

# A Mathematical Approach to Order Book Modeling

Frédéric Abergel\* and Aymen Jedidi\*

August, 2011

## Abstract

We present a mathematical study of the order book as a multidimensional continuous-time Markov chain where the order flow is modeled by independent Poisson processes. Our aim is to bridge the gap between the microscopic description of price formation (agent-based modeling), and the Stochastic Differential Equations approach used classically to describe price evolution at macroscopic time scales. To do this, we rely on the theory of infinitesimal generators and Foster-Lyapunov stability criteria for Markov chains. We motivate our approach using an elementary example where the spread is kept constant (“perfect market making”). Then we compute the infinitesimal generator associated with the order book in a general setting, and link the price dynamics to the instantaneous state of the order book. In the last section, we prove that the order book is *ergodic*—in particular it has a *stationary distribution*—that it converges to its stationary state *exponentially fast*, and that the large-scale limit of the price process is a *Brownian motion*.

**Keywords:** Limit order book; agent-based modeling; order flow; bid-ask spread; Markov chain; stochastic stability; FCLT; geometric mixing.

## 1 Introduction and Background

The emergence of electronic trading as a major means of trading financial assets makes the study of the order book central to understanding the mechanisms of price formation. In order-driven markets, buy and sell orders are matched continuously subject to price and time priority. The *order book* is the list of all buy and sell limit orders, with their corresponding price and size, at a given instant of time. Essentially, three types of orders can be submitted:

- *Limit order*: Specify a price (also called “quote”) at which one is willing to buy or sell a certain number of shares;
- *Market order*: Immediately buy or sell a certain number of shares at the best available opposite quote;

---

\*Chaire de Finance Quantitative, Laboratoire de Mathématiques Appliquées aux Systèmes, École Centrale Paris, 92290 Châtenay-Malabry, France. E-mail: frederic.abergel@ecp.fr, aymen.jedidi@ecp.fr.

- *Cancellation order*: Cancel an existing limit order.

In the econophysics literature, “agents” who submit exclusively limit orders are referred to as *liquidity providers*. Those who submit market orders are referred to as *liquidity takers*.

Limit orders are stored in the order book until they are either executed against an incoming market order or canceled. The *ask* price  $P^A$  (or simply the ask) is the price of the best (i.e. lowest) limit sell order. The *bid* price  $P^B$  is the price of the best (i.e. highest) limit buy order. The gap between the bid and the ask

$$S := P^A - P^B, \quad (1)$$

is always positive and is called the *spread*. Prices are not continuous, but rather have a discrete resolution  $\Delta P$ , the *tick*, which represents the smallest quantity by which they can change. We define the *mid-price* as the average between the bid and the ask

$$P := \frac{P^A + P^B}{2}. \quad (2)$$

The price dynamics is the result of the interplay between the incoming order flow and the order book [2]. Figure 1 is a schematic illustration of this process [4]. Note that we chose to represent quantities on the bid side of the book by non-positive numbers.

Although in reality orders can have any size, we shall assume throughout this paper that all orders have a fixed unit size  $q$ . This assumption is convenient to carry out our analysis and is, for now, of secondary importance to the problem we are interested in.

## 2 An Elementary Approximation: Perfect Market Making

We start with the simplest agent-based market model:

- The order book starts in a full state: All limits above  $P^A(0)$  and below  $P^B(0)$  are filled with one limit order of unit size  $q$ . The spread starts equal to 1 tick;
- The flow of market orders is modeled by two independent Poisson processes  $M^+(t)$  (buy orders) and  $M^-(t)$  (sell orders) with constant arrival rates (or intensities)  $\lambda^+$  and  $\lambda^-$ ;
- There is one liquidity provider, who reacts immediately after a market order arrives so as to maintain the spread constantly equal to 1 tick. He places a limit order on the same side as the market order (i.e. a buy limit order after a buy market order and vice versa) with probability  $u$  and on the opposite side with probability  $1 - u$ .

The mid-price dynamics can be written in the following form

$$dP(t) = \Delta P(dM^+(t) - dM^-(t))Z, \quad (3)$$

where  $Z$  is a Bernoulli random variable

$$\begin{cases} Z = 0 & \text{with probability } (1 - u), \\ Z = 1 & \text{with probability } u. \end{cases} \quad (4)$$

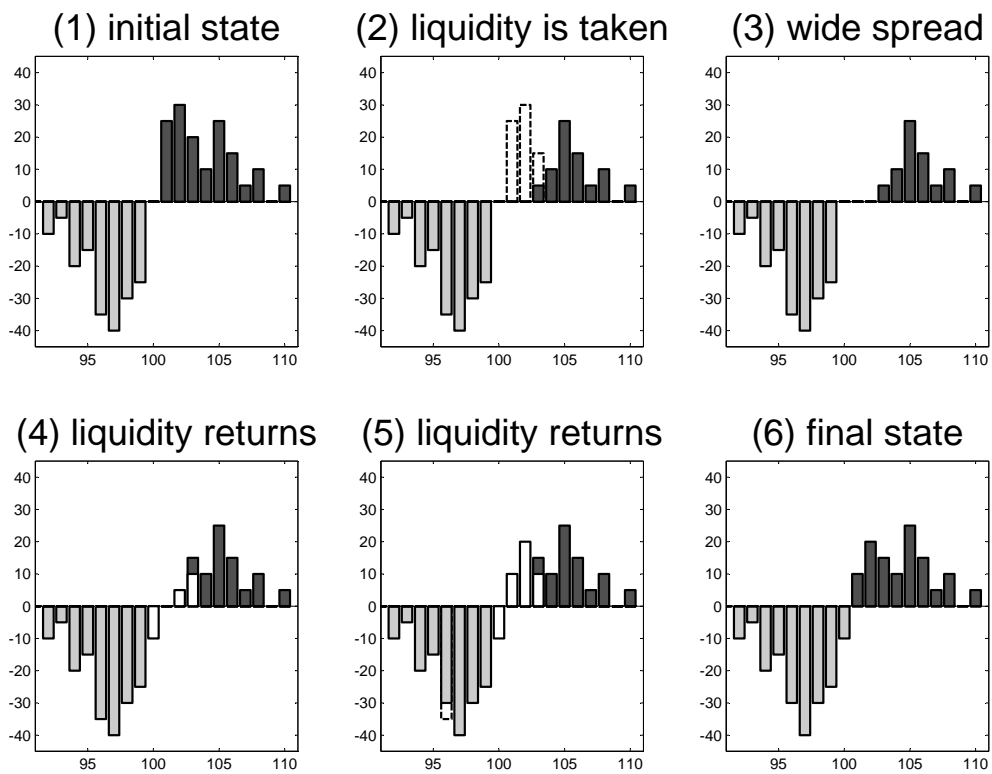


Figure 1: Order book schematic illustration: a buy market order arrives and removes liquidity from the ask side, then sell limit orders are submitted and liquidity is restored.

The infinitesimal generator<sup>1</sup> associated with this dynamics is

$$\mathcal{L}f(P) = u(\lambda^+(f(P + \Delta P) - f) + \lambda^-(f(P - \Delta P) - f)). \quad (6)$$

It is well known that a continuous limit is obtained under suitable assumptions on the intensity and tick size. Noting that (6) can be rewritten as

$$\begin{aligned} \mathcal{L}f(P) &= \frac{1}{2}u(\lambda^+ + \lambda^-)(\Delta P)^2 \frac{f(P + \Delta P) - 2f + f(P - \Delta P)}{(\Delta P)^2} \\ &+ u(\lambda^+ - \lambda^-)\Delta P \frac{f(P + \Delta P) - f(P - \Delta P)}{2\Delta P}, \end{aligned} \quad (7)$$

and under the following assumptions

$$\begin{cases} u(\lambda^+ + \lambda^-)(\Delta P)^2 \rightarrow \sigma^2 & \text{as } \Delta P \rightarrow 0, \\ u(\lambda^+ - \lambda^-)\Delta P \rightarrow \mu & \text{as } \Delta P \rightarrow 0, \end{cases} \quad (8)$$

the generator converges to the classical diffusion operator

$$\frac{\sigma^2}{2} \frac{\partial^2 f}{\partial P^2} + \mu \frac{\partial f}{\partial P}, \quad (9)$$

corresponding to a Brownian motion with drift. This simple case is worked out as an example of the type of limit theorems that we will be interested in in the sequel. One should also note that a more classical approach using the Functional Central limit Theorem (FCLT) as in [1] or [12] yields similar results ; For given fixed values of  $\lambda^+$ ,  $\lambda^-$  and  $\Delta P$ , the rescaled-centered price process

$$\frac{P(nt) - n\mu t}{\sqrt{n}\sigma} \quad (10)$$

converges as  $n \rightarrow \infty$ , to a standard Brownian motion ( $B(t)$ ) where

$$\begin{cases} \sigma = \Delta P \sqrt{(\lambda^+ + \lambda^-)u}, \\ \mu = \Delta P(\lambda^+ - \lambda^-)u. \end{cases} \quad (11)$$

Let us mention that one can easily achieve more complex diffusive limits such as a local volatility model by imposing that the limit is a function of  $P$  and  $t$

$$\begin{cases} u(\lambda^+ + \lambda^-)(\Delta P)^2 \rightarrow \sigma^2(P, t), \\ u(\lambda^+ - \lambda^-)\Delta P \rightarrow \mu(P, t). \end{cases} \quad (12)$$

This would be the case if the original intensities are functions of  $P$  and  $t$  themselves.

<sup>1</sup>The infinitesimal generator of a time-homogeneous Markov process  $(\mathbf{X}(t))_{t \geq 0}$  is the operator  $\mathcal{L}$ , if exists, defined to act on sufficiently regular functions  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , by

$$\mathcal{L}f(\mathbf{x}) := \lim_{t \downarrow 0} \frac{\mathbb{E}[f(\mathbf{X}(t)) | \mathbf{X}(0) = \mathbf{x}] - f(\mathbf{x})}{t}. \quad (5)$$

It provides an analytical tool to study  $(\mathbf{X}(t))$  [6].

### 3 Order Book Dynamics

#### Model setup: Poissonian arrivals, reference frame and boundary conditions

We now consider the dynamics of a general order book under a Poisson type assumption for the arrival of new market orders, limit orders and cancellations. We shall assume that each side of the order book is fully described by a *finite* number of limits  $N$ , ranging from 1 to  $N$  ticks away from the best available opposite quote. We will use the notation

$$\mathbf{X}(t) := (\mathbf{a}(t); \mathbf{b}(t)) := (a_1(t), \dots, a_N(t); b_1(t), \dots, b_N(t)), \quad (13)$$

where  $\mathbf{a} := (a_1, \dots, a_N)$  designates the ask side of the order book and  $a_i$  the number of shares available at price level  $i$  (i.e.  $i$  ticks away from the best opposite quote), and  $\mathbf{b} := (b_1, \dots, b_N)$  designates the bid side of the book. By doing so, we adopt the representation described e.g. in [3] or [11]<sup>2</sup>, but depart slightly from it by adopting a *finite moving frame*, as we think it is realistic and more convenient when scaling in tick size will be addressed.

Let us now recall the events that may happen:

- arrival of a new market order;
- arrival of a new limit order;
- cancellation of an already existing limit order.

These events are described by *independent* Poisson processes:

- $M^\pm(t)$ : arrival of new market order, with intensity  $\lambda^{M^+} \mathbb{I}(\mathbf{a} \neq \mathbf{0})$  and  $\lambda^{M^-} \mathbb{I}(\mathbf{b} \neq \mathbf{0})$ ;
- $L_i^\pm(t)$ : arrival of a limit order at level  $i$ , with intensity  $\lambda_i^{L^\pm}$ ;
- $C_i^\pm(t)$ : cancellation of a limit order at level  $i$ , with intensity  $\lambda_i^{C^+} \frac{a_i}{q}$  and  $\lambda_i^{C^-} \frac{|b_i|}{q}$ .

$q$  is the size of any new incoming order, and the superscript “+” (respectively “-”) refers to the ask (respectively bid) side of the book. Note that the intensity of the cancellation process at level  $i$  is proportional to the available quantity at that level. That is to say, each order at level  $i$  has a lifetime drawn from an exponential distribution with intensity  $\lambda_i^{C^\pm}$ . Note also that buy limit orders  $L_i^-(t)$  arrive below the ask price  $P^A(t)$ , and sell limit orders  $L_i^+(t)$  arrive above the bid price  $P^B(t)$ .

We impose constant boundary conditions outside the moving frame of size  $2N$ : Every time the moving frame leaves a price level, the number of shares at that level is set to  $a_\infty$  (or  $b_\infty$  depending on the side of the book). Our choice of a finite moving frame and constant<sup>3</sup> boundary conditions has three motivations. Firstly, it assures that the order book does not empty and that  $P^A$ ,  $P^B$  are always well defined. Secondly, it keeps the spread  $S$  and the increments of

<sup>2</sup>See also [5] for an interesting discussion.

<sup>3</sup>Actually, taking for  $a_\infty$  and  $|b_\infty|$  independent positive random variables would not change much our analysis. We take constants for simplicity.

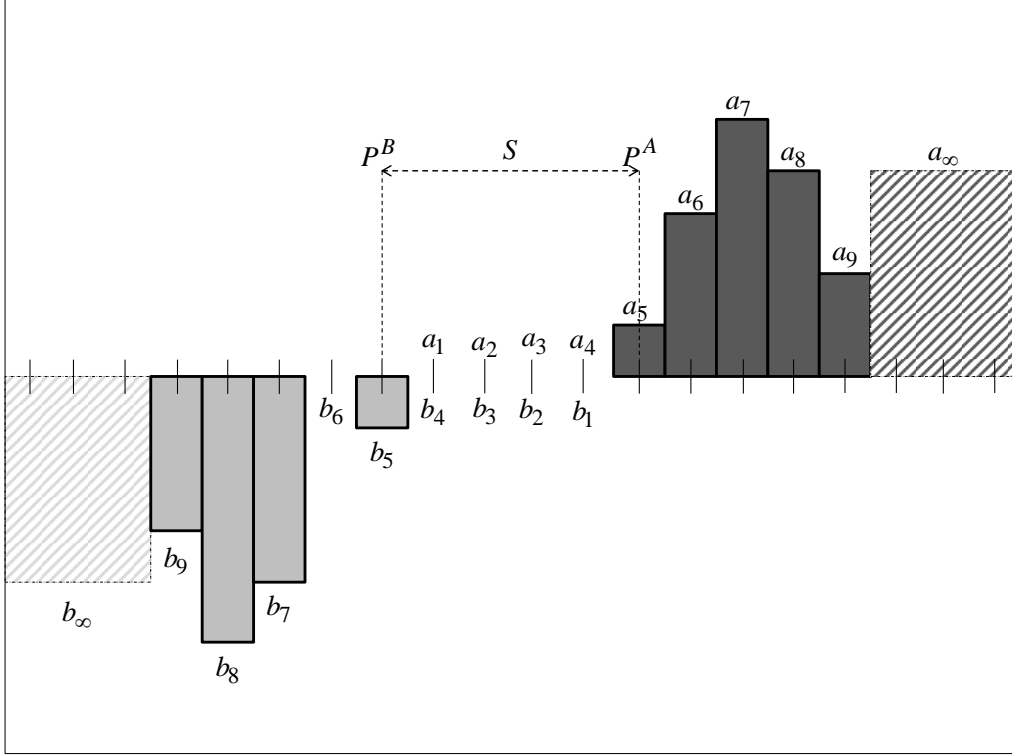


Figure 2: Order book dynamics: in this example,  $N = 9$ ,  $q = 1$ ,  $a_\infty = 4$ ,  $b_\infty = -4$ . The shape of the order book is such that  $\mathbf{a}(t) = (0, 0, 0, 0, 1, 3, 5, 4, 2)$  and  $\mathbf{b}(t) = (0, 0, 0, 0, -1, 0, -4, -5, -3)$ . The spread  $S(t) = 5$  ticks. Assume that at time  $t' > t$  a sell market order  $dM^-(t')$  arrives, then  $\mathbf{a}(t') = (0, 0, 0, 0, 0, 0, 1, 3, 5)$ ,  $\mathbf{b}(t') = (0, 0, 0, 0, 0, 0, -4, -5, -3)$  and  $S(t') = 7$ . Assume instead that at  $t' > t$  a buy limit order  $dL_1^-(t')$  arrives one tick away from the best opposite quote, then  $\mathbf{a}(t') = (1, 3, 5, 4, 2, 4, 4, 4, 4)$ ,  $\mathbf{b}(t') = (-1, 0, 0, 0, -1, 0, -4, -5, -3)$  and  $S(t') = 1$ .

$P^A$ ,  $P^B$  and  $P = \frac{P^A + P^B}{2}$  bounded—This will be important when addressing the diffusive limit of the price. Thirdly, it makes the model Markovian as we do not keep track of the price levels that have been visited (then left) by the moving frame at some prior time. Figure 2 is a representation of the order book using the above notations.

## Evolution of the order book

We can now write the following coupled SDEs for the quantities of outstanding limit orders in each side of the order book<sup>4</sup>

$$\left\{ \begin{array}{l} da_i(t) = - \left( q - \sum_{k=1}^{i-1} a_k \right)_+ dM^+(t) + qdL_i^+(t) - qdC_i^+(t) \\ \quad + (J^{M^-}(\mathbf{a}) - \mathbf{a})_i dM^-(t) + \sum_{i=1}^N (J^{L_i^-}(\mathbf{a}) - \mathbf{a})_i dL_i^-(t) + \sum_{i=1}^N (J^{C_i^-}(\mathbf{a}) - \mathbf{a})_i dC_i^-(t), \\ db_i(t) = \left( q - \sum_{k=1}^{i-1} |b_k| \right)_+ dM^-(t) - qdL_i^-(t) + qdC_i^-(t) \\ \quad + (J^{M^+}(\mathbf{b}) - \mathbf{b})_i dM^+(t) + \sum_{i=1}^N (J^{L_i^+}(\mathbf{b}) - \mathbf{b})_i dL_i^+(t) + \sum_{i=1}^N (J^{C_i^+}(\mathbf{b}) - \mathbf{b})_i dC_i^+(t), \end{array} \right. \quad (14)$$

where the  $J$ 's are *shift operators* corresponding to the renumbering of the ask side following an event affecting the bid side of the book and vice versa. For instance the shift operator corresponding to the arrival of a sell market order  $dM^-(t)$  of size  $q$  is<sup>5</sup>

$$J^{M^-}(\mathbf{a}) = \left( \underbrace{0, 0, \dots, 0}_{k \text{ times}}, a_1, a_2, \dots, a_{N-k} \right), \quad (15)$$

with

$$k := \inf \left\{ p : \sum_{j=1}^p |b_j| > q \right\} - \inf \{ p : |b_p| > 0 \}. \quad (16)$$

Similar expressions can be derived for the other events affecting the order book.

In the next sections, we will study some general properties of such models, starting with the generator associated with this  $2N$ -dimensional continuous-time Markov chain.

<sup>4</sup>Remember that, by convention, the  $b_i$ 's are non-positive.

<sup>5</sup>For notational simplicity, we write  $J^{M^-}(\mathbf{a})$  instead of  $J^{M^-}(\mathbf{a}; \mathbf{b})$  etc. for the shift operators.

## 4 Infinitesimal Generator

Let us now work out the infinitesimal generator associated with the jump process described above. We have

$$\begin{aligned}
\mathcal{L}f(\mathbf{a}; \mathbf{b}) &= \lambda^{M^+} (f([a_i - (q - A(i-1))_+]_+; J^{M^+}(\mathbf{b})) - f) \\
&+ \sum_{i=1}^N \lambda_i^{L^+} (f(a_i + q; J^{L^+}_i(\mathbf{b})) - f) \\
&+ \sum_{i=1}^N \lambda_i^{C^+} \frac{a_i}{q} (f(a_i - q; J^{C^+}_i(\mathbf{b})) - f) \\
&+ \lambda^{M^-} (f(J^{M^-}(\mathbf{a}); [b_i + (q - B(i-1))_+]_-) - f) \\
&+ \sum_{i=1}^N \lambda_i^{L^-} (f(J^{L^-}_i(\mathbf{a}); b_i - q) - f) \\
&+ \sum_{i=1}^N \lambda_i^{C^-} \frac{|b_i|}{q} (f(J^{C^-}_i(\mathbf{a}); b_i + q) - f),
\end{aligned} \tag{17}$$

where, to ease the notations, we note  $f(a_i; \mathbf{b})$  instead of  $f(a_1, \dots, a_i, \dots, a_N; \mathbf{b})$  etc. and

$$x_+ := \max(x, 0), \quad x_- := \min(x, 0), x \in \mathbb{R}. \tag{18}$$

The operator above, although cumbersome to put in writing, is simple to decipher: a series of standard difference operators corresponding to the ‘‘deposition-evaporation’’ of orders at each limit, combined with the shift operators expressing the moves in the best limits and therefore, in the origins of the frames for the two sides of the order book. Note the coupling of the two sides: the shifts on the  $a$ ’s depend on the  $b$ ’s, and vice versa. More precisely the shifts depend on the profile of the order book on the other side, namely the cumulative depth up to level  $i$  defined by

$$\begin{cases} A(i) := \sum_{k=1}^i a_k, \\ B(i) := \sum_{k=1}^i |b_k|, \end{cases} \tag{19}$$

and the generalized inverse functions thereof

$$\begin{cases} A^{-1}(q') := \inf \{p : \sum_{j=1}^p a_j > q'\}, \\ B^{-1}(q') := \inf \{p : \sum_{j=1}^p |b_j| > q'\}, \end{cases} \tag{20}$$

where  $q'$  designates a certain quantity of shares. Note that a more rigorous notation would be

$$A(i, \mathbf{a}(t)) \text{ and } A^{-1}(q', \mathbf{a}(t))$$

for the depth and inverse depth functions respectively. We drop the dependence on the last variable as it is clear from the context.

**Remark 4.1** *The index corresponding to the best opposite quote equals the spread  $S$  in ticks, that is*

$$\begin{cases} i_A := A^{-1}(0) = \inf\{p : \sum_{j=1}^p a_j > 0\} = \frac{S}{\Delta P} := i_S, \\ i_B := B^{-1}(0) = \inf\{p : \sum_{j=1}^p |b_j| > 0\} = \frac{S}{\Delta P} := i_S = i_A. \end{cases} \quad (21)$$

## 5 Price Dynamics

We now focus on the dynamics of the best ask and bid prices, denoted by  $P^A(t)$  and  $P^B(t)$ . One can easily see that they satisfy the following SDEs

$$\begin{cases} dP^A(t) = \Delta P[(A^{-1}(q) - A^{-1}(0))dM^+(t) \\ \quad - \sum_{i=1}^N (A^{-1}(0) - i)_+ dL_i^+(t) + (A^{-1}(q) - A^{-1}(0))dC_{i_A}^+(t)], \\ dP^B(t) = -\Delta P[(B^{-1}(q) - B^{-1}(0))dM^-(t) \\ \quad - \sum_{i=1}^N (B^{-1}(0) - i)_+ dL_i^-(t) + (B^{-1}(q) - B^{-1}(0))dC_{i_B}^-(t)], \end{cases} \quad (22)$$

which describe the various events that affect them: change due to a market order, change due to limit orders inside the spread, and change due to the cancellation of a limit order at the best price. One can summarize these two equations in order to highlight, in a more traditional fashion, the respective dynamics of the mid-price and the spread

$$\begin{aligned} dP(t) &= \frac{\Delta P}{2} [(A^{-1}(q) - A^{-1}(0))dM^+(t) - (B^{-1}(q) - B^{-1}(0))dM^-(t) \\ &\quad - \sum_{i=1}^N (A^{-1}(0) - i)_+ dL_i^+(t) + \sum_{i=1}^N (B^{-1}(0) - i)_+ dL_i^-(t) \\ &\quad + (A^{-1}(q) - A^{-1}(0))dC_{i_A}^+(t) - (B^{-1}(q) - B^{-1}(0))dC_{i_B}^-(t)], \end{aligned} \quad (23)$$

$$\begin{aligned} dS(t) &= \Delta P [(A^{-1}(q) - A^{-1}(0))dM^+(t) + (B^{-1}(q) - B^{-1}(0))dM^-(t) \\ &\quad - \sum_{i=1}^N (A^{-1}(0) - i)_+ dL_i^+(t) - \sum_{i=1}^N (B^{-1}(0) - i)_+ dL_i^-(t) \\ &\quad + (A^{-1}(q) - A^{-1}(0))dC_{i_A}^+(t) + (B^{-1}(q) - B^{-1}(0))dC_{i_B}^-(t)]. \end{aligned} \quad (24)$$

The equations above are interesting in that they relate in an explicit way the profile of the order book to the size of an increment of the mid-price or the spread, therefore linking the

price dynamics to the order flow. For instance the “infinitesimal” drifts of the mid-price and the spread, conditional on the shape of the order book at time  $t$ , are given by

$$\begin{aligned} \mathbb{E}[dP(t)|(\mathbf{a}; \mathbf{b})] &= \frac{\Delta P}{2} \left[ (A^{-1}(q) - A^{-1}(0))\lambda^{M^+} - (B^{-1}(q) - B^{-1}(0))\lambda^{M^-} \right. \\ &\quad - \sum_{i=1}^N (A^{-1}(0) - i)_+ \lambda_i^{L^+} + \sum_{i=1}^N (B^{-1}(0) - i)_+ \lambda_i^{L^-} \\ &\quad \left. + (A^{-1}(q) - A^{-1}(0))\lambda_{i_A}^{C^+} \frac{a_{i_A}}{q} - (B^{-1}(q) - B^{-1}(0))\lambda_{i_B}^{C^-} \frac{|b_{i_B}|}{q} \right] dt, \end{aligned} \quad (25)$$

and

$$\begin{aligned} \mathbb{E}[dS(t)|(\mathbf{a}; \mathbf{b})] &= \Delta P \left[ (A^{-1}(q) - A^{-1}(0))\lambda^{M^+} + (B^{-1}(q) - B^{-1}(0))\lambda^{M^-} \right. \\ &\quad - \sum_{i=1}^N (A^{-1}(0) - i)_+ \lambda_i^{L^+} - \sum_{i=1}^N (A^{-1}(0) - i)_+ \lambda_i^{L^-} \\ &\quad \left. + (A^{-1}(q) - A^{-1}(0))\lambda_{i_A}^{C^+} \frac{a_{i_A}}{q} + (B^{-1}(q) - B^{-1}(0))\lambda_{i_B}^{C^-} \frac{|b_{i_B}|}{q} \right] dt. \end{aligned} \quad (26)$$

## 6 Ergodicty and Diffusive Limit

In this section, our interest lies in the following questions:

1. Is the order book model defined above stable?
2. What is the stochastic-process limit of the price at large time scales?

The notions of “stability” and “large scale limit” will be made precise below. We first need some useful definitions from the theory of Markov chains and stochastic stability. Let  $(Q^t)_{t \geq 0}$  be the Markov transition probability function of the order book at time  $t$ , that is

$$Q^t(\mathbf{x}, E) := \mathbb{P}[\mathbf{X}(t) \in E | \mathbf{X}(0) = \mathbf{x}], \quad t \in \mathbb{R}_+, \mathbf{x} \in \mathcal{S}, E \subset \mathcal{S}, \quad (27)$$

where  $\mathcal{S} \subset \mathbb{Z}^{2N}$  is the state space of the order book. We recall that a (aperiodic, irreducible) Markov process is *ergodic* if an invariant probability measure  $\pi$  exists and

$$\lim_{t \rightarrow \infty} \|Q^t(\mathbf{x}, \cdot) - \pi(\cdot)\| = 0, \quad \forall \mathbf{x} \in \mathcal{S}, \quad (28)$$

where  $\|\cdot\|$  designates for a signed measure  $\nu$  the *total variation norm*<sup>6</sup> defined as

$$\|\nu\| := \sup_{f: |f| < 1} |\nu(f)| = \sup_{E \in \mathcal{B}(\mathcal{S})} \nu(E) - \inf_{E \in \mathcal{B}(\mathcal{S})} \nu(E). \quad (30)$$

<sup>6</sup>The convergence in total variation norm implies the more familiar pointwise convergence

$$\lim_{t \rightarrow \infty} |Q^t(\mathbf{x}, \mathbf{y}) - \pi(\mathbf{y})| = 0, \quad \mathbf{x}, \mathbf{y} \in \mathcal{S}. \quad (29)$$

Note that since the state space  $\mathcal{S}$  is countable, one can formulate the results without the need of a “measure-theoretic” framework. We prefer to use this setting as it is more flexible, and can accommodate possible generalizations of our results.

In (30),  $\mathcal{B}(\mathcal{S})$  is the Borel  $\sigma$ -field generated by  $\mathcal{S}$ , and for a measurable function  $f$  on  $\mathcal{S}$ ,  $\nu(f) := \int_{\mathcal{S}} f d\nu$ .

*V-uniform ergodicity.* A Markov process is said *V-uniformly ergodic* if there exists a coercive<sup>7</sup> function  $V > 1$ , an invariant distribution  $\pi$ , and constants  $0 < r < 1$ , and  $R < \infty$  such that

$$\|Q^t(\mathbf{x}, \cdot) - \pi(\cdot)\| \leq Rr^t V(\mathbf{x}), \mathbf{x} \in \mathcal{S}, t \in \mathbb{R}_+. \quad (31)$$

$V$ -uniform ergodicity can be characterized in terms of the infinitesimal generator of the Markov process. Indeed, it is shown in [7, 8] that it is equivalent to the existence of a coercive function  $V$  (the ‘‘Lyapunov test function’’) such that

$$\mathcal{L}V(\mathbf{x}) \leq -\beta V(\mathbf{x}) + \gamma, \quad (\text{Geometric Drift condition.}) \quad (32)$$

for some positive constants  $\beta$  and  $\gamma$ . (Theorems 6.1 and 7.1 in [8].) Intuitively, condition (32) says that the larger  $V(\mathbf{X}(t))$  the stronger  $\mathbf{X}$  is pulled back towards the center of the state space  $\mathcal{S}$ . A similar drift condition is available for discrete-time Markov processes  $(\mathbf{X}_n)_{n \in \mathbb{N}}$  and reads

$$\mathcal{D}V(\mathbf{x}) \leq -\beta V(\mathbf{x}) + \gamma \mathbb{1}_{\mathcal{C}}(\mathbf{x}), \quad (33)$$

where  $\mathcal{D}$  is the *drift operator*

$$\mathcal{D}V(\mathbf{x}) := \mathbb{E}[V(\mathbf{X}_{n+1}) - V(\mathbf{X}_n) | \mathbf{X}_n = \mathbf{x}]. \quad (34)$$

and  $\mathcal{C} \subset \mathcal{S}$  a finite set. (Theorem 16.0.1 in [7].) We refer to [7] for further details.

## Ergodicity of the order book and rate of convergence to the stationary state

Of utmost interest is the behavior of the order book in its stationary state. We have the following result:

**Theorem 6.1** *If  $\underline{\lambda}_{\mathcal{C}} = \min_{1 \leq i \leq N} \{\lambda_i^{C^\pm}\} > 0$ , then  $(\mathbf{X}(t))_{t \geq 0} = (\mathbf{a}(t); \mathbf{b}(t))_{t \geq 0}$  is an ergodic Markov process. In particular  $(\mathbf{X}(t))$  has a stationary distribution  $\pi$ . Moreover, the rate of convergence of the order book to its stationary state is exponential. That is, there exist  $r < 1$  and  $R < \infty$  such that*

$$\|Q^t(\mathbf{x}, \cdot) - \pi(\cdot)\| \leq Rr^t V(\mathbf{x}), t \in \mathbb{R}^+, \mathbf{x} \in \mathcal{S}. \quad (35)$$

**Proof.** Let

$$V(\mathbf{x}) := V(\mathbf{a}; \mathbf{b}) := \sum_{i=1}^N a^i + \sum_{i=1}^N |b^i| + q \quad (36)$$

<sup>7</sup>That is, a function such that  $V(\mathbf{x}) \rightarrow \infty$  as  $|\mathbf{x}| \rightarrow \infty$ .

be the total number of shares in the book (+ $q$  shares). Using the expression of the infinitesimal generator (17) we have

$$\begin{aligned} \mathcal{L}V(\mathbf{x}) &\leq -(\lambda^{M^+} + \lambda^{M^-})q + \sum_{i=1}^N (\lambda_i^{L^+} + \lambda_i^{L^-})q - \sum_{i=1}^N (\lambda_i^{C^+} a^i + \lambda_i^{C^-} |b^i|) \\ &\quad + \sum_{i=1}^N \lambda_i^{L^+} (i_S - i)_+ a^\infty + \sum_{i=1}^N \lambda_i^{L^+} (i_S - i)_+ |b^\infty| \end{aligned} \quad (37)$$

$$\begin{aligned} &\leq -(\lambda^{M^+} + \lambda^{M^-})q + (\Lambda^{L^-} + \Lambda^{L^+})q - \underline{\lambda}^C V(\mathbf{x}) \\ &\quad + N(\Lambda^{L^-} a^\infty + \Lambda^{L^+} |b^\infty|), \end{aligned} \quad (38)$$

where

$$\Lambda^{L^\pm} := \sum_{i=1}^N \lambda_i^{L^\pm} \quad \text{and} \quad \underline{\lambda}^C := \min_{1 \leq i \leq N} \{\lambda_i^{C^\pm}\} > 0. \quad (39)$$

The first three terms in the right hand side of inequality (37) correspond respectively to the arrival of a market, limit or cancellation order—ignoring the effect of the shift operators. The last two terms are due to shifts occurring after the arrival of a limit order inside the spread. The terms due to shifts occurring after market or cancellation orders (which we do not put in the r.h.s. of (37)) are negative, hence the inequality. To obtain inequality (38), we used the fact that the spread  $i_S$  is bounded by  $N + 1$ —a consequence of the boundary conditions we impose— and hence  $(i_S - i)_+$  is bounded by  $N$ .

The drift condition (38) can be rewritten as

$$\mathcal{L}V(\mathbf{x}) \leq -\beta V(\mathbf{x}) + \gamma, \quad (40)$$

for some positive constants  $\beta, \gamma$ . Inequality (40) together with theorem 7.1 in [8] let us assert that  $(\mathbf{X}(t))$  is  $V$ -uniformly ergodic, hence (35).  $\square$

**Corollary 6.1** *The spread  $S(t) = A^{-1}(0, \mathbf{a}(t))\Delta P = S(\mathbf{X}(t))$  has a well-defined stationary distribution—This is expected as by construction the spread is bounded by  $N + 1$ .*

## The embedded Markov chain

Let  $(\mathbf{X}_n)$  denote the embedded Markov chain associated with  $(\mathbf{X}(t))$ . In event time, the probabilities of each event are “normalized” by the quantity

$$\Lambda(\mathbf{x}) := \lambda^{M^+} + \lambda^{M^-} + \Lambda^{L^+} + \Lambda^{L^-} + \sum_{i=1}^N \lambda_i^{C^+} a_i + \sum_{i=1}^N \lambda_i^{C^-} |b_i|. \quad (41)$$

For instance, the probability of a buy market order when the order book is in state  $\mathbf{x}$ , is

$$\mathbb{P}[\text{“Buy market order at time } n\text{”} | \mathbf{X}_{n-1} = \mathbf{x}] := p^{M^+}(\mathbf{x}) = \frac{\lambda^{M^+}}{\Lambda(\mathbf{x})}. \quad (42)$$

The choice of the test function  $V(\mathbf{x}) = \sum_i a_i + \sum_i b_i + q$  does not yield a geometric drift condition, and more care should be taken to obtain a suitable test function. Let  $z > 1$  be a fixed real number and consider the function<sup>8</sup>

$$V(\mathbf{x}) := z^{\sum_i a_i + \sum_i |b_i|} := z^{\varphi(\mathbf{x})}. \quad (43)$$

We have

**Theorem 6.2**  $(\mathbf{X}_n)$  is  $V$ -uniformly ergodic. Hence, there exist  $r_2 < 1$  and  $R_2 < \infty$  such that

$$\|U^n(\mathbf{x}, \cdot) - \nu(\cdot)\| \leq R_2 r_2^n V(\mathbf{x}), n \in \mathbb{N}, \mathbf{x} \in \mathcal{S}. \quad (44)$$

where  $(U^n)_{n \in \mathbb{N}}$  is the transition probability function of  $(\mathbf{X}_n)_{n \in \mathbb{N}}$  and  $\nu$  its stationary distribution.

**Proof.**

$$\begin{aligned} \mathcal{D}V(\mathbf{x}) &\leq \frac{\lambda^{M^+}}{\Lambda(\mathbf{x})} (z^{\sum_i a_i - q + \sum_i |b_i|} - V(\mathbf{x})) \\ &+ \sum_j \frac{\lambda_j^{L^+}}{\Lambda(\mathbf{x})} (z^{\sum_i a_i + q + \sum_i |b_i| + N|b_\infty|} - V(\mathbf{x})) \\ &+ \sum_j \frac{\lambda_j^{C^+} a_j}{\Lambda(\mathbf{x})} (z^{\sum_i a_i - q + \sum_i |b_i|} - V(\mathbf{x})) \\ &+ \frac{\lambda^{M^-}}{\Lambda(\mathbf{x})} (z^{\sum_i a_i + \sum_i |b_i| - q} - V(\mathbf{x})) \\ &+ \sum_j \frac{\lambda_j^{L^-}}{\Lambda(\mathbf{x})} (z^{\sum_i a_i + N a_\infty + \sum_i |b_i| + q} - V(\mathbf{x})) \\ &+ \sum_j \frac{\lambda_j^{C^-} |b_j|}{\Lambda(\mathbf{x})} (z^{\sum_i a_i + \sum_i |b_i| - q} - V(\mathbf{x})). \end{aligned} \quad (45)$$

If we factor out  $V(\mathbf{x}) = z^{\sum a_i + \sum b_i}$  in the r.h.s of (45), we get

$$\begin{aligned} \frac{\mathcal{D}V(\mathbf{x})}{V(\mathbf{x})} &\leq \frac{\lambda^{M^+} + \lambda^{M^-}}{\Lambda(\mathbf{x})} (z^{-q} - 1) \\ &+ \frac{\Lambda^{L^-} + \Lambda^{L^+}}{\Lambda(\mathbf{x})} (z^{q + N d_\infty} - 1) \\ &+ \frac{\sum_j \lambda_j^{C^+} a_j + \sum_j \lambda_j^{C^-} |b_j|}{\Lambda(\mathbf{x})} (z^{-q} - 1), \end{aligned} \quad (46)$$

where

$$d_\infty := \max\{a_\infty, |b_\infty|\}. \quad (47)$$

<sup>8</sup>To save notations, we always use the letter  $V$  for the test function.

Then

$$\begin{aligned} \frac{\mathcal{D}V(\mathbf{x})}{V(\mathbf{x})} &\leq \frac{\lambda^{M^+} + \lambda^{M^-}}{\lambda^{M^+} + \lambda^{M^-} + \Lambda^{L^+} + \Lambda^{L^-} + \overline{\lambda^C} \varphi(\mathbf{x})} (z^{-q} - 1) \\ &+ \frac{\Lambda^{L^+} + \Lambda^{L^-}}{\lambda^{M^+} + \lambda^{M^-} + \Lambda^{L^+} + \Lambda^{L^-} + \underline{\lambda^C} \varphi(\mathbf{x})} (z^{q+Nd_\infty} - 1) \\ &+ \frac{\underline{\lambda^C} \varphi(\mathbf{x})}{\lambda^{M^+} + \lambda^{M^-} + \Lambda^{L^+} + \Lambda^{L^-} + \overline{\lambda^C} \varphi(\mathbf{x})} (z^{-q} - 1), \end{aligned} \quad (48)$$

with the usual notations

$$\underline{\lambda^C} := \min \lambda_i^{C^\pm} \quad \text{and} \quad \overline{\lambda^C} := \max \lambda_i^{C^\pm}. \quad (49)$$

Denote the r.h.s of (48)  $B(\mathbf{x})$ . Clearly

$$\lim_{\varphi(\mathbf{x}) \rightarrow \infty} B(\mathbf{x}) = \frac{\underline{\lambda^C} (z^{-q} - 1)}{\overline{\lambda^C}} < 0, \quad (50)$$

hence there exists  $A > 0$  such that for  $\mathbf{x} \in \mathcal{S}$  and  $\varphi(\mathbf{x}) > A$

$$\frac{\mathcal{D}V(\mathbf{x})}{V(\mathbf{x})} \leq \frac{\underline{\lambda^C} (z^{-q} - 1)}{2\overline{\lambda^C}} := -\beta < 0. \quad (51)$$

Let  $\mathcal{C}$  denote the finite set

$$\mathcal{C} = \{\mathbf{x} \in \mathcal{S} : \varphi(\mathbf{x}) = \sum_i a_i + \sum_i b_i \leq A\}. \quad (52)$$

We have

$$\mathcal{D}V(\mathbf{x}) \leq -\beta V(\mathbf{x}) + \gamma \mathbb{1}_{\mathcal{C}}(\mathbf{x}), \quad (53)$$

with

$$\gamma := \max_{\mathbf{x} \in \mathcal{C}} \mathcal{D}V(\mathbf{x}). \quad (54)$$

Therefore  $(\mathbf{X}_n)_{n \geq 0}$  is  $V$ -uniformly ergodic, by theorem 16.0.1 in [7].  $\square$

## The case of non-proportional cancellation rates

The proof above can be applied to the case where the cancellation rates are independent of the state of order book  $\mathbf{X}'(t)$ —We shall denote the order book  $\mathbf{X}'(t)$  in order to highlight that the assumption of proportional cancellation rates is relaxed. The probability of a cancellation  $dC_i^\pm(t)$  in  $[t, t + \delta t]$  is now

$$\mathbb{P}[C_i^\pm(t + \delta t) - C_i^\pm(t) = 1 | \mathbf{X}'(t) = \mathbf{x}'] = \lambda_i^{C^\pm} \delta t + o(\delta t), \quad (55)$$

where  $\lim_{\delta t \rightarrow 0} o(\delta t)/\delta t = 0$ . Since  $\Lambda = \lambda^{M^+} + \lambda^{M^-} + \Lambda^{L^+} + \Lambda^{L^-} + \sum_{i=1}^N \lambda_i^{C^+} + \sum_{i=1}^N \lambda_i^{C^-}$  does not depend on  $\mathbf{x}'$ , the analysis of the stability of the continuous-time process  $(\mathbf{X}'(t))$  and its discrete-time counterpart  $(\mathbf{X}'_n)$  are essentially the same.

We have the following result:

**Theorem 6.3** *Set*

$$\Lambda^{C^\pm} := \sum_{i=1}^N \lambda_i^{C^\pm} \quad \text{and} \quad \Lambda^{L^\pm} := \sum_{i=1}^N \lambda_i^{L^\pm}. \quad (56)$$

*Under the condition*

$$\lambda^{M^+} + \lambda^{M^-} + \Lambda^{C^+} + \Lambda^{C^-} > (\Lambda^{L^+} + \Lambda^{L^-})(1 + N \frac{d_\infty}{q}), \quad (57)$$

$(\mathbf{X}'_n)$  is  $V$ -uniformly ergodic. There exist  $r_3 < 1$  and  $R_3 < \infty$  such that

$$\|U^n(\mathbf{x}, \cdot) - \nu'(\cdot)\| \leq R_3 r_3^n, n \in \mathbb{N}, \mathbf{x} \in \mathcal{S}. \quad (58)$$

*The same is true for  $(\mathbf{X}'(t))$ .*

**Proof.** Let us prove the result for  $(\mathbf{X}'_n)$ . Inequality (46) is still valid by the same arguments, but this time the arrival rates are independent of  $\mathbf{x}'$

$$\begin{aligned} \frac{\mathcal{D}V(\mathbf{x}')}{V(\mathbf{x}')} &\leq \frac{\lambda^{M^+} + \lambda^{M^-}}{\Lambda} (z^{-q} - 1) \\ &+ \frac{\Lambda^{L^+} + \Lambda^{L^-}}{\Lambda} (z^{q+Nd_\infty} - 1) \\ &+ \frac{\Lambda^{C^+} + \Lambda^{C^-}}{\Lambda} (z^{-q} - 1). \end{aligned} \quad (59)$$

Set

$$z =: 1 + \epsilon > 1. \quad (60)$$

A Taylor expansion in  $\epsilon$  gives

$$\begin{aligned} \Lambda \frac{\mathcal{D}V(\mathbf{x}')}{V(\mathbf{x}')} &\leq (\lambda^{M^+} + \lambda^{M^-})(-q\epsilon) \\ &+ (\Lambda^{L^+} + \Lambda^{L^-})(q + Nd_\infty)\epsilon \\ &+ (\Lambda^{C^+} + \Lambda^{C^-})(-q\epsilon) + o(\epsilon). \end{aligned} \quad (61)$$

For  $\epsilon > 0$  small enough, the sign of (61) is determined by the quantity

$$-(\lambda^{M^+} + \lambda^{M^-}) + (\Lambda^{L^+} + \Lambda^{L^-})(1 + N \frac{d_\infty}{q}) - (\Lambda^{C^+} + \Lambda^{C^-}). \quad (62)$$

Hence, if (57) holds

$$\mathcal{D}V(\mathbf{x}') \leq -\beta V(\mathbf{x}') \quad \text{for some } \beta > 0, \quad (63)$$

and a geometric drift condition is obtained for  $\mathbf{X}'$ .  $\square$

If for concreteness we set  $q = 1$  share, and all the arrival rates are constant and symmetric, then condition (57) can be rewritten as

$$\lambda^M + N\lambda^C > N\lambda^L(1 + Nd_\infty). \quad (64)$$

where  $N$  is the size of the order book and  $d_\infty$  is the depth far away from the mid-price. Note that the above is a *sufficient* condition for ( $V$ -uniform) stability.

## Large scale limit of the price process

We are now able to answer the main question of this paper. Let us define the process  $e(t) \in \{1, \dots, 2(2N + 1)\}$  which indicates the last event

$$\{M^\pm, L_i^\pm, C_i^\pm\}_{i \in \{1, \dots, N\}},$$

that has occurred before time  $t$ .

**Lemma 6.1** *If we append  $e(t)$  to the order book  $(\mathbf{X}(t))$ , we get a Markov process*

$$\mathbf{Y}(t) := (\mathbf{X}(t), e(t)) \quad (65)$$

*which still satisfies the drift condition (32).*

**Proof.** Since  $e(t)$  takes its values in a finite set, the arguments of the previous sections are valid with minor modifications, and with the test functions

$$V(\mathbf{y}) := q + \sum a^i + \sum |b^i| + e, \quad (\text{continuous-time setting}) \quad (66)$$

$$V(\mathbf{y}) := e^{\sum a_i + \sum |b_i| + e}. \quad (\text{discrete-time setting}) \quad (67)$$

The  $V$ -uniform ergodicity of  $(\mathbf{Y}(t))$  and  $(\mathbf{Y}_n)$  follows.  $\square$

Given the state  $\mathbf{X}_{n-1}$  of the order book at time  $n - 1$  and the event  $e_n$ , the price increment at time  $n$  can be determined. (See equation (23).) We define the sequence of random variables

$$\eta_n := \Psi(\mathbf{X}_{n-1}, e_n) := \Phi(\mathbf{Y}_n, \mathbf{Y}_{n-1}), \quad (68)$$

as the price increment at time  $n$ .  $\Psi$  is a deterministic function giving the elementary “price-impact” of event  $e_n$  on the order book at state  $\mathbf{X}_{n-1}$ . Let  $\mu$  be the stationary distribution of  $(\mathbf{Y}_n)$ , and  $M$  its transition probability function. We are interested in the random sums

$$P_n := \sum_{k=1}^n \bar{\eta}_n = \sum_{k=1}^n \bar{\Phi}(\mathbf{Y}_k, \mathbf{Y}_{k-1}), \quad (69)$$

where

$$\bar{\eta}_k := \eta_k - \mathbb{E}_\mu[\eta_k] = \bar{\Phi}_k = \Phi_k - \mathbb{E}_\mu[\Phi_k], \quad (70)$$

and the asymptotic behavior of the rescaled-centered price process

$$\tilde{P}^{(n)}(t) := \frac{P_{\lfloor nt \rfloor}}{\sqrt{n}}, \quad (71)$$

as  $n$  goes to infinity.

**Theorem 6.4** *The series*

$$\sigma^2 = \mathbb{E}_\mu[\bar{\eta}_0^2] + 2 \sum_{n=1}^{\infty} \mathbb{E}_\mu[\bar{\eta}_0 \bar{\eta}_n] \quad (72)$$

*converges absolutely, and the rescaled-centered price process is a Brownian motion in the limit of  $n$  going to infinity. That is*

$$\tilde{P}^{(n)}(t) \xrightarrow{n \rightarrow \infty} \sigma B(t), \quad (73)$$

*where  $(B(t))$  is a standard Brownian motion.*

**Proof.** The idea is to apply the Functional Central Limit Theorem for (stationary and ergodic) sequences of *weakly dependent* random variables with *finite variance*. First, we note that the variance of the price increments  $\eta_n$  is finite since it is bounded by  $N + 1$ . Second, the  $V$ -uniform ergodicity of  $(\mathbf{Y}_n)$  is equivalent to

$$\|M^n(\mathbf{x}, \cdot) - \mu(\cdot)\| \leq R\rho^n V(\mathbf{x}), n \in \mathbb{N}, \quad (74)$$

for some  $R < \infty$  and  $\rho < 1$ . This implies thanks to theorem 16.1.5 in [7] that for any  $g^2, h^2 \leq V$ ,  $k, n \in \mathbb{N}$ , and any initial condition  $\mathbf{y}$

$$|\mathbb{E}_{\mathbf{y}}[g(\mathbf{Y}_k)h(\mathbf{Y}_{n+k})] - \mathbb{E}_{\mathbf{y}}[g(\mathbf{Y}_k)]\mathbb{E}_{\mathbf{y}}[h(\mathbf{Y}_k)]| \leq R\rho^n[1 + \rho^k V(\mathbf{y})], \quad (75)$$

where  $\mathbb{E}_{\mathbf{y}}[\cdot]$  means  $\mathbb{E}[\cdot | \mathbf{Y}_0 = \mathbf{y}]$ . This in turn implies

$$|\mathbb{E}_{\mathbf{y}}[\bar{h}(\mathbf{Y}_k)\bar{g}(\mathbf{Y}_{k+n})]| \leq R_1\rho^n[1 + \rho^k V(\mathbf{y})] \quad (76)$$

for some  $R_1 < \infty$ , where  $\bar{h} = h - \mathbb{E}_{\mu}[h]$ ,  $\bar{g} = g - \mathbb{E}_{\mu}[g]$ . By taking the expectation over  $\mu$  on both sides of (76) and noting that  $\mathbb{E}_{\mu}[V(\mathbf{Y}_0)]$  is finite by theorem 14.3.7 in [7], we get

$$|\mathbb{E}_{\mu}[\bar{h}(\mathbf{Y}_k)\bar{g}(\mathbf{Y}_{k+n})]| \leq R_2\rho^n =: \rho(n), k, n \in \mathbb{N}. \quad (77)$$

Hence the stationary version of  $(\mathbf{Y}_n)$  satisfies a *geometric mixing condition*, and in particular

$$\sum_n \rho(n) < \infty. \quad (78)$$

Theorems 19.2 and 19.3 in [1] on functions of mixing processes<sup>9</sup> let us conclude that

$$\sigma^2 := \mathbb{E}_{\mu}[\bar{\eta}_0^2] + 2 \sum_{n=1}^{\infty} \mathbb{E}_{\mu}[\bar{\eta}_0 \bar{\eta}_n] \quad (79)$$

is well-defined—the series in (79) converges absolutely—and coincides with the asymptotic variance

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}_{\mu} \left[ \sum_{k=1}^n (\bar{\eta}_k)^2 \right] = \sigma^2. \quad (80)$$

Moreover

$$\tilde{P}^{(n)}(t) \xrightarrow{n \rightarrow \infty} \sigma B(t), \quad (81)$$

where  $(B(t))$  is a standard Brownian motion. The convergence in (81) happens in  $D[0, \infty)$ , the space of  $\mathbb{R}$ -valued càdlàg functions, equipped with the Skorohod topology.  $\square$

**Remark 6.1** Obviously, theorem 6.4 is also true with non-proportional cancellation rates under condition (57). In this case the result holds both in event time and physical time. Indeed, let  $(N(t))_{t \in \mathbb{R}_+}$  denote a Poisson process with intensity  $\Lambda = \lambda^{M^{\pm}} + \lambda^{L^{\pm}} + \sum_{i=1}^N \lambda_i^{C^{\pm}}$ . The price process in physical time  $(P_c(t))_{t \in \mathbb{R}_+}$  can be linked to the price in event time  $(P_n)_{n \in \mathbb{N}}$  by

$$P_c(t) = P_{N(t)}. \quad (82)$$

<sup>9</sup>See also theorem 4.4.1 in [12] and discussion therein.

Then

$$\frac{P_{\lfloor kt \rfloor}}{\sqrt{k}} \xrightarrow{k \rightarrow \infty} \sigma B(t) \text{ as in theorem 6.4,} \quad (83)$$

and since  $\frac{N(u)}{\Lambda u} \xrightarrow{u \rightarrow \infty} 1$  a.s.,

$$\frac{P_c(kt)}{\sqrt{k}} = \frac{P_{N(kt)}}{\sqrt{k}} \xrightarrow{k \rightarrow \infty} \frac{P_{\lfloor \Lambda kt \rfloor}}{\sqrt{k}} \xrightarrow{k \rightarrow \infty} \sqrt{\Lambda} \sigma B(t). \quad (84)$$

**Remark 6.2** In the large scale limit, the mid-price  $P$ , the ask price  $P^A = P + \frac{S}{2}$ , and the bid price  $P^B = P - \frac{S}{2}$  converge to the same process ( $\sigma B(t)$ ).

## Numerical illustration

Figures 5–8 are obtained by numerical simulation of the order book. We note in particular the asymptotic normality of price increments, the fast decay of autocorrelation and the linear scaling of variance with time, in accordance with the theoretical analysis.

## 7 Conclusion and Prospects

This paper provides a simple Markovian framework for order book modeling, in which elementary changes in the price and spread processes are explicitly linked to the instantaneous shape of the order book and the order flow parameters. Two basic properties were investigated: the ergodicity of the order book and the large scale limit of the price process. The first property, which we answered positively, is desirable in that it assures the stability of the order book in the long run. The scaling limit of the price process is, as anticipated, a Brownian motion. A key ingredient in this result is the convergence of the order book to its stationary state at an exponential rate, a property equivalent to a geometric mixing condition satisfied by the stationary version of the order book. This short memory effect, plus a constraint on the variance of price increments guarantee a diffusive limit at large time scales. We hope that our approach offers a plausible microscopic justification to the much celebrated Bachelier model of asset prices.

We conclude with a final remark regarding two possible extensions: The assumption of a finite order book size—and hence a bounded spread—may seem artificial, and one can seek more general stability conditions for an order book model in which the spread is unbounded *a priori*. In addition, richer price dynamics (heavy tailed return distributions, long memory, more realistic spread distribution etc.) can be achieved with more complex assumptions on the order flow (e.g. feedback loops [10], or mutually exciting arrival rates [9]). These extensions may, however, render the model less amenable to mathematical analysis, and we leave the investigation of such interesting (but difficult) questions for future research.

## References

- [1] P. Billingsley, *Convergence of Probability Measures*, Wiley, 2<sup>nd</sup> ed. (1999).
- [2] J.P. Bouchaud, M. Mézard, M. Potters, Statistical properties of stock order books: empirical results and models, *Quantitative Finance* **2** (2002) 251–256.

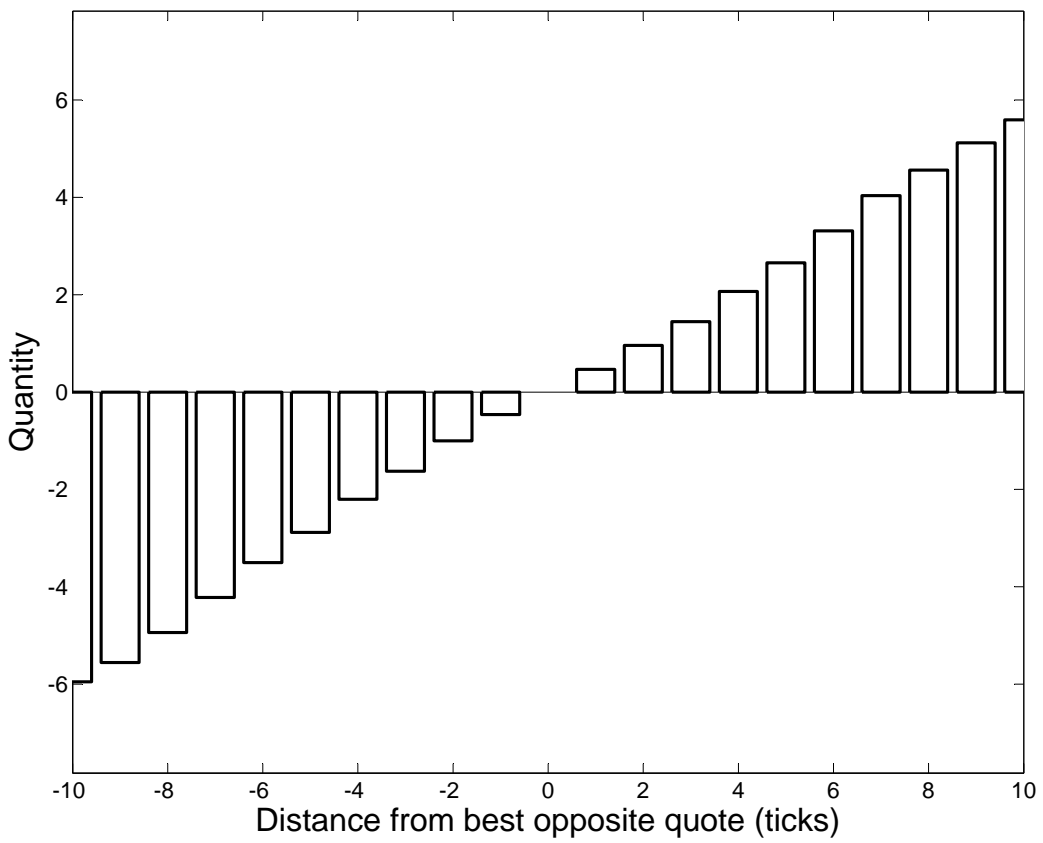


Figure 3: Average depth profile.

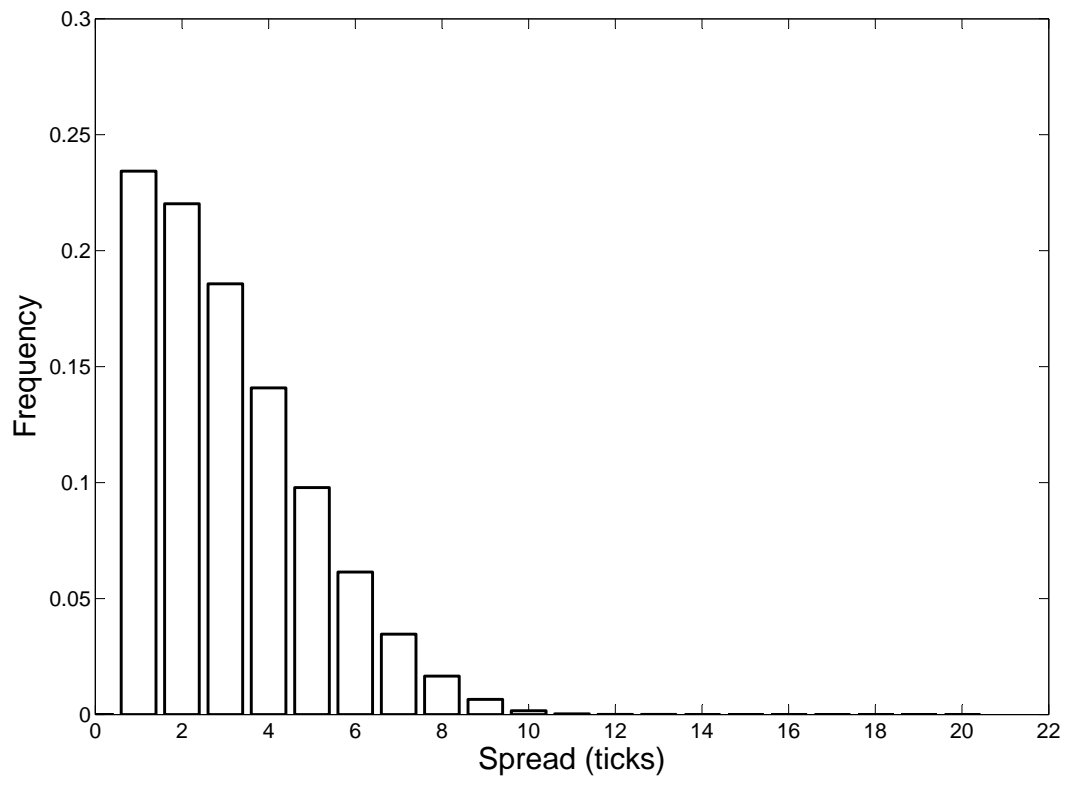


Figure 4: Histogram of the spread.

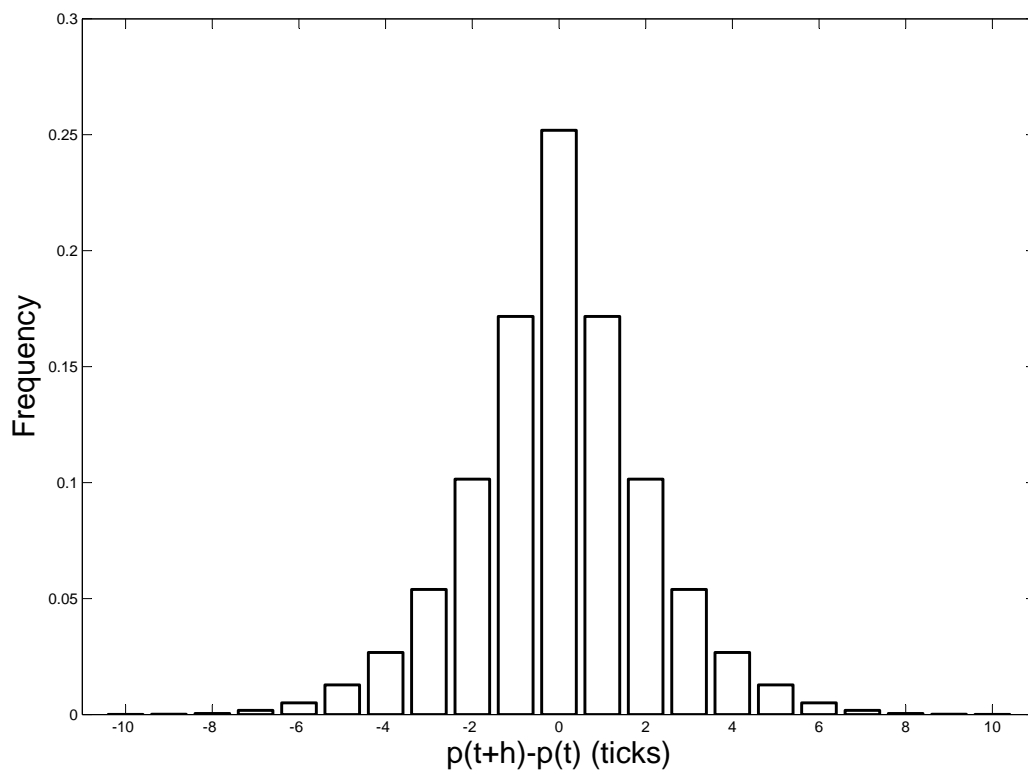


Figure 5: Histogram of price increments.

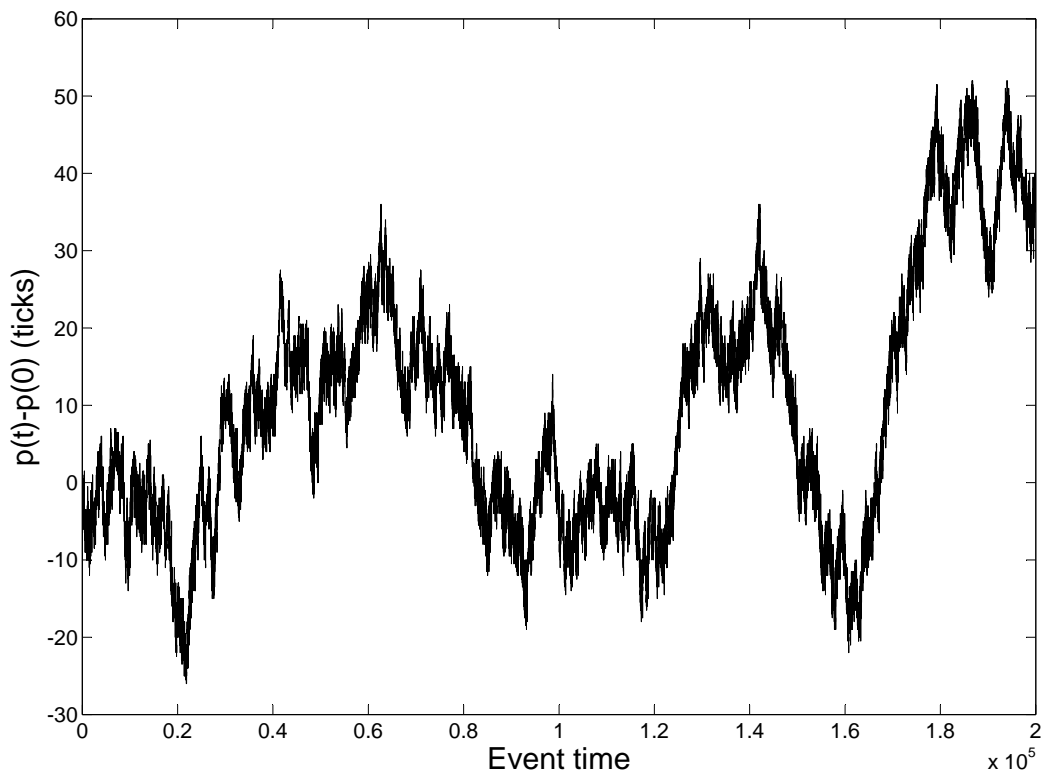


Figure 6: Price sample path.

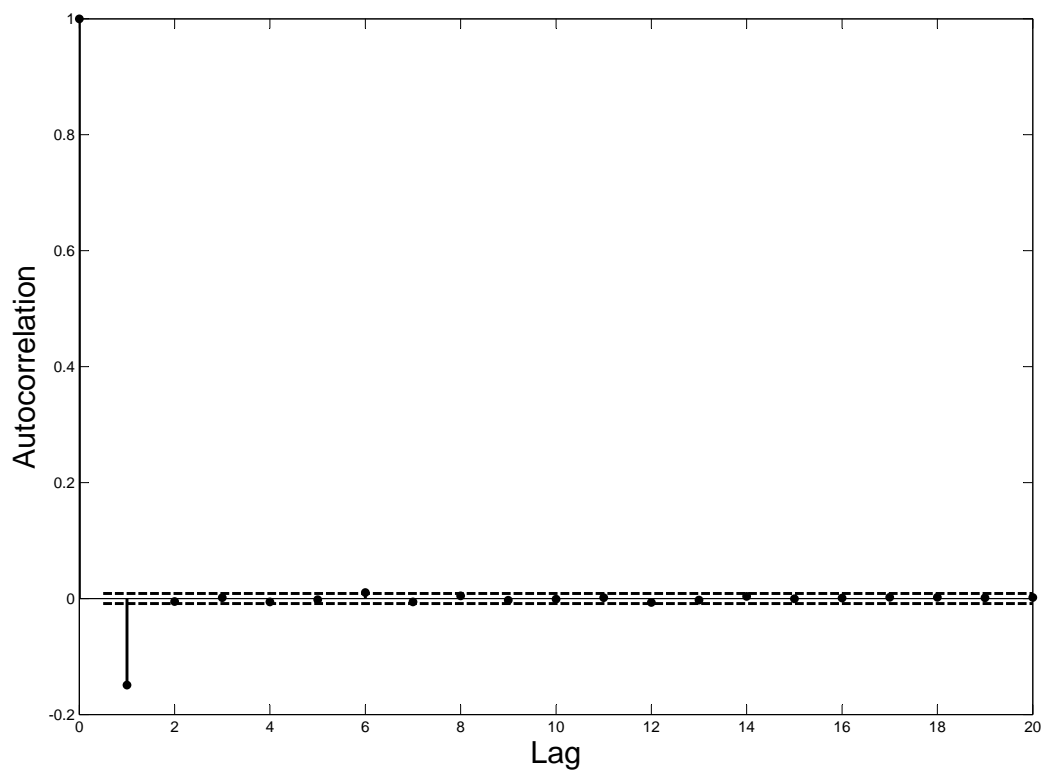


Figure 7: Autocorrelation of price increments.

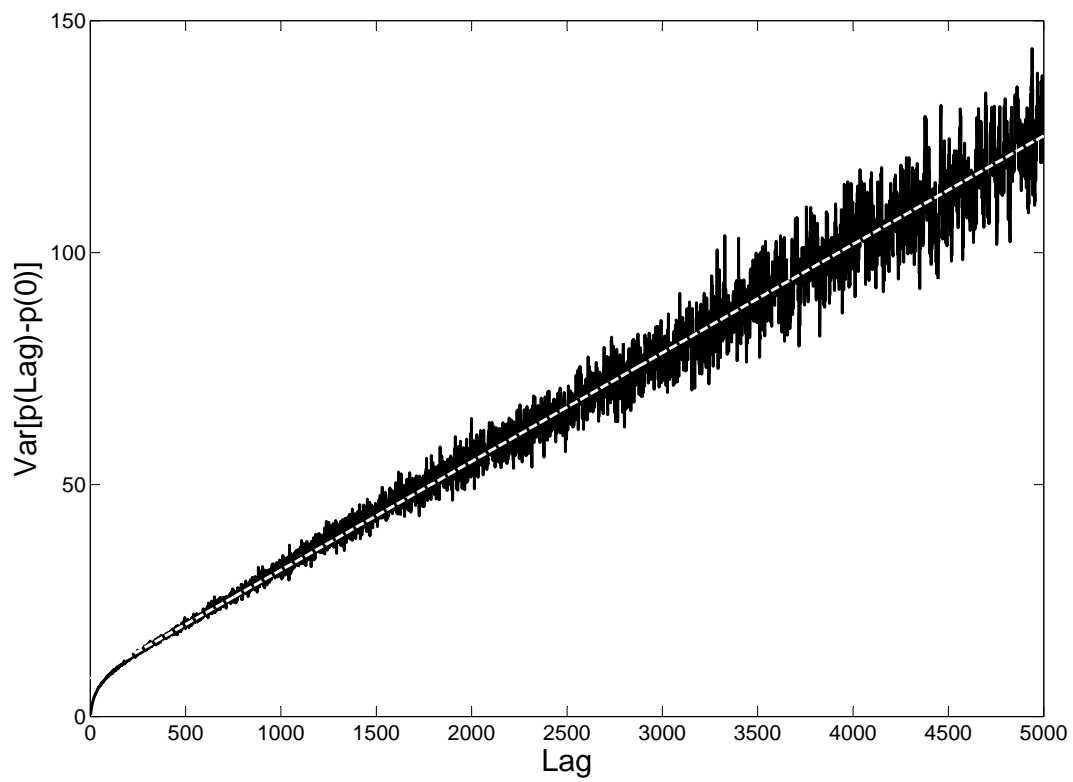


Figure 8: Variance in event time. The dashed line is a linear fit.

- [3] R. Cont, S. Stoikov and R. Talreja, A stochastic model for order book dynamics, *Operations Research* **58** (2010) 549–563.
- [4] A. Ferraris, Equity market impact models, Presentation (2008).
- [5] J. Gatheral and R. Oomen, Zero-intelligence realized variance estimation, *Finance and Stochastics* **14** (2010) 249–283.
- [6] F. Klebaner, *Introduction to Stochastic Calculus with Applications*, Imperial College Press, 2<sup>nd</sup> ed. (2005).
- [7] S. Meyn and R.L. Tweedie, *Markov Chains and Stochastic Stability*, Cambridge University Press, 2<sup>nd</sup> ed. (2009).
- [8] S. Meyn and R.L. Tweedie, Stability of Markovian processes III: Foster-Lyapunov criteria for continuous-time processes, *Advances in Applied Probability*, **25** (1993) 518–548.
- [9] I. Muni Toke, Market making behavior in an order book model and its impact on the spread, *Econophysics of Order-driven Markets*, Springer (2011).
- [10] T. Preis, S. Golke, W. Paul and J.J. Schneider, Multi-agent-based order book model for financial markets, *Europhysics Letters*, **75** (2006) 510–516.
- [11] E. Smith, J.D. Farmer, L. Guillemot and S. Krishnamurthy, Statistical theory of the continuous double auction, *Quantitative Finance* **3** (2003) 481–514.
- [12] W. Whitt, *Stochastic-Process Limits*, Springer (2002).