infty} \frac{1}{i!} \left(\frac{\lambda}{\theta} \right)^i = e^{\lambda/\theta} - 1 < \infty, \quad (7)$$

and

$$\sum_{i=1}^{\infty} \frac{\mu_1 \cdots \mu_i}{\lambda^i} > \sum_{i=1}^{M} \frac{\mu_1 \cdots \mu_i}{\lambda^i} + \sum_{i=M+1}^{\infty} \left(\frac{2\mu + M\theta}{\lambda} \right)^i = \infty, \quad (8)$$

for $M > 0$ chosen large enough so that $2\mu + M\theta > \lambda$. Therefore, by Corollary 2.5 of Asmussen (2003), $\tilde{N}$ is ergodic so that $\mathbb{E}[\tilde{T}_{-0}^k] < \infty$. Combining this with the bound (6) and the fact that for each $t \geqslant 0$ $X(t) = (0, \ldots, 0)$ if and only if $N(t) = 0$ shows that X is positive recurrent. Because X is clearly also irreducible, it follows that X is ergodic. $\square$

The ergodicity of X is a desirable feature of theoretical interest: it allows comparison of time averages of various quantities in simulations (average shape of the order book, average price impact, etc.) to unconditional expectations of these quantities computed in the model. The steady-state behavior of X will be further discussed in §4.1. We note, however, that our results involving conditional probabilities in §3 and applications discussed in §4.3 do not rely on this ergodicity result.

2. Parameter Estimation

2.1. Description of the Data Set

Our data consist of time-stamped sequences of trades (market orders) and quotes (prices and quantities of outstanding limit orders) for the five best price levels on each side of the order book, for stocks traded on the Tokyo stock exchange over a period of 125 days (Aug.–Dec. 2006). This data set, referred to as Level II order book data, provides a more detailed view of price dynamics than the trade and quotes (TAQ) data often used for high-frequency data analysis, which consist of prices and sizes of trades (market orders) and time-stamped updates in the price and size of the bid and ask quotes.

In Table 1, we display a sample of three consecutive trades for Sky Perfect Communications. Each row provides the time, size, and price of a market order. We also display a sample of Level II bid-side quotes. Each row displays the five bid prices (pb1, pb2, pb3, pb4, pb5), as well as the quantity of shares bid at these respective prices (qb1, qb2, qb3, qb4, qb5).

2.2. Estimation Procedure

Recall that in our stylized model we assume orders to be of "unit" size. In the data set, we first compute the average sizes of market orders S_m, limit orders S_l, and canceled orders S_c and choose the size unit to be the average size of a limit order S_l. The limit order arrival rate function for $1 \leqslant i \leqslant 5$ can be estimated by

$$\hat{\lambda}(i) = \frac{N_l(i)}{T_*},$$

where $N_l(i)$ is the total number of limit orders that arrived at a distance i from the opposite best quote, and T_* is the total trading time in the sample (in minutes). $N_l(i)$ is obtained by enumerating the number of times that a quote increases in size at a distance of $1 \leqslant i \leqslant 5$ ticks from the opposite best quote. We then extrapolate by fitting a power law function of the form

$$\hat{\lambda}(i) = \frac{k}{i^\alpha}$$

(suggested by Zovko and Farmer 2002 or Bouchaud et al. 2002). The power law parameters k and α are obtained by a least-squares fit

$$\min_{k, \alpha} \sum_{i=1}^{5} \left(\hat{\lambda}(i) - \frac{k}{i^\alpha} \right)^2.$$

Estimated arrival rates at distances $0 \leqslant i \leqslant 10$ from the opposite best quote are displayed in Figure 1(a).

The arrival rate of market orders is then estimated by

$$\hat{\mu} = \frac{N_m}{T_*} \frac{S_m}{S_l},$$

Table 1. A sample of three trades and five quotes for Sky Perfect Communications.

Time	Price	Size
9:11:01	74,300	1
9:11:04	74,600	2
9:11:19	74,400	1

Time	pb1	pb2	pb3	pb4	pb5	qb1	qb2	qb3	qb4	qb5
9:11:01	74,300	74,200	74,000	73,900	73,800	12	13	1	52	11
9:11:03	74,400	74,300	74,200	74,000	73,900	20	12	13	1	52
9:11:04	74,400	74,300	74,200	74,000	73,900	21	11	13	1	52
9:11:05	74,400	74,300	74,200	74,000	73,900	34	4	13	1	52
9:11:19	74,400	74,300	74,200	74,000	73,900	33	4	13	1	52

where T_* is the total trading time in the sample (in minutes) and N_m is the number of market orders. Note that we ignore market orders that do not affect the best quotes, as is the case when a market order is matched by a hidden order.

Because the cancellation rate in our model is proportional to the number of orders at a particular price level,

Figure 1. The arrival rates as a function of the distance from the opposite quote.
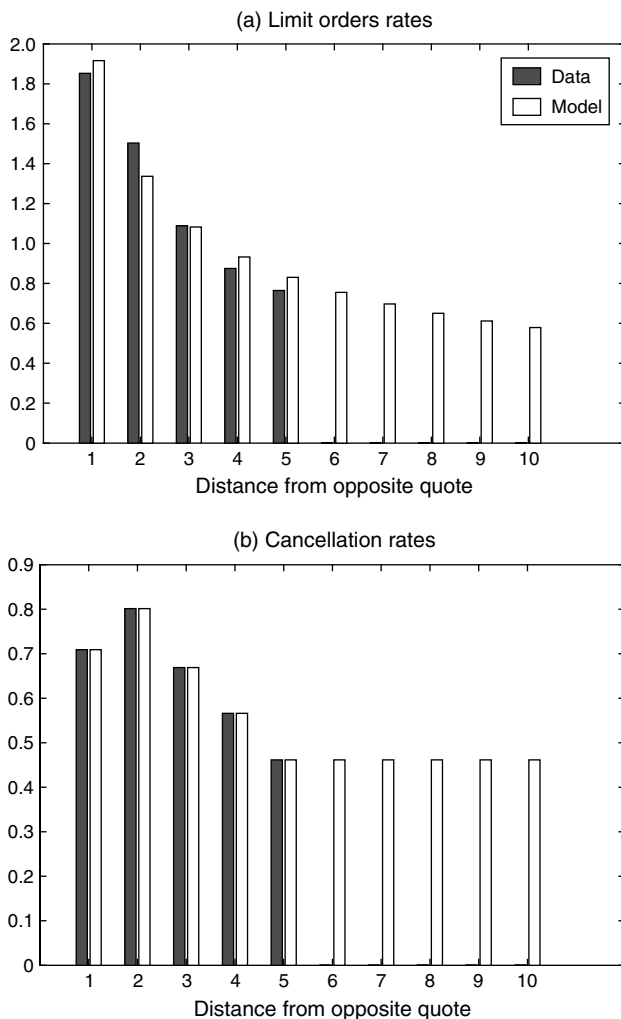
in order to estimate the cancellation rates we first need to estimate the steady-state shape of the order book Q_i, which is the average number of orders at a distance of i ticks from the opposite best quote, for $1 \leqslant i \leqslant 5$. If M is the number of quote rows and $S_i^B(j)$ the number of shares bid at a distance of i ticks from the ask on the jth row, for $1 \leqslant j \leqslant M$, we have

$$Q_i^B = \frac{1}{S_l} \frac{1}{M} \sum_{j=1}^{M} S_i^B(j).$$

The vector Q_i^A is obtained analogously, and Q_i is the average of Q_i^A and Q_i^B.

An estimator for the cancellation rate function is then given by

$$\hat{\theta}(i) = \frac{N_c(i)}{TQ_i} \frac{S_c}{S_l} \quad \text{for } i \leqslant 5 \quad \text{and}$$

$$\hat{\theta}(i) = \hat{\theta}(5) \quad \text{for } i > 5, \tag{9}$$

where $N_c(i)$ is obtained by counting the number of times that a quote decreases in size at a distance of $1 \leqslant i \leqslant 5$ ticks from the opposite best quote, excluding decreases due to market orders. The fitted values are displayed in Figure 1(b).

Estimated parameter values for Sky Perfect Communications are given in Table 2.

3. Laplace Transform Methods for Computing Conditional Probabilities

As noted above, an important motivation for modelling high-frequency dynamics of order books is to use the information provided by the limit order book for predicting

Table 2. Estimated parameters: Sky Perfect Communications.

			i		
	1	2	3	4	5
$\hat{\lambda}(i)$	1.85	1.51	1.09	0.88	0.77
$\hat{\theta}(i)$	0.71	0.81	0.68	0.56	0.47
$\hat{\mu}$	0.94				
k	1.92				
α	0.52				

short-term behavior of various quantities that are useful in trade execution and algorithmic trading, for instance, the probability of the midprice moving up versus down, the probability of executing a limit order at the bid before the ask quote moves, and the probability of executing both a buy and a sell order at the best quotes before the price moves. These quantities can be expressed in terms of *conditional probabilities* of events, given the state of the order book. In this section we show that the model proposed in §1 allows such conditional probabilities to be analytically computed using Laplace methods. After presenting some background on Laplace transforms in §3.1, we give various examples of these computations. The probability of an increase in the midprice is discussed in §3.2, the probability that a limit order executes before the price moves is discussed in §3.3, and the probability of executing both a buy and a sell limit order before the price moves is discussed in §3.4. Laplace transform methods allow efficient computation of these quantities, bypassing the need for Monte Carlo simulation.

3.1. Laplace Transforms and First-Passage Times of Birth-Death Processes

We first recall some basic facts about two-sided Laplace transforms and discuss the computation of Laplace transforms for first-passage times of birth-death processes (Abate and Whitt 1999). Given a function $f: \mathbb{R} \to \mathbb{R}$, its two-sided Laplace transform is given by

$$\hat{f}(s) = \int_{-\infty}^{\infty} e^{-st} f(t) \, dt,$$

where s is a complex numbers. When f is the probability density function (pdf) of some random variable X, we also say that $\hat{f}$ is the two-sided Laplace transform of the random variable X. We work with *two-sided* Laplace transforms here because for our purposes the function f will usually correspond to the pdf of a random variable with both positive and negative support. From now on, we drop the prefix "two-sided" when referring to two-sided Laplace transforms. When we say *conditional* Laplace transform of the random variable X conditional on the event A, we mean the Laplace transform of the conditional pdf of X given A. Recall that if X and Y are independent random variables with well-defined Laplace transforms, then

$$\hat{f}_{X+Y}(s) = \mathbb{E}[e^{-s(X+Y)}] = \mathbb{E}[e^{-sX}]\mathbb{E}[e^{-sY}] = \hat{f}_X(s)\hat{f}_Y(s). \quad (10)$$

If for some $\gamma \in \mathbb{R}$ we have $\int_{-\infty}^{\infty} |\hat{f}(\gamma + i\omega)| \, d\omega < \infty$ and $f(t)$ is continuous at t, then the inverse transform is given by the Bromwich contour integral

$$f(t) = \frac{1}{2\pi i} \int_{\gamma - i\infty}^{\gamma + i\infty} e^{ts} \hat{f}(s) \, ds. \quad (11)$$

The continued fraction associated with a sequence $\{a_n, n \geq 1\}$ of partial numerators and $\{b_n, n \geq 1\}$ of partial

denominators, which are complex numbers with $a_n \neq 0$ for all $n \geq 1$, is the sequence $\{w_n, n \geq 1\}$, where

$$w_n = t_1 \circ t_2 \circ \cdots \circ t_n(0), \quad n \geq 1, \qquad t_k(u) = \frac{a_k}{b_k + u}, \quad k \geq 1,$$

and $\circ$ denotes the composition operator. If $w \equiv \lim_{n \to \infty} w_n$, then the continued fraction is said to be convergent and the limit w is said to be the value of the continued fraction (Abate and Whitt 1999). In this case, we write

$$w \equiv \Phi_{n=1}^{\infty} \frac{a_n}{b_n}.$$

Consider now a birth-death process with constant birth rate λ and death rates μ_i in state $i \geq 1$, and let σ_b denote the first-passage time of this process to 0 given that it begins in state b. Next, notice that we can write σ_B as the sum

$$\sigma_b = \sigma_{b,b-1} + \sigma_{b-1,b-2} + \cdots + \sigma_{1,0},$$

where $\sigma_{i,i-1}$ denotes the first-passage time of the birth-death process from the state i to the state $i - 1$, for $i = 1, \ldots, b$, and all terms on the right-hand side are independent. If $\hat{f}_b$ denotes the Laplace transform of σ_b and $\hat{f}_{i,i-1}$ denotes the Laplace transform of $\sigma_{i,i-1}$ for $i = 1, \ldots, b$, then we have by (10),

$$\hat{f}_b(s) = \prod_{i=1}^{b} \hat{f}_{i,i-1}(s). \quad (12)$$

Therefore, in order to compute $\hat{f}_b$, it suffices to compute the simpler Laplace transforms $\hat{f}_{i,i-1}$, for $i = 1, \ldots, b$. By Equation (4.9) of Abate and Whitt (1999), we see that the Laplace transform of $\hat{f}_{i,i-1}$ is given by

$$\hat{f}_{i,i-1}(s) = -\frac{1}{\lambda} \Phi_{k=i}^{\infty} \frac{-\lambda\mu_k}{\lambda + \mu_k + s}. \quad (13)$$

The computation there is based on a recursive relationship between the $\hat{f}_{i,i-1}$, $i = 1, \ldots, b$, which is derived by considering the first transition of the birth-death process. Combining (12) and (13), we obtain

$$\hat{f}_b(s) = \left(-\frac{1}{\lambda}\right)^b \left(\prod_{i=1}^{b} \Phi_{k=i}^{\infty} \frac{-\lambda\mu_k}{\lambda + \mu_k + s}\right). \quad (14)$$

We will use this result in all our computations below.

3.2. Direction of Price Moves

We now compute the probability that the midprice increases at its next move. The first move in the midprice occurs at the first-passage time of the bid or ask queue to zero or, if the bid/ask spread is greater than one, the first time a limit order arrives inside the spread. Throughout this section, let $X_A \equiv X_{p_A(\cdot)}(\cdot)$ and $X_B \equiv |X_{p_B(\cdot)}(\cdot)|$. Furthermore, let $W_B \equiv \{W_B(t), t \geq 0\}$ ($W_A \equiv \{W_A(t), t \geq 0\}$), where $W_B(t)$ ($W_A(t)$) denotes the number of orders remaining at the bid

(ask) at time t of the initial $X_B(0)$ ($X_A(0)$) orders and let ϵ_B (ϵ_A) be the first-passage time of W_B (W_A) to 0. Furthermore, let T be the time of the first change in midprice:

$$T \equiv \inf\{t \geqslant 0, \, p_M(t) \neq p_M(0)\}.$$

Given an initial configuration of the book, the probability that the next change in midprice is an increase can then be written as

$$\mathbb{P}[p_M(T) > p_M(0) \,|\, X_A(0) = a, X_B(0) = b, p_S(0) = S], \quad (15)$$

where $S > 0$. For ease of notation, we will omit the condition in (15) in all proofs below.

The idea for computing (15) is to use a coupling argument.

LEMMA 3. *Let $p_S(0) = S$. Then*
1. *There exist independent birth-death processes $\tilde{X}_A$ and $\tilde{X}_B$ with constant birth rates $\lambda(S)$ and death rates $\mu + i\theta(S)$, $i \geqslant 1$, such that for all $0 \leqslant t \leqslant T$, $\tilde{X}_A(t) = X_A(t)$, and $\tilde{X}_B(t) = X_B(t)$.*
2. *There exist independent pure death processes $\widetilde{W}_A$ and $\widetilde{W}_B$ with death rate $\mu + i\theta(S)$ in state $i \geqslant 1$, such that for all $0 \leqslant t \leqslant T$, $\widetilde{W}_A(t) = W_A(t)$ and $\widetilde{W}_B(t) = W_B(t)$. Furthermore, $\widetilde{W}_A$ is independent of $\tilde{X}_B$, $\widetilde{W}_B$ is independent of $\tilde{X}_A$, $\widetilde{W}_A \leqslant \tilde{X}_A$, and $\widetilde{W}_B \leqslant \tilde{X}_B$.*

PROOF. We prove Part 1. Part 2 can be proven analogously. X is a continuous-time Markov chain, with transition rates given by (1.2). For $0 \leqslant t \leqslant T$, $p_A(t) = p_A(0)$ and $p_B(t) = p_B(0)$, so substituting in (1.2) yields that $X_A(t)$ and $X_B(t)$ have the following (identical) transition rates for $0 \leqslant t \leqslant T$

$$n \to n + 1 \quad \text{with rate } \lambda(S) \quad (16)$$

$$n \to n - 1 \quad \text{with rate } \mu + n\theta(S). \quad (17)$$

Define $\tilde{X}_A$ and $\tilde{X}_B$ such that
- $\tilde{X}_A(t) = X_A(t)$ and $\tilde{X}_B(t) = X_B(t)$ for $t \leqslant T$ and
- $\tilde{X}_A(t), \tilde{X}_B(t), t \geqslant T$ follow independent birth-death processes with rates given by (16) and (17).

The above remarks show that in fact $(\tilde{X}_A(t))_{t \geqslant 0}$ (respectively $(\tilde{X}_B(t))_{t \geqslant 0}$) has the same law as a birth-death process with rates (16)–(17). To show that $\tilde{X}_A$ and $\tilde{X}_B$ are independent, we note that because the transition rates of X_A (respectively X_B) do not depend on $(X_p(t), p \neq p_A(0))$ (respectively $(X_p(t), p \neq p_B(0))$) for $0 \leqslant t \leqslant T$, we have, in particular, conditional independence of $X_A(t)$ and $X_B(t)$ given $X(0)$ and $\{t \leqslant T\}$. $\square$

Henceforth, we let σ_A and σ_B denote the first-passage times of $\tilde{X}_A$ and $\tilde{X}_B$ to 0, respectively. The conditional probability (15) can then be computed as follows:

PROPOSITION 4 (PROBABILITY OF INCREASE IN MIDPRICE). *Let $\hat{f}_j^S$ be given by*

$$\hat{f}_j^S(s) = \left(-\frac{1}{\lambda(S)}\right)^j \left(\prod_{i=1}^{b} \Phi_{k=i}^{\infty} \frac{-\lambda(S)(\mu + k\theta(S))}{\lambda(S) + \mu + k\theta(S) + s}\right), \quad (18)$$

for $j \geqslant 1$, and let $\Lambda_S \equiv \sum_{i=1}^{S-1} \lambda(i)$. Then (15) *is given by the inverse Laplace transform of*

$$\hat{F}_{a,b}^S(s) = \frac{1}{s}\left(\hat{f}_a^S(\Lambda_S + s) + \frac{\Lambda_S}{\Lambda_S + s}(1 - \hat{f}_a^S(\Lambda_S + s))\right)$$
$$\cdot\left(\hat{f}_b^S(\Lambda_S - s) + \frac{\Lambda_S}{\Lambda_S - s}(1 - \hat{f}_b^S(\Lambda_S - s))\right), \quad (19)$$

evaluated at 0. When $S = 1$, (19) *reduces to*

$$\hat{F}_{a,b}^1(s) = \frac{1}{s}\hat{f}_a^1(s)\hat{f}_b^1(-s). \quad (20)$$

PROOF. We will first focus on the special case when $S = 1$ and then extend the analysis to the case $S > 1$, using Lemma 5 below. Construct the independent birth-death processes $\tilde{X}_A$ and $\tilde{X}_B$ as in Lemma 3. When $S = 1$, the price changes for the first time exactly when one of the two processes $\tilde{X}_A$ and $\tilde{X}_B$ reaches the state 0 for the first time. Thus, given our initial conditions, the distribution of T is given by the minimum of the independent first-passage times σ_A and σ_B. Furthermore, the quantity (15) is given by $\mathbb{P}[\sigma_A < \sigma_B]$. By (14), the conditional Laplace transform of $\sigma_A - \sigma_B$ given the initial conditions is given by $\hat{f}_a^1(s)\hat{f}_b^1(-s)$ so that the conditional Laplace transform of the cumulative distribution function (cdf) of $\sigma_A - \sigma_B$ is given by (20). Thus, our desired probability is given by the inverse Laplace transform of (20) evaluated at 0.

We now move on to the case where $S > 1$. Let σ_A^i denote the first time an ask order arrives i ticks away from the bid and σ_B^i denote the first time a bid order arrives i ticks away from the ask, for $i = 1, \ldots, S - 1$. The time of the first change in midprice is now given by

$$T = \sigma_A \wedge \sigma_B \wedge \min\{\sigma_A^i, \sigma_B^i, i = 1, \ldots, S - 1\}.$$

Notice that $\tilde{X}_A$ and $\tilde{X}_B$ are independent of the mutually independent arrival times σ_A^i, σ_B^i, for $i = 1, \ldots, S-1$. Also, notice that σ_A^i and σ_B^i are exponentially distributed with rates $\lambda(i)$ for $i = 1, \ldots, S-1$. The first change in midprice is an increase if there is an arrival of a limit bid order within $S - 1$ ticks of the best ask or $\tilde{X}_A$ hits zero, before there is an arrival of a limit ask order within $S - 1$ ticks of the best bid or $\tilde{X}_B$ hits zero. Thus, the quantity (15) can be written as

$$\mathbb{P}[\sigma_A \wedge \sigma_B^1 \wedge \cdots \wedge \sigma_B^{S-1} < \sigma_B \wedge \sigma_A^1 \wedge \cdots \wedge \sigma_A^{S-1}]$$
$$= \mathbb{P}[\sigma_A \wedge \sigma_B^\Sigma < \sigma_B \wedge \sigma_A^\Sigma], \quad (21)$$

where σ_A^Σ and σ_B^Σ are independent exponential random variables, both with rate Λ_S. To compute (21), we first need to compute the conditional Laplace transform of the minimum $\sigma_B \wedge \sigma_A^\Sigma$. This is given in Lemma 5, substituting σ_A^Σ for Z. The conditional Laplace transform of the random variable $\sigma_B \wedge \sigma_A^\Sigma - \sigma_A \wedge \sigma_B^\Sigma$ can then be computed using (10), and the probability (15) can be computed by inverting the conditional Laplace transform of the cdf of this random variable and evaluating at 0 as in the case $S = 1$. $\square$

LEMMA 5. *Let Z be an exponentially distributed random variable with parameter Λ. Then the Laplace transform of the random variable $\sigma_B \wedge Z$ is given by*

$$\hat{f}_b^1(\Lambda + s) + \frac{\Lambda}{\Lambda + s}(1 - \hat{f}_b^1(\Lambda + s)),$$

where $\hat{f}_b^1$ is given in (18).

PROOF. We first compute the density $f_{\sigma_B \wedge Z}$ of the random variable $\sigma_B \wedge Z$ in terms of the density f_b of the random variable σ_B. Because Z is exponential with rate Λ, we have for all $t \geq 0$,

$$\mathbb{P}[\sigma_B \wedge Z < t] = 1 - \mathbb{P}[\sigma_B > t]\mathbb{P}[Z > t]$$
$$= 1 - (1 - F_{\sigma_B}(t))e^{-\Lambda t}.$$

Taking derivatives with respect to t gives

$$f_{\sigma_B \wedge Z}(t) = f_b^1(t)e^{-\Lambda t} + \Lambda(1 - F_b^1(t))e^{-\Lambda t}, \qquad (22)$$

for $t \geq 0$, where $F_b^1(t)$ $(f_b^1(t))$ is the cdf (pdf) of σ_B. Also, $f_{\sigma_B \wedge Z}(t) = 0$ for $t < 0$. The Laplace transform of $\sigma_B \wedge Z$ is thus given by

$$\hat{f}_{\sigma_B \wedge Z}(s) = \int_{-\infty}^{\infty} e^{-st} f_{\sigma_B \wedge \sigma_B^\Sigma}(t)\, dt$$
$$= \int_0^{\infty} e^{-st}(f_b^1(t)e^{-\Lambda t} + \Lambda(1 - F_b^1(t))e^{-\Lambda t})\, ds$$
$$= \int_0^{\infty} e^{-t(s+\Lambda)} f_b^1(t)\, dt + \Lambda \int_0^{\infty}(1 - F_b^1(t))e^{-t(s+\Lambda)}\, dt$$
$$= \hat{f}_b^1(s+\Lambda) + \frac{\Lambda}{\Lambda + s}(1 - \hat{f}_b^1(s+\Lambda)),$$

where the last equality follows from integration by parts. □

Proposition 4 yields a numerical procedure for computing the probability that the next change in the midprice will be an increase. We discuss implementation of the procedure in §4.2.2.

3.3. Executing an Order Before the Mid-Price Moves

A trader that submits a limit order at a given time obtains a better price than a trader that submits a market order at that same time, but faces the risk of nonexecution and the "winner's curse." Whereas a market order executes with certainty, a limit order stays in the order book until either a matching order is entered or the order is canceled. The probability that a limit order is executed before the price moves is therefore useful in quantifying the choice between placing a limit order and placing a market order. We now compute the probability that an order placed at the bid price is executed before any movement in the midprice, given that the order is not canceled. Our result holds for initial spread $S \equiv p_S(0) \geq 1$, but we remark that in the case where

$S = 1$ the probability we are interested in is equal to the probability that the order is executed before the midprice moves *away* from the desired price, given that the order is not canceled. Although we focus here on an order placed at the bid price, because our model is symmetric in bids and asks, our result also holds for orders placed at the ask price.

We introduce some new notation that we will use in this subsection as well as the next. Let NC_b (NC_a) denote the event that an order that never gets canceled is placed at the bid (ask) at time 0. Then, the probability that an order placed at the bid is executed before the midprice moves is given by

$$\mathbb{P}[\epsilon_B < T \mid X_B(0) = b, X_A(0) = a, p_S(0) = S, NC_b]. \quad (23)$$

PROPOSITION 6 (PROBABILITY OF ORDER EXECUTION BEFORE MIDPRICE MOVES). *Define $\hat{f}_a^S(s)$ as in* (18), *let $\hat{g}_j^S$ be given by*

$$\hat{g}_j^S(s) = \prod_{i=1}^{j} \frac{\mu + \theta(S)(i-1)}{\mu + \theta(S)(i-1) + s}, \qquad (24)$$

for $j \geq 1$, and let $\Lambda_S \equiv \sum_{i=1}^{S-1} \lambda(i)$. Then the quantity (23) *is given by the inverse Laplace transform of*

$$\hat{F}_{a,b}^S(s) = \frac{1}{s}\hat{g}_b^S(s)\bigg(\hat{f}_b^S(2\Lambda_S - s)$$
$$+ \frac{2\Lambda_S}{2\Lambda_S - s}(1 - \hat{f}_b^S(2\Lambda_S - s))\bigg), \quad (25)$$

evaluated at 0. *When $S = 1$,* (25) *reduces to*

$$\hat{F}_{a,b}^1(s) = \frac{1}{s}\hat{g}_b^1(s)\hat{f}_a^1(-s). \qquad (26)$$

PROOF. Construct $\widetilde{X}_A$ and $\widetilde{W}_B$ using Lemma 3. Let us first consider the case $S = 1$. Let $T' \equiv \epsilon_B \wedge T$ denote the first time when either the process $\widetilde{W}_B$ hits 0 or the midprice changes. Conditional on an infinitely patient order being placed at the bid price at time 0, T' is the first time when either that order gets executed or the midprice changes. Notice that conditional on our initial conditions, ϵ_B is given by a sum of b independent exponentially distributed random variables with parameters $\mu + (i-1)\theta(1)$, for $i = 1, \ldots, b$, and independent of $\widetilde{X}_A$. Thus, the conditional Laplace transform of ϵ_B given our initial conditions is given by (24). Because in the case $S = 1$ the midprice can change before time ϵ_B if and only if $\sigma_A < \epsilon_B$, the quantity (23) can be written simply as $\mathbb{P}[\epsilon_B < \sigma_A]$. Using (10) with the conditional Laplace transforms of ϵ_B and σ_A, given in (24) and (18), respectively, we obtain (26).

This analysis can be extended to the case where $S > 1$ just as in the proof of Proposition 4. When $S > 1$, our desired quantity can be written as $\mathbb{P}[\epsilon_B < \sigma_A \wedge \sigma_B^\Sigma \wedge \sigma_A^\Sigma]$. Because the conditional distribution of $\sigma_B^\Sigma \wedge \sigma_A^\Sigma$ is exponential with parameter $2\Lambda_S$, Lemma 5 then yields the result. □

3.4. Making the Spread

We now compute the probability that two orders, one placed at the bid price and one placed at the ask price, are both executed before the midprice moves, given that the orders are not canceled. If the probability of executing both a buy and a sell limit order before the price moves is high, a statistical arbitrage strategy can be designed by submitting limit orders at the bid and the ask and wait for both orders to execute. If both orders execute before the price moves, the strategy has paid off the bid-ask spread: we refer to this situation as "making the spread." Otherwise, losses may be minimized by submitting a market order and losing the bid-ask spread. We restrict attention to the case where the initial spread is one tick: $S = 1$. The probability of making the spread can be expressed as

$$\mathbb{P}[\max\{\epsilon_A, \epsilon_B\} < T \mid X_B(0) = b, X_A(0) = a,$$

$$p_S(0) = 1, NC_a, NC_b]. \tag{27}$$

The following result allows one to compute this probability using Laplace transform methods:

PROPOSITION 7. *The probability* (27) *of making the spread is given by* $h_{a,b} + h_{b,a}$, *where*

$$h_{a,b} = \sum_{i=0}^{\infty} \sum_{j=1}^{a} \mathbb{P}[\epsilon_j < \sigma_i] \int_0^{\infty} P_{0,i}^X(t) P_{a,j}^W(t) g_b^1(t) \, dt, \tag{28}$$

where

$$P_{0,i}^X(t) \equiv \frac{e^{-\lambda^X(t)} \lambda^X(t)^i}{i!}, \quad \lambda^X(t) \equiv \frac{\lambda}{\theta}(1 - e^{-\theta t}), \tag{29}$$

$$P_{a,j}^W(t) \equiv (e^{Q_a^W t})_{a,j} \equiv \left(\sum_{k=0}^{\infty} \frac{t^k}{k!}(Q_a^W)^k\right)_{a,j}, \tag{30}$$

$$Q_a^W \equiv \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ \mu & -\mu & 0 & \cdots & 0 \\ 0 & \mu+\theta & -\mu-\theta & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \mu+(a-1)\theta & -\mu-(a-1)\theta \end{bmatrix}, \tag{31}$$

and g_b^1 *is the inverse Laplace transform of* $\hat{g}_b^1$, *which is given in* (24).

PROOF. Because $S = 1$, $T = \min\{\sigma_A, \sigma_B\}$, and the quantity (27) can be written as

$$\mathbb{P}[\max\{\epsilon_B, \epsilon_A\} < \min\{\sigma_B, \sigma_A\}]. \tag{32}$$

Construct $\widetilde{X}_A$, $\widetilde{X}_B$, $\widetilde{W}_A$, and $\widetilde{W}_B$ using Lemma 3. Let $T' = \max\{\epsilon_A, \epsilon_B\} \wedge T$ denote the first time when either both of the processes $\widetilde{W}_A$ and $\widetilde{W}_B$ have hit 0, or the midprice has changed. Conditional on infinitely patient orders being

placed at the best bid and ask prices at time 0, T' is the first time when either both the orders get executed or the midprice changes. Furthermore, by Lemma 3, $\widetilde{W}_A$ and $\widetilde{W}_B$ are independent pure death processes with death rate $\mu + i\theta(1)$ in state $i \geqslant 1$, and $\widetilde{W}_A(t) \leqslant \widetilde{X}_A(t)$ and $\widetilde{W}_B(t) \leqslant \widetilde{X}_B(t)$. This implies that ϵ_A and ϵ_B are independent of each other and σ_A and σ_B are independent of each other with $\epsilon_A \leqslant \sigma_A$ and $\epsilon_B \leqslant \sigma_B$. Using these properties, we obtain

$$\mathbb{P}[\max\{\epsilon_B, \epsilon_A\} < \min\{\sigma_B, \sigma_A\}]$$

$$= \mathbb{P}[\epsilon_B < \sigma_A, \epsilon_A < \sigma_B]$$

$$= \mathbb{P}[\epsilon_B < \sigma_A, \epsilon_A < \sigma_B, \epsilon_B < \epsilon_A]$$

$$\quad + \mathbb{P}[\epsilon_B < \sigma_A, \epsilon_A < \sigma_B, \epsilon_A < \epsilon_B]$$

$$= \mathbb{P}[\epsilon_A < \sigma_B, \epsilon_B < \epsilon_A] + \mathbb{P}[\epsilon_B < \sigma_A, \epsilon_A < \epsilon_B]$$

$$= h_{a,b} + h_{b,a}, \tag{33}$$

where we define $h_{a,b} \equiv \mathbb{P}[\epsilon_B < \epsilon_A < \sigma_B]$, the probability that the order placed at the bid is executed before the order placed at the ask, and the order at the ask is executed before the bid quote disappears. We now focus on computing $h_{a,b}$. Conditioning on the value of ϵ_B gives

$$h_{a,b} = \int_0^{\infty} \mathbb{P}[\epsilon_B < \epsilon_A < \sigma_B \mid \epsilon_B = t] g_b^1(t) \, dt. \tag{34}$$

Focusing on the first factor in the integrand in (34) and conditioning on the values of $\widetilde{X}_B(t)$ and $\widetilde{W}_A(t)$ gives us

$$\mathbb{P}[\epsilon_B < \epsilon_A < \sigma_B \mid \epsilon_B = t]$$

$$= \sum_{i=0}^{\infty} \sum_{j=0}^{a} \mathbb{P}[\epsilon_B < \epsilon_A < \sigma_B \mid \epsilon_B = t, \widetilde{X}_B(t) = i, \widetilde{W}_A(t) = j]$$

$$\cdot \mathbb{P}[\widetilde{X}_B(t) = i, \widetilde{W}_A(t) = j \mid \epsilon_B = t]. \tag{35}$$

The first conditional probability on the right hand of (35) can now be simplified as follows. For $i = 0$ or $j = 0$ it is simply 0. For $i, j \geqslant 1$, under the condition of the probability, at time t there are j orders in the ask queue that have been placed before time 0 that have yet to be executed, and there are a total of i orders in the bid queue. Thus, the probability of interest is simply the probability that the j ask orders get executed before the number of orders in the bid queue hits 0. Thus,

$$\mathbb{P}[\epsilon_B < \epsilon_A < \sigma_B \mid \epsilon_B = t, \widetilde{X}_B(t) = i, \widetilde{W}_A(t) = j]$$

$$= \mathbb{P}[\epsilon_j < \sigma_i]. \tag{36}$$

Furthermore, the second probability on the right-hand side of (35) can be written as

$$\mathbb{P}[\widetilde{X}_B(t) = i, \widetilde{W}_A(t) = j \mid \epsilon_B = t]$$

$$= \mathbb{P}[\widetilde{X}_B(t) = i \mid \epsilon_B = t] \mathbb{P}[\widetilde{W}_A(t) = j \mid \epsilon_B = t]$$

$$= \mathbb{P}[\widetilde{X}_B(t) = i \mid \epsilon_B = t] \mathbb{P}[\widetilde{W}_A(t) = j]. \tag{37}$$

Combining Equations (33)–(35) and using Tonelli's theorem to interchange the integral and the summation gives us

$$h_{a,b} = \sum_{i=0}^{\infty} \sum_{j=1}^{a} \mathbb{P}[\epsilon_j < \sigma_i] \int_0^{\infty} \mathbb{P}[\tilde{X}_B(t) = i \mid \epsilon_B = t]$$

$$\cdot \mathbb{P}[\tilde{W}_A(t) = j] g_b^1(t) \, dt.$$

The quantity $\mathbb{P}[\tilde{X}_B(t) = i \mid \epsilon_B = t]$ can be computed using an analogy with the $M/M/\infty$ queue. The number of orders in the bid queue at the time when the bid order placed at time 0 has executed is simply the number of customers at time t in an initially empty $M/M/\infty$ queue with arrival rate λ and service rate θ, which has a Poisson distribution with mean given by $\lambda^X(t)$ in (29). This leads to the expression for $P_{0,i}^X(t)$ in (29).

The quantity $\mathbb{P}[\tilde{W}_A(t) = j]$ is the probability that a pure death process with death rate $\mu + (k-1)\theta(1)$ in state $k \geqslant 1$ is in state j at time t, given that it begins in state a. The infinitesimal generator of this pure death process is given by (31). Thus, by Corollary II.3.5 of Asmussen (2003), $\mathbb{P}[\tilde{W}_A(t) = j]$ is given by (30). □

REMARK 1. We note here that the probabilities computed in this section can also be computed using transition matrices of appropriately defined transient discrete-time Markov chains. In general, for a continuous-time Markov chain the probability of hitting state i before state j can be determined by constructing a corresponding embedded discrete-time Markov chain with states i and j absorbing states and computing the fundamental matrices of this Markov chain (see, for example, §4.4 of Ross 1996). However, our Laplace transform approach has the advantage of computing full distributions of random variables such as σ_A, σ_B, ϵ_A, and ϵ_B. This could be used, for example, to compute probabilities such as $\mathbb{P}[\sigma_A + \delta < \sigma_B]$, for $\delta > 0$, which are useful when latency in order processing is an issue.

4. Numerical Results

Our stochastic model allows one to compute various quantities of interest both by simulating the evolution of the order book and by using the Laplace transform methods presented in §3, based on parameters μ, λ, and θ estimated from the order flow. In this section we compute these quantities—for example, of Sky Perfect Communications—and compare them to empirically observed values in order to assess the precision of the description provided by our model.

In §4.1, we compare empirically observed long-term behavior (e.g., unconditional properties) of the order book to simulations of the fitted model. Although these quantities may not be particularly important for traders who are interested in trading in a short time scale, they indicate how well the model reproduces the average properties of the order book. In §4.2, we compare conditional probabilities of various events in our model to frequencies of the events in the data. We also compare results using the Laplace transform methods developed in §3 to our simulation results.

4.1. Long-Term Behavior

Recent empirical studies on order books (Bouchaud et al. 2002, 2008) have focused mainly on *average* properties of the order book, which in our context correspond to unconditional expectations of quantities under the stationary measure of X: the steady-state shape of the book and the volatility of the midprice. The ergodicity of the Markov chain X, shown in Proposition 2, implies that such expectations $E[f(X_\infty)]$ can be computed in the model by simulating the order book over a large horizon T and averaging $f(X(t))$ over the simulated path:

$$\frac{1}{T} \int_0^T f(X(t)) \, dt \to E[f(X_\infty)] \quad \text{a.s.} \quad \text{as } T \to \infty.$$

4.1.1. Steady-State Shape of the Book. We simulate the order book over a long horizon ($n = 10^6$ events) and observe the mean number of orders Q_i at distances $1 \leqslant i \leqslant 30$ ticks from the opposite best quote. The results are displayed in Figure 2. The steady-state profile of the order book describes the average market impact of trades (Farmer et al. 2004, Bouchaud et al. 2008). Figure 2 shows that the average profile of the order book displays a hump (in this case, at two ticks from the bid/ask), as observed in empirical studies (Bouchaud et al. 2008). Note that this hump feature does not result from any fine-tuning of model parameters or additional ingredients such as correlation between order flow and past price moves.

4.1.2. Volatility. Define the realized volatility of the asset over a day by

$$RV_n = \sqrt{\sum_{i=1}^n \left(\log\left(\frac{P_{i+1}}{P_i} \right) \right)^2}, \tag{38}$$

Figure 2. Simulation of the steady-state profile of the order book: Sky Perfect Communications.
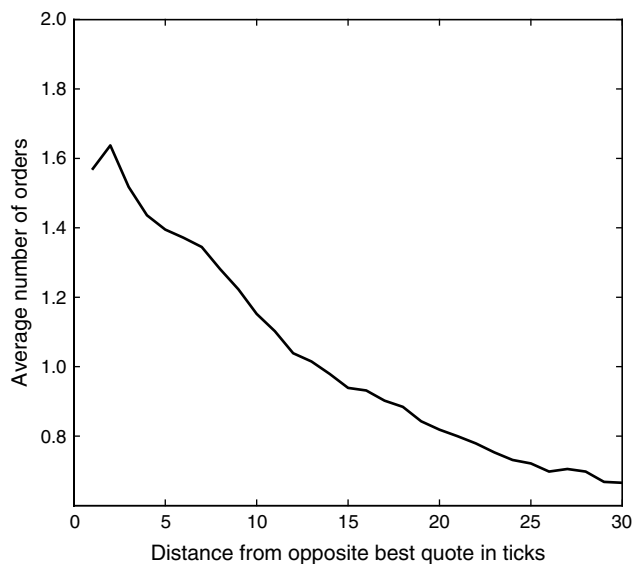
where n is the number of quotes in the day and the prices P_i represent the midprice of the stock, for $i = 1, \ldots, n$. In the first day of the sample, we compute a realized volatility of 0.0219 after a total of 370 trades. After repeatedly simulating our model for 370 trades, we obtained a 95% confidence interval for realized volatility of 0.0228 ± 0.0003. Interestingly, this estimator yields the correct order of magnitude for realized volatility based solely on intensity parameters for the order flow (λ, μ, θ).

4.2. Conditional Distributions

As discussed in the introduction, conditional distributions are the main quantities of interest for applications in high-frequency trading. A good description of conditional distributions of variables characterizing the order book gives one the ability to predict their behavior in the short term, which is of obvious interest in optimal trade execution and the design of trading strategies.

4.2.1. One-Step Transition Probabilities. In order to assess the model's usefulness for short-term prediction of order book behavior, we compare one-step transition probabilities implied by our model to corresponding empirical frequencies. In particular, we consider the probability that the number of orders at a given price level increases given that it changes.

Define T_m as the time of the mth event in the order book:

$$T_0 = 0, \quad T_{m+1} \equiv \inf\{t \geqslant T_m \mid X(t) \neq X(T_m)\}. \tag{39}$$

The probability that the number of orders at a distance i from the opposite best quote moves from n to $n+1$ at the next change is given by

$$P_i(n) \equiv \mathbb{P}[Q_i^A(T_{m+1}) = n + 1 \mid Q_i^A(T_m) = n, Q_i^A(T_{m+1}) \neq n]$$

$$= \begin{cases} \dfrac{\lambda(1)}{\lambda(1) + \mu + n\theta(1)}, & i = 1, \\[2mm] \dfrac{\lambda(i)}{\lambda(i) + n\theta(i)}, & i > 1. \end{cases} \tag{40}$$

To see how the above expression arises, consider the case $i = 1$. The next change in Q_1^A is an increase if an arrival of a limit order at price Q_1^A occurs before any of the limit orders at Q_1^A cancel or a market buy order occurs. However, because an arrival of a limit order at price Q_1^A occurs with rate $\lambda(1)$ and a cancellation or market buy order occurs at rate $\mu + n\theta(1)$, the probability that an arrival of a limit order occurs first is given by $\lambda(1)/(\lambda(1) + \mu + n\theta(1))$.

Denoting empirical quantities with a hat, e.g., $\hat{Q}_i^B(t)$ is the empirically observed number of bid orders at a distance of i units from the ask price at time t, an estimator for the above probability is given by

$$\hat{P}_i(n) \equiv \frac{\hat{B}_{up} + \hat{A}_{up}}{\hat{B}_{change} + \hat{A}_{change}},$$

where

$$\hat{B}_{up} = |\{m \mid \hat{Q}_i^B(\hat{T}_m) = n, \hat{Q}_i^B(\hat{T}_{m+1}) > n\}|,$$

$$\hat{A}_{up} = |\{m \mid \hat{Q}_i^A(\hat{T}_m) = n, \hat{Q}_i^A(\hat{T}_{m+1}) > n\}|,$$

$$\hat{B}_{change} = |\{m \mid \hat{Q}_i^B(\hat{T}_m) = n, \hat{Q}_i^B(\hat{T}_{m+1}) \neq n\}|, \quad \text{and}$$

$$\hat{A}_{change} = |\{m \mid \hat{Q}_i^A(\hat{T}_m) = n, \hat{Q}_i^A(\hat{T}_{m+1}) \neq n\}|.$$

In Figure 3, $P_i(n)$ and $\hat{P}_i(n)$ for $1 \leqslant i \leqslant 5$ are shown for Sky Perfect Communications. We see that these probabilities are reasonably close in most cases, indicating that the transition probabilities of the order book are well described by the model.

4.2.2. Direction of Price Moves. This subsection and the next two are devoted to the computation of conditional probabilities using the Laplace transform methods described in §3. These computations require the numerical inversion of Laplace transforms. The inversions are performed by shifting the random variable X under study by a constant c such that $\mathbb{P}[X + c \geqslant 0] \approx 1$, then inverting the corresponding one-sided Laplace transform using the methods proposed in Abate and Whitt (1992, 1995). When computing the probability of an increase in midprice, one can find a good shift c by using the fact that when $a = b$ the probability of an increase in midprice is 0.5. This shift c should also serve well for cases where $a \neq b$.

Table 3 compares the empirical frequencies of an increase in midprice to model-implied probabilities, given an initial configuration of b orders at the bid price, a orders at the ask price, and a spread of 1, for various values of a and b. We computed these quantities using Monte Carlo simulation (using 30,000 replications) and the Laplace transform methods described in §3. The simulation results, reported as 95% confidence intervals, agree with the Laplace transform computations and show that the probability of an increase in the midprice is well captured by the model.

4.2.3. Executing an Order Before the Midprice Moves. Table 4 gives probabilities computed using both simulation and our Laplace transform method for executing a bid order before a change in midprice for various values of a and b and for $S = 1$. Because our data set does not allow us to track specific orders, empirical values for these quantities, as well as the quantities in §4.2.4, are not obtainable.

4.2.4. Making the Spread. Table 5 gives probabilities computed using both simulation and our Laplace transform method for executing both a bid and an ask order at the best quotes before the midprice changes. One interesting observation here is that for a fixed value of a, as b is increased, the probability of making the spread is not monotone. Thus, for a fixed number of orders at the ask price the probability of making the spread is maximized for a nontrivial optimal number of orders at the bid price.

Figure 3. Probability of an increase in the number of orders at distance i from the opposite best quote in the next change, for $i = 1, \ldots, 5$.
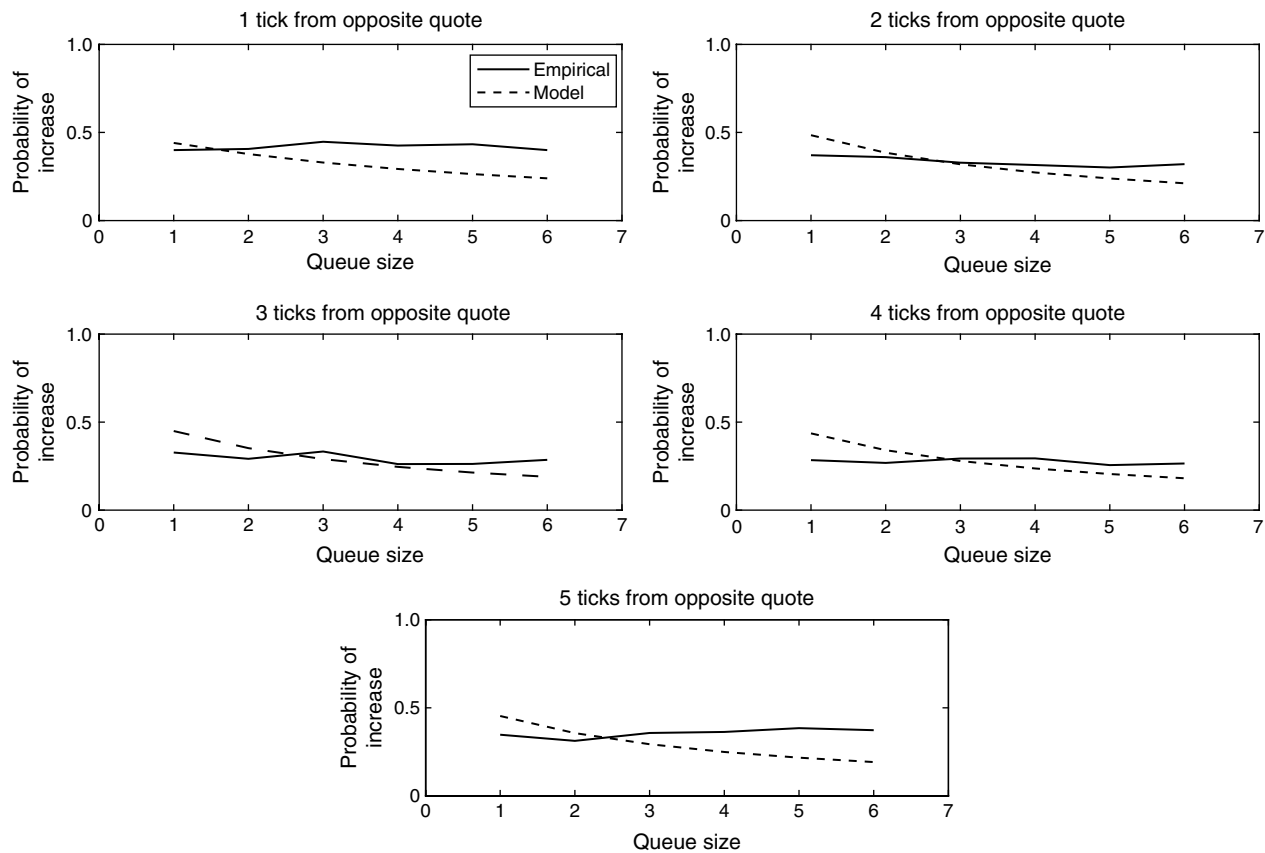
4.3. An Application to High-Frequency Trading

The conditional probabilities described in the above section may be used as a building block to construct systematic trading strategies. Such strategies fall into the realm of statistical arbitrage because they do not guarantee a profit, but lead to trades with positive expected returns and bounded losses. As a final exercise, we provide the reader with one such example based on our results in §3.2 on the probability that the midprice increases, conditional on the configuration of the book. In particular, using Equation (19), we can compute the probability that the midprice increases given that the spread is $S = 2$, the number of orders at the bid is $X_B(0) = b = 3$, and the number of orders at the ask $X_A(0) = a = 1$. A simple application of our Laplace transform results, with our estimated parameters for Sky Perfect Communications given in Table 2, yields a probability 0.62 of the midprice increasing. We use this as the basis for the following strategy, which we test in simulation:

Entering the position. If the spread is $S = 1$, the number of orders at the bid is $X_B(0) \geqslant 3$, the number of orders at the ask is $X_A(0) = 1$ and the number of orders at the second-best ask is $X_{p_A(0)+1}(0) \leqslant 1$, then submit a market buy order. Right after this trade, if $X_{p_A(0)+1}(0) = 1$, the new configuration of the order book will have $X_B(0+) = X_B(0) \geqslant 3$, $X_A(0+) = X_{p_A(0)+1}(0) = 1$,

and the spread will be $S = 2$. In this scenario, the probability of the midprice increasing is now 0.62, as stated above, and we have entered the position at the current midprice. Thus, we are in a good position to make a profit. In the case where $X_{p_A(0)+1}(0) = 0$, the order was bought at a price $X_A(0)$, which is strictly lower than the new midprice $(X_B(0+) + X_A(0+))/2 \geqslant (X_A(0) - 1 + X_A(0) + 2)/2 = X_A(0) + \frac{1}{2}$. In order for the trade to be welldefined, we must define an exit strategy.

Exiting the position. We submit a market sell order at the first time τ such that either

1. $p_B(\tau) > p_A(0)$, in which case we are selling at a price that is strictly greater than our buying price, or

2. $p_B(\tau) = p_B(0)$ and $X_B(\tau) = 1$, which results in a loss of one tick.

The probability of success of this round-trip transaction need not be recomputed in real time: if an "offline" computation (for example, using Laplace transform methods described in §3) indicates that the probability in (19) is large, this suggests that this strategy would perform well. Comparing this probability across different stocks may be a good indicator of the profitability of this strategy.

After running our simulation for 15,788 trades, roughly the equivalent of 30 days of trading, our algorithm does a total of 2,376 round-trip trades, and we display the P&L

Table 3. Probability of an increase in mid-price: empirical frequencies (top), simulation results (95% confidence intervals, middle), and Laplace transform method results (bottom).

			a		
b	1	2	3	4	5
1	0.512	0.304	0.263	0.242	0.226
2	0.691	0.502	0.444	0.376	0.359
3	0.757	0.601	0.533	0.472	0.409
4	0.806	0.672	0.580	0.529	0.484
5	0.822	0.731	0.640	0.714	0.606

			a		
b	1	2	3	4	5
1	0.499 ± 0.006	0.333 ± 0.005	0.258 ± 0.005	0.213 ± 0.005	0.187 ± 0.005
2	0.663 ± 0.005	0.495 ± 0.006	0.411 ± 0.006	0.346 ± 0.005	0.307 ± 0.005
3	0.743 ± 0.006	0.589 ± 0.006	0.506 ± 0.006	0.434 ± 0.006	0.389 ± 0.006
4	0.788 ± 0.005	0.652 ± 0.006	0.564 ± 0.006	0.503 ± 0.006	0.452 ± 0.006
5	0.811 ± 0.004	0.693 ± 0.005	0.615 ± 0.006	0.547 ± 0.006	0.504 ± 0.006

			a		
b	1	2	3	4	5
1	0.500	0.336	0.259	0.216	0.188
2	0.664	0.500	0.407	0.348	0.307
3	0.741	0.593	0.500	0.437	0.391
4	0.784	0.652	0.563	0.500	0.452
5	0.812	0.693	0.609	0.548	0.500

distribution in Figure 4. Note that the computed probability of 0.62 is not directly linked to the probability of the trade being successful, which may only be computed through simulation. Indeed, the probability of success of each round-trip transaction is less than 0.5, although the average profit of each trade was 0.068 ticks, or 6.8 yen. The analysis of the above trading strategy does not take into account transaction costs, but these can easily be included in the analysis.

5. Conclusion

We have proposed a stylized stochastic model describing the dynamics of a limit order book, where the occurrences of market events—market orders, limit orders

Table 4. Probability of executing a bid order before a change in midprice: simulation results (95% confidence intervals, top) and Laplace transform method results (bottom).

			a		
b	1	2	3	4	5
1	0.498 ± 0.004	0.642 ± 0.004	0.709 ± 0.004	0.748 ± 0.004	0.779 ± 0.004
2	0.299 ± 0.004	0.451 ± 0.004	0.536 ± 0.004	0.592 ± 0.004	0.632 ± 0.004
3	0.204 ± 0.004	0.335 ± 0.004	0.422 ± 0.004	0.484 ± 0.004	0.532 ± 0.004
4	0.152 ± 0.003	0.264 ± 0.004	0.344 ± 0.004	0.403 ± 0.004	0.450 ± 0.004
5	0.117 ± 0.003	0.213 ± 0.004	0.291 ± 0.004	0.342 ± 0.004	0.394 ± 0.004

			a		
b	1	2	3	4	5
1	0.497	0.641	0.709	0.749	0.776
2	0.302	0.449	0.535	0.591	0.631
3	0.206	0.336	0.422	0.483	0.528
4	0.152	0.263	0.344	0.404	0.452
5	0.118	0.213	0.287	0.346	0.393

Table 5. Probability of making the spread: simulation results (95% confidence intervals, top) and Laplace transform method results (bottom).

	a				
b	1	2	3	4	5
1	0.268 ± 0.004	0.306 ± 0.004	0.312 ± 0.004	0.301 ± 0.004	0.286 ± 0.004
2	0.306 ± 0.004	0.384 ± 0.004	0.406 ± 0.004	0.411 ± 0.004	0.401 ± 0.004
3	0.312 ± 0.004	0.406 ± 0.004	0.441 ± 0.004	0.455 ± 0.004	0.456 ± 0.004
4	0.301 ± 0.004	0.411 ± 0.004	0.455 ± 0.004	0.473 ± 0.004	0.485 ± 0.004
5	0.286 ± 0.004	0.401 ± 0.004	0.456 ± 0.004	0.485 ± 0.004	0.491 ± 0.004

	a				
b	1	2	3	4	5
1	0.266	0.308	0.309	0.300	0.288
2	0.308	0.386	0.406	0.406	0.400
3	0.309	0.406	0.441	0.452	0.452
4	0.300	0.406	0.452	0.471	0.479
5	0.288	0.400	0.452	0.479	0.491

and cancellations—are governed by independent Poisson processes.

The formulation of the model, which can be viewed as a queuing system, is entirely based on observable quantities so that its parameters can be easily estimated from observations of events in an actual order book. The model is simple enough to allow semianalytical computation of various conditional probabilities of order book events via Laplace transform methods, yet rich enough to adequately capture the short-term behavior of the order book: conditional distributions of various quantities of interest show good agreement with the corresponding empirical distributions for parameters estimated from data sets from the Tokyo Stock Exchange. The ability of our model to compute conditional distributions is useful for short-term prediction and design of automated trading strategies. Finally, simulation results illustrate that our model also yields realistic features for long-term (steady-state) average behavior of the order book profile and of price volatility.

One by-product of this study is to show how far a stochastic model can go in reproducing the dynamic properties of a limit order book without resorting to detailed behavioral assumptions about market participants or introducing unobservable parameters describing agent preferences, as in the market microstructure literature.
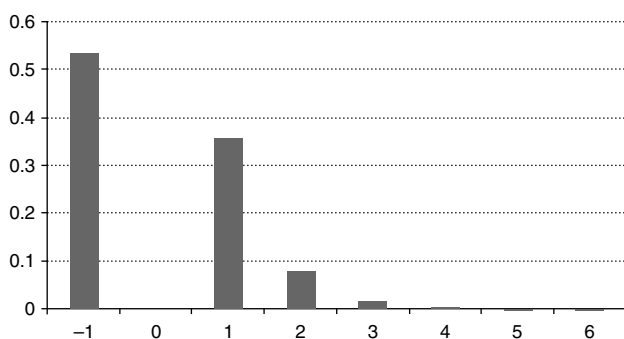
This model can be extended in various ways to take into account a richer set of empirically observed properties (Bouchaud et al. 2008). Correlation of the order flow with recent price behavior can be modeled by introducing state-dependent intensities of order arrivals. The heterogeneity of order sizes, which appears to be an important ingredient in actual order book dynamics, can be incorporated by making order sizes independent and identically distributed random variables. Both of these features would conserve the Markovian nature of the process. A more realistic distribution of interevent times may also be introduced by modelling the event arrivals via renewal processes. It remains to be seen whether the analytical tractability of the model can be preserved when such generalities are introduced. We look forward to exploring some of these extensions in future work.

Acknowledgments

Figure 4. Probability distribution of P&L per round-trip trade, in ticks.

References

Abate, J., W. Whitt. 1992. The Fourier-series method for inverting transforms of probability distributions. *Queueing Systems* **10** 5–88.

Abate, J., W. Whitt. 1995. Numerical inversion of Laplace transforms of probability distributions. *ORSA J. Comput.* **7**(1) 36–43.

Abate, J., W. Whitt. 1999. Computing Laplace transforms for numerical inversion via continued fractions. *INFORMS J. Comput.* **11**(4) 394–405.

Alfonsi, A., A. Schied, A. Schulz. 2010. Optimal execution strategies in limit order books with general shape functions. *Quant. Finance* **10**(2) 143–157.

Asmussen, S. 2003. *Applied Probability and Queues*. Springer-Verlag, New York.

Bouchaud, J.-P., D. Farmer, F. Lillo. 2008. How markets slowly digest changes in supply and demand. T. Hens, K. Schenk-Hoppe, eds. *Handbook of Financial Markets: Dynamics and Evolution*. North-Holland Publishers, Amsterdam, 57–160.

Bouchaud, J.-P., M. Mézard, M. Potters. 2002. Statistical properties of stock order books: Empirical results and models. *Quant. Finance* **2** 251–256.

Bovier, A., J. Černý, O. Hryniv. 2006. The opinion game: Stock price evolution from microscopic market modelling. *Internat. J. Theoret. Appl. Finance* **9** 91–111.

Farmer, J. D., L. Gillemot, F. Lillo, S. Mike, A. Sen. 2004. What really causes large price changes? *Quant. Finance* **4** 383–397.

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

Hollifield, B., R. A. Miller, P. Sandas. 2004. Empirical analysis of limit order markets. *Rev. Econom. Stud.* **71**(4) 1027–1063.

Luckock, H. 2003. A steady-state model of the continuous double auction. *Quant. Finance* **3** 385–404.

Maslov, S., M. Mills. 2001. Price fluctuations from the order book perspective—Empirical facts and a simple model. *PHYSICA A* **299** 234–246.

Obizhaeva, A., J. Wang. 2006. Optimal trading strategy and supply/demand dynamics. Working paper, Massachusetts Institute of Technology, Cambridge, MA.

Parlour, Ch. A. 1998. Price dynamics in limit order markets. *Rev. Financial Stud.* **11**(4) 789–816.

Ross, S. 1996. *Stochastic Processes*. John Wiley & Sons, New York.

Rosu, I. 2009. A dynamic model of the limit order book. *Rev. Financial Stud.* **22** 4601–4641.

Smith, E., J. D. Farmer, L. Gillemot, S. Krishnamurthy. 2003. Statistical theory of the continuous double auction. *Quant. Finance* **3**(6) 481–514.

Zovko, I., J. D. Farmer. 2002. The power of patience; A behavioral regularity in limit order placement. *Quant. Finance* **2** 387–392.