Price dynamics in a Markovian limit order market

Rama CONT & Adrien de LARRARD
IEOR Dept, Columbia University, New York
&
Laboratoire de Probabilités et Modèles Aléatoires
CNRS-Université Pierre et Marie Curie, Paris.

December 2010.

We propose and study a simple stochastic model for the dynamics of a limit order book, in which arrivals of market order, limit orders and order cancellations are described in terms of a Markovian queueing system. Through its analytical tractability, the model allows to obtain analytical expressions for various quantities of interest such as the distribution of the duration between price changes, the distribution and autocorrelation of price changes, and the probability of an upward move in the price, *conditional* on the state of the order book. We study the diffusion limit of the price process and express the volatility of price changes in terms of parameters describing the arrival rates of buy and sell orders and cancelations. These analytical results provide some insight into the relation between order flow and price dynamics in order-driven markets.

Key words: limit order book, market microstructure, queueing, diffusion limit, high-frequency data, liquidity, duration analysis, point process

---

Contents

1 Introduction **2**

2 A Markov model of limit order book dynamics **3**
2.1 Simplified representation of a limit order book . . . 3
2.2 Order book dynamics . . . 4
2.3 Price dynamics . . . 6
2.4 Summary . . . 6
2.5 Quantities of interest . . . 7

3 Analytical results **7**
3.1 Duration until the next price change . . . 7
3.2 Probability of upward move in the price for a balanced limit order book . . . 9
3.3 Dynamics of the price . . . 10

4 Diffusion limit of the price process **12**
4.1 Balanced order book . . . 13
4.2 Case when market orders and cancelations dominate . . . 14
4.3 Conclusion . . . 15

5 Proofs of technical lemmas **15**

Electronic copy available at: http://ssrn.com/abstract=1735338

1. Introduction

An increasing number of stocks are traded in electronic, order-driven markets, in which orders to buy and sell are centralized in a *limit order book* available to to market participants and market orders are executed against the best available offers in the limit order book. The dynamics of prices in such markets are not only interesting from the viewpoint of market participants –for trading and order execution (Alfonsi et al. (2010), Predoiu et al. (2010))– but also from a fundamental perspective, since they provide a rare glimpse into the dynamics of supply and demand and their role in price formation.

Equilibrium models of price formation in limit order markets (Parlour (1998), Rosu (2009)) have shown that the evolution of the price in such markets is rather complex and depends on the state of the order book. On the other hand, empirical studies on limit order books (Bouchaud et al. (2008), Farmer et al. (2004), Gourieroux et al. (1999), Hollifield et al. (2004), Smith et al. (2003)) provide an extensive list of statistical features of order book dynamics that are challenging to incorporate in a single model. While most of these studies have focused on unconditional/steady–state distributions of various features of the order book, empirical studies (see e.g. Harris and Panchapagesan (2005)) show that the state of the order book contains information on short-term price movements so it is of interest to provide forecasts of various quantities *conditional* on the state of the order book. Providing analytically tractable models which enable to compute and/or reproduce *conditional* quantities which are relevant for trading and intraday risk management has proven to be challenging, given the complex relation between order book dynamics and price behavior.

The search for tractable models of limit order markets has led to the development of stochastic models which aim to retain the main statistical features of limit order books while remaining computationally manageable. Stochastic models also serve to illustrate how far one 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 more detailed market microstructure models.

Starting from a description of order arrivals and cancelations as point processes Engle and Lunde (2003), the dynamics of a limit order book is naturally described in the language of queueing theory. Cont et al. (2010b) model the dynamics of a limit order book as a Markovian multiclass queueing system and compute various transition probabilities of the price conditional on the state of the order book, using Laplace transform methods.

We propose in this work a Markovian model of a limit order market, which captures some salient features of the dynamics of market orders and limit orders, yet is even simpler than the model of Cont et al. (2010b) and enables a wide range of properties of the price process to be computed analytically.

Our approach is motivated by the observation that, if one is primarily interested in the dynamics of the price, it is sufficient to focus on the dynamics of the (best) bid and ask queues. Indeed, empirical evidence shows that most of the order flow is directed at the best bid and ask prices (Biais et al. (1995)) and the imbalance between the order flow at the bid and at the ask appears to be the main driver of price changes (Cont et al. (2010a)).

Motivated by this remark, we propose a parsimonious model in which the limit order book is represented by the number of limit orders (q_t^b, q_t^a) sitting at the bid and the ask, represented as a system of two interacting Markovian queues. The remaining levels of the order book are treated as a 'reservoir' of limit orders represented by the distribution of the size of the queues at the 'next-to-best' price levels. Through its analytical tractability, our model allows to obtain analytical expressions for various quantities of interest such as the distribution of the duration until the next price change, the distribution and autocorrelation of price changes, and the probability of an upward move in the price, *conditional* on the state of the order book.

Electronic copy available at: http://ssrn.com/abstract=1735338

Order arrivals and cancelations are very frequent and occur at millisecond time scale, whereas, in many applications such as order execution, the metric of success is the volume-weighted average price (VWAP) so one is interested in the dynamics of order flow over a large time scale, typically tens of seconds or minutes. As shown in Table 1, thousands of order book events may occur over such time scales. This *aggregation* of events actually simplifies much of the analysis and enables us to use asymptotic methods. We study the link between price volatility and order flow in this model by studying the *diffusion limit* of the price process. In particular, we express the volatility of price changes in terms of parameters describing the arrival rates of buy and sell orders and cancelations. These analytical results provide some insight into the relation between order flow and price dynamics in order-driven markets.

	Average no. of orders in 10s	Price changes in 1 day
Citigroup	4469	12499
General Electric	2356	7862
General Motors	1275	9016

Table 1 Average number of orders in 10 seconds and number of price changes (June 26th, 2008).

The paper is organized as follows. Section 2 introduces a reduced-form representation of a limit order book and presents a Markovian model in which limit orders, market orders and cancellations occur according to Poisson processes. Section 3 presents various analytical results for this model: we compute the distribution of the duration until the next price change (section 3.1), the probability of upward move in the price (section 3.2) and the dynamics of the price (section 3.3). In Section 4, we show that the price behaves, at longer time scales, as a Brownian motion whose variance is expressed in terms of the parameters describing the order flow, thus establishing a link between volatility and order flow statistics.

2. A Markov model of limit order book dynamics

2.1. Simplified representation of a limit order book

Empirical studies of limit order markets suggest that the major component of the order flow occurs at the (best) bid and ask price levels (see e.g. Biais et al. (1995)). Furthermore, studies on the price impact of order book events show that the net effect of orders on the bid and ask queue sizes is the main factor driving price variations (Cont et al. (2010a)). These observations, together with the fact that queue sizes at the best bid and ask ("Level I" order book) are more easily obtainable (from trades and best quotes) than Level II data, motivate a reduced-form modeling approach in which we represent the state of the limit order book by

- the bid price s_t^b and the ask price s_t^a
- the size of the bid queue q_t^b representing the outstanding limit buy orders at the bid, and
- the size of the ask queue q_t^a representing the outstanding limit sell orders at the ask

Figure 1 summarizes this representation.

The bid and ask prices are multiples of the tick size δ. As shown in Table 2.1, for liquid stocks the *bid-ask spread* $s_t^a - s_t^b$ is equal to one tick for more than 98% of observations. We will therefore make the simplifying assumption that the spread is equal to one tick, i.e. $s_t^a = s_t^b + \delta$, resulting in a further reduction of dimension in the model.

The state of the limit order book is thus described by the triplet $X_t = (s_t^b, q_t^b, q_t^a)$ which takes values in the discrete state space $\delta.\mathbb{Z} \times \mathbb{N}^2$.

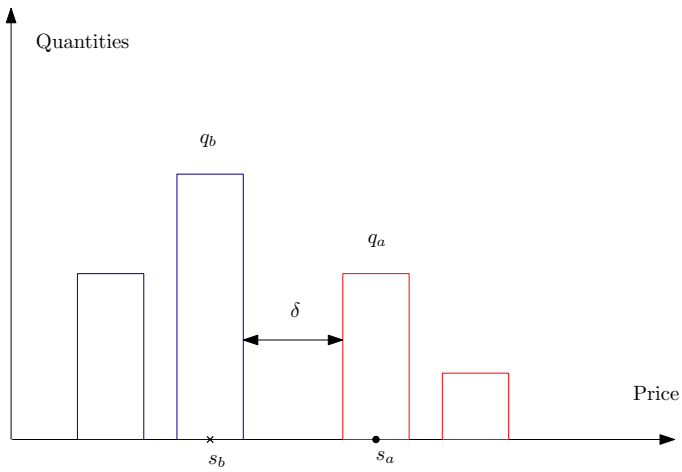

Figure 1 Simplified representation of a limit order book.

Bid-ask spread	1 tick	2 tick	$\geq$ 3 tick
Citigroup	98.82	1.18	0
General Electric	98.80	1.18	0.02
General Motors	98.71	1.15	0.14

Table 2 Percentage of observations with a given bid-ask spread (June 26th, 2008).

2.2. Order book dynamics

The state X_t of the order book is modified by *order book events*: limit orders (at the bid or ask), market orders and cancelations (see Cont et al. (2010b,a), Smith et al. (2003)). A limit buy (resp. sell) order of size x increases the size of the bid (resp. ask) queue by x, while a market buy (resp. sell) order decreases the corresponding queue size by x. Cancellation of x orders in a given queue reduces the queue size by x. Given that we are interested in the queue sizes at the best bid/ask levels, market orders and cancellations have the same effect on the state variable X_t.

We will assume that these events occur according to independent Poisson processes:

- Market buy (resp. sell) orders arrive at independent, exponential times with rate μ,
- Limit buy (resp. sell) orders at the (best) bid (resp. ask) arrive at independent, exponential times with rate λ,
- Cancellations occur at independent, exponential times with rate θ.
- These events are mutually independent.
- All orders sizes are equal (assumed to be 1 without loss of generality).

Denoting by $(T_i^a, i \geq 1)$ (resp. T_i^b) the times at which the size of ask (resp. the bid) queue changes and V_i^a (resp. V_i^a) the size of the associated change in queue size, the above assumptions translate into the following properties for the sequences $T_i^a, T_i^b, V_i^a, V_i^b$:

(i) $(T_{i+1}^a - T_i^a)_{i\geq 0}$ is a sequence of independent random variables with exponential distribution with parameter $\lambda + \theta + \mu$,

(ii) $(T_{i+1}^b - T_i^b)_{i\geq 0}$ is a sequence of independent random variables with exponential distribution with parameter $\lambda + \theta + \mu$,

(iii) $(V_i^a)_{i\geq 0}$ is a sequence of independent random variables with

$$\mathbb{P}[V_i^a = 1] = \frac{\lambda}{\lambda + \mu + \theta} \quad \textit{and} \quad \mathbb{P}[V_i^a = -1] = \frac{\mu + \theta}{\lambda + \mu + \theta}, \tag{1}$$

(iv) $(V_i^b)_{i\geq 0}$ is a sequence of independent random variables with

$$\mathbb{P}[V_i^b = 1] = \frac{\lambda}{\lambda+\mu+\theta} \quad and \quad \mathbb{P}[V_i^b = -1] = \frac{\mu+\theta}{\lambda+\mu+\theta} \tag{2}$$

- All the previous sequences are independent.

Once the bid (resp. the ask) queue is depleted, the price will move to the queue at the next level, which we assume to be one tick below (resp. above). The new queue size then corresponds to what was previously the number of orders sitting at the price immediately below (resp. above) the best bid (resp. ask). Instead of keeping track of these queues (and the corresponding order flow) at all price levels (as in Cont et al. (2010b), Smith et al. (2003)), we treat these sizes as stationary variables drawn from a certain distribution f on $\mathbb{N}^2$. Here $f(x,y)$ represents the probability of observing $(q_t^b, q_t^a) = (x,y)$ right after a price increase. Similarly, we denote $\tilde{f}(x,y)$ the probability of observing $(q_t^b, q_t^a) = (x,y)$ right after a price decrease. More precisely, denoting by $\mathcal{F}_t$ the history of prices and order book events on $[0,t]$,

- if $q_{t-}^a = 0$ then (q_t^b, q_t^a) is a random variable with distribution f, independent from $\mathcal{F}_{t-}$.
- if $q_{t-}^b = 0$ then (q_t^b, q_t^a) is a random variable with distribution $\tilde{f}$, independent from $\mathcal{F}_{t-}$.

Given the independence assumptions on event types, the probability that these two situations occur simultaneously is zero.

The distributions f and $\tilde{f}$ summarize the interaction of the queues at the best bid/ask levels with the rest of the order book, viewed here as a 'reservoir' of limit orders. For simplicity we shall assume $\tilde{f}(x,y) = f(y,x)$ i.e. events occurring on the bid and on the ask side have similar statistical properties but our analysis may be readily extended to the asymmetric case. Figure 2 shows the (joint) empirical distribution of bid and ask queue sizes after a price move for Citigroup stock on June 26th 2008.

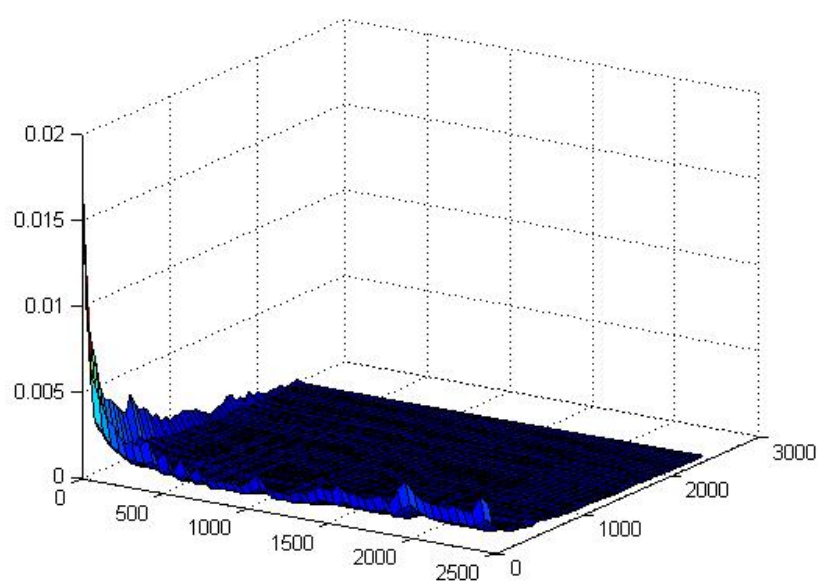

Figure 2 Joint density of bid and ask queue sizes after a price move (Citigroup, June 26th 2008).

Under these assumptions $q_t = (q_t^b, q_t^a)$ is thus a Markov process, taking values in $\mathbb{N}^2$, whose transitions correspond to the order book events $\{T_i^a, i \geq 1\} \cup \{T_i^b, i \geq 1\}$:

- At the arrival of a new limit buy (resp. sell) order the bid (resp. ask) queue increases by one unit. This occurs at rate λ.
- At each cancellation or market order, which occurs at rate $\theta+\mu$, either:
  (a) the corresponding queue decreases by one unit if it is >1, or
  (b) if the ask queue is depleted then q_t is an independent draw from the distribution f.
  (c) if the bid queue is depleted then q_t is an independent draw from the distribution $\tilde{f}$.

The values of λ and $\mu+\theta$ are readily estimated from high-frequency records of order books (see Cont et al. (2010b) for a description of the estimation procedure). Table 2.2 gives examples of such parameter estimates for the stocks mentioned above. We note that in all cases $\lambda<\mu+\theta$ but that the difference is small.

	$\hat{\lambda}$	$\hat{\mu}+\hat{\theta}$
Citigroup	2204	2331
General Electric	317	325
General Motors	102	104

Table 3 Estimates for the intensity of limit orders and market orders+cancellations, in number of batches per second (each batch representing 100 shares) on June 26th, 2008).

2.3. Price dynamics

When the bid or ask queue is depleted, the price moves up or down to the next level of the order book. We will assume that the order book contains no 'gaps' (empty levels) so that these price increments are equal to one tick:

- When the bid queue is depleted, the price decreases by one tick.
- When the ask queue is depleted, the price increases by one tick.

If there are gaps in the order book, this results in 'jumps' (i.e. variations of more than one tick) in the price. The price process s_t^b is thus a piecewise constant process whose transitions correspond to hitting times of the $\{(0,y), y\in\mathbb{N}\}\cup\{(x,0), x\in\mathbb{N}\}$ by the Markov process $q_t=(q_t^a,q_t^b)$.

2.4. Summary

In summary, the process $X_t=(s_t^b,q_t^b,q_t^a)$ is a continuous-time process with right-continuous, piecewise constant sample paths whose transitions correspond to the order book events $\{T_i^a, i\geq 1\}\cup\{T_i^b, i\geq 1\}$. At each event:

- If an order or cancelation arrives on the ask side i.e. $T\in\{T_i^a, i\geq 1\}$:

$$(s_T^b,q_T^b,q_T^a)=(s_{T-}^b,q_{T-}^b,q_{T-}^a+V_i^a)1_{q_{T-}^a>-V_i^a}+(S_{T-}^b+\delta,R_i^b,R_i^a)1_{q_{T-}^a\leq -V_i^a},$$

- If an order or cancelation arrives on the bid side i.e. $T\in\{T_i^b, i\geq 1\}$:

$$(s_T^b,q_T^b,q_T^a)=(s_{T-}^b,q_{T-}^b+V_i^b,q_{T-}^a)1_{q_{T-}^b>-V_i^b}+(s_{T-}^b-\delta,\tilde{R}_i^b,\tilde{R}_i^a)1_{q_{T-}^b\leq -V_i^b},$$

where $(V_i^a)_{i\geq 1}$ and $(V_i^b)_{i\geq 1}$ are sequences of IID variables with distribution given by (1)-(2), $(R_i)_{i\geq 1}=(R_i^b,R_i^a)_{i\geq 1}$ is a sequence of IID variables with (joint) distribution f, and $(\tilde{R}_i)_{i\geq 1}=(\tilde{R}_i^b,\tilde{R}_i^a)_{i\geq 1}$ is a sequence of IID variables with (joint) distribution $\tilde{f}$.

2.5. Quantities of interest

In applications, one is interested in computing various quantities that intervene in high frequency trading such as:

- the conditional distribution of the duration between price moves, given the state of the order book (Section 3.1),
- the probability of a price increase, given the state of the order book (Section 3.2),
- the dynamics of the price :autocorrelations and distribution and autocorrelations of price changes (section 3.3), and
- the volatility of the price (section 4).

We will show that all these quantities may be characterized analytically in this model, in terms of order flow statistics.

3. Analytical results

The high-frequency dynamics of the price may be described in terms of *durations* between successive price changes and the magnitude of these price changes. Given that the state of the (Level I) order book is observable, it is of interest to examine what information the current state of the order book gives about the dynamics of the price. We now proceed to show how the model presented above may be used to compute the conditional distributions of durations and price changes, given the current state of the order book, in terms of the arrival rates of market orders, limit orders and cancellations.

3.1. Duration until the next price change

We consider first the distribution of the duration until the next price change, starting from a given configuration of the order book. We define

- σ_a the first time when the ask queue $(q_t^a, t \geq 0)$ is depleted,
- σ_b the first time when the bid queue $(q_t^b, t \geq 0)$ is depleted

Since the queue sizes are constant between events, one can express these stopping times as:

$$\sigma^a = \inf\{T_i^a,\ q^a_{T_i^a -} + V_i^a = 0\} \qquad \sigma^b = \inf\{T_i^b,\ q^b_{T_i^b -} + V_i^b = 0\}$$

The price $(s_t, t \geq 0)$ moves when the queue $q_t = (q_t^b, q_t^a)$ hits one of the axes: the duration until the next price move is thus

$$\tau = \sigma_a \wedge \sigma_b.$$

The following theorem gives the distribution of the duration τ, conditional on the initial queue sizes:

Proposition 1 **(Distribution of duration until next price move)**. *The distribution of τ conditioned on the state of the order book is given by:*

$$\mathbb{P}[\tau > t | q_0^a = a,\ q_0^b = b] = \sqrt{(\frac{\mu+\theta}{\lambda})^{a+b}} \psi_{a,\lambda,\theta+\mu}(t) \psi_{b,\lambda,\theta+\mu}(t) \tag{3}$$

$$\text{where} \qquad \psi_{n,\lambda,\theta+\mu}(t) = \int_t^\infty \frac{n}{u} I_n(2\sqrt{\lambda(\theta+\mu)}u) e^{-u(\lambda+\theta+\mu)} du \tag{4}$$

and I_n is the modified Bessel function of the first kind. The conditional law of τ has a regularly varying tail

- *with tail exponent 2 if $\lambda < \mu + \theta$*
- *with tail exponent 1 if $\lambda = \mu + \theta$. In particular, if $\lambda = \mu + \theta$, $E[\tau | q_0^a = a, q_0^b = b] = \infty$ whenever $a > 0, b > 0$.*

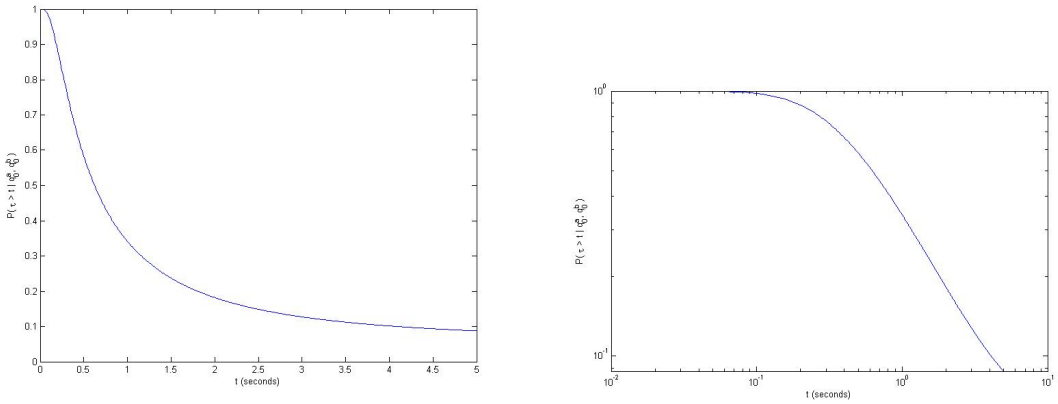

Figure 3 Left: $P(\tau > t | q_0^a = 4, q_0^b = 5)$ as a function of t for $\lambda = 12, \mu + \theta = 13$. Right: same figure in log-log coordinates. Note the Pareto tail which decays as t^{-2}.

Proof. Since $(q_t^a, t \geq 0)$ follows a birth and death process with birth rate λ and death rate $\mu+\theta$, $\mathcal{L}(s,x) := \mathbb{E}[e^{-s\sigma_a} | q_0^a = x]$ satisfies:

$$\mathcal{L}(s,x) = \frac{\lambda \mathcal{L}(s,x+1) + (\mu+\theta)\mathcal{L}(s,x-1)}{\lambda+\mu+\theta+s}.$$

We can find the roots of the polynomial: $\lambda X^2 - (\lambda+\mu+\theta+s)X + \mu+\theta$; one root is >1, the other is <1; since $\mathcal{L}(s,0) = 1$ and $\lim_{x\to\infty} \mathcal{L}(s,x) = 0$,

$$\mathcal{L}(s,x) = \left(\frac{(\lambda+\mu+\theta+s) - \sqrt{((\lambda+\mu+\theta+s))^2 - 4\lambda(\mu+\theta)}}{2\lambda}\right)^x.$$

Moreover if we use the relation $\mathbb{P}[\tau > t | q_0^a = x, q_0^b = y] = \mathbb{P}[\sigma^a > t | q_0^a = x]\mathbb{P}[\sigma^b > t | q_0^b = y]$,

$$\mathbb{P}[\tau > t | q_0^a = x, q_0^b = y] = \int_t^\infty \hat{\mathcal{L}}(u,x)du \int_t^\infty \hat{\mathcal{L}}(u,y)du.$$

This Laplace transform may be inverted (see (Feller 1971, XIV.7)) and the inversion yields

$$\hat{\mathcal{L}}(t,x) = \frac{x}{t}\sqrt{\left(\frac{\mu+\theta}{\lambda}\right)^x} \quad I_x(2\sqrt{\lambda(\theta+\mu)}t)e^{-t(\lambda+\theta+\mu)},$$

which gives us the expected result.

Tail behavior of τ:

- If $\lambda < \mu+\theta$:

$$\mathcal{L}(s,x) = \alpha(s)^x \underset{s\to 0}{\sim} 1 - \frac{x(\lambda+\mu+\theta)}{2\lambda(\mu+\theta-\lambda)}s,$$

so Karamata's Tauberian theorem (Feller 1971, XIII.5) yields

$$\mathbb{P}[\sigma^a > t | q_0^a = x] \underset{t\to\infty}{\sim} \frac{x(\lambda+\mu+\theta)}{2\lambda(\mu+\theta-\lambda)}\frac{1}{t};$$

therefore the conditional law of the duration τ is a regularly varying with tail index 2

$$\mathbb{P}[\tau > t | q_0^a = x, q_0^b = y] \underset{t\to\infty}{\sim} \frac{xy(\lambda+\mu+\theta)^2}{\lambda^2(\mu+\theta-\lambda)^2}\frac{1}{4t^2}. \tag{5}$$

- If the order flow is balanced i.e. $\lambda = \mu + \theta$ then

$$\mathcal{L}(s,x) = (\alpha(s))^x \underset{s\to 0}{\sim} 1 - \frac{x}{\sqrt{\lambda}}\sqrt{s},$$

the law of σ^a is regularly-varying with tail index $1/2$ and

$$\mathbb{P}[\sigma^a > t | q_0^a = x] \underset{t\to\infty}{\sim} \frac{x}{\sqrt{\pi\lambda}}\frac{1}{\sqrt{t}}.$$

The duration then follows a heavy-tailed distribution with infinite first moment:

$$\mathbb{P}[\tau > t | q_0^a = x, q_0^b = y] \underset{t\to\infty}{\sim} \frac{xy}{\pi\lambda}\frac{1}{t}; \tag{6}$$

The expression given in (3) is easily computed by discretizing the integral in (4). Plotting (3) for a fine grid of values of t typically takes less than a second on a laptop. Figure 3 gives a numerical example, with $\lambda = 12$ sec^{-1}, $\mu + \theta = 13$ sec^{-1}, $q_0^a = 4, q_0^b = 5$ (queue sizes are given in multiples of average batch size).

3.2. Probability of upward move in the price for a balanced limit order book

Assume now that $\lambda = \mu + \theta$, i.e. that the flow of limit orders is balanced by the flow of market orders and cancellations. Therefore for all $t \le \tau$, $q_t = M_{N_{2\lambda t}}$, where $(M_n, n \ge 0)$ is a symmetric random walk on $\mathbb{Z}^2$ killed when it hits either the x-axis or the y-axis and $(N_{2\lambda t}, t \ge 0)$ is a Poisson process with parameter 2λ. Hence the probability of an upward move in the price starting from a configuration $q_t^b = n, q_t^a = p$ for the order book is equal to the probability that the random walk M starting from (n,p) hits the x-axis before the y-axis. This probability is given by the following proposition:

Proposition 2. *For $(n,p) \in \mathbb{N}^2$, the probability $\phi(n,p)$ that the next price move is an increase, conditioned on having the n orders on the bid side and p orders on the ask side is:*

$$\phi(n,p) = \frac{1}{\pi}\int_0^\pi (2 - \cos(t) - \sqrt{(2-\cos(t))^2 - 1})^p \frac{\sin(nt)\cos(\frac{t}{2})}{\sin(\frac{t}{2})} dt. \tag{7}$$

Proof. The generator of the bivariate random walk $(M_n, n \ge 1)$ is the discrete Laplacian so $\phi(n,p) = \mathbb{P}[\sigma_a < \sigma_b | q_{0-}^b = n, q_{0-}^a = p]$ satisfies, for all $n \ge 1$ and $p \ge 1$,

$$4\phi(n,p) = \phi(n+1,p) + \phi(n-1,p) + \phi(n,p+1) + \phi(n,p-1), \tag{8}$$

with the boundary conditions: $\phi(0,p) = 0$ for all $p \ge 1$ and $\phi(n,0) = 1$ for all $n \ge 1$. This problem is known as the *discrete Dirichlet* problem; solutions of (8) are called discrete harmonic functions. (Lawler and Limic 2010, Ch. 8) show that for all $t \ge 0$, the functions

$$f_t(x,y) = e^{xr(t)}\sin(yt), \qquad \text{and} \qquad \tilde{f}_t(x,y) = e^{-xr(t)}\sin(yt) \qquad \text{with} \quad r(t) = \cosh^{-1}(2-\cos t)$$

are solutions of (8). In (Lawler and Limic 2010, Corollary 8.1.8) it is shown that the probability that a simple random walk $(M_k, k \ge 1)$ starting at $(n,p) \in \mathbb{Z}^+ \times \mathbb{Z}^+$ reaches the axes at $(x,0)$ is

$$\frac{2}{\pi}\int_0^\pi e^{-r(t)p}\sin(nt)\sin(tx)dt,$$

therefore

$$\phi(n,p) = \sum_{k=1}^\infty \frac{2}{\pi}\int_0^\pi e^{-r(t)p}\sin(tn)\sin(tk)dt.$$

Since

$$\sum_{k=1}^{m} \sin(kt) = \frac{\sin(\frac{mt}{2})\sin(\frac{(m+1)t}{2})}{\sin(t/2)} = \frac{\cos(\frac{t}{2}) - \cos((m+\frac{1}{2})t)}{2\sin(t/2)},$$

using integration by parts we see that the second term leads to the integral:

$$\int_0^\pi \underbrace{\frac{e^{-r(t)p}\sin(nt)}{\sin(t/2)}}_{g(t)} \cos((m+1/2)t)dt = -\frac{1}{m+\frac{1}{2}} \int_0^\pi g'(t)\sin((m+\frac{1}{2})t)dt \underset{m\to\infty}{\to} 0.$$

since g' is bounded. So finally:

$$\phi(n,p) = \frac{1}{\pi}\int_0^\pi e^{-r(t)p}\sin(tn)\frac{\cos(\frac{t}{2})}{\sin(\frac{t}{2})}dt.$$

Noting that $e^{-r(t)} = (2-\cos(t) - \sqrt{(2-\cos(t))^2 - 1})$ we obtain the result.

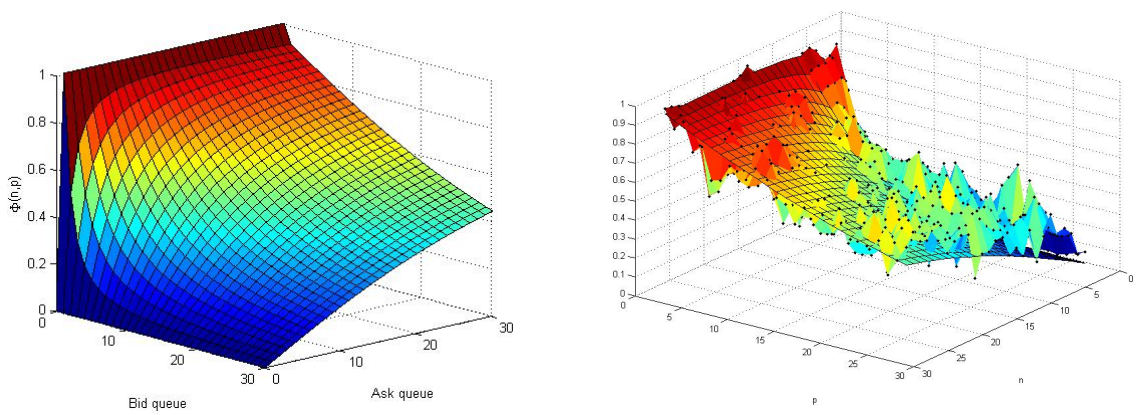

Figure 4 Left: Conditional probability of a price increase, as a function of the bid and ask queue size. Right: comparison with transition frequencies for CitiGroup tick-by-tick data on June 26, 2008.

It is remarkable that the conditional probabilities (7) are, in the case of a balanced order book, independent of the parameters describing the order flow. The expression (7) is easily computed numerically: Figure 4 displays the shape of the function Φ. The comparison with the corresponding empirical transition frequencies for CitiGroup tick-by-tick data on June 26, 2008 (Figure 4, right) shows good agreement between the theoretical conditional probabilities and their empirical counterparts.

3.3. Dynamics of the price

The high-frequency dynamics of the price in this model is described by a piecewise constant, right continuous process $(s_t, t \geq 0)$ whose jumps times correspond to times when the order book process $(q_t, t \geq 0)$ hits one of the axes. Denote by $(\tau_1, \tau_2, ...)$ the successive durations between price changes. The number of price changes that occur during $[0,t]$ is given by

$$N_t := \max\{\ n \geq 0, \quad \tau_1 + ... + \tau_n \leq t\ \}$$

At $t=\tau_i$, $s_{\tau_i}=s_{\tau_{i-}}+1$ if $q^a_{\tau_{i-}}=0$ and $s_{\tau_i}=s_{\tau_{i-}}-1$ if $q^b_{\tau_{i-}}=0$. $(X_1,X_2,X_3,...,X_n,...)$ are the successive moves in the price. Note that in general this is not a sequence of independent random variables. We define for $n\geq 1$,

$$Z_n=\sum_{i=1}^{n}X_i$$

the value of the price, after n changes. Hence, for all $t\geq 0$, $s_t=Z_{N_t}$.

PROPOSITION 3. *Let $p_{cont}=\mathbb{P}[X_2=\delta|X_1=\delta]=\mathbb{P}[X_2=-\delta|X_1=-\delta]$ be the probability of two successive price moves in the same direction.*

- $\forall k\geq 1,\quad Cov(X_1,X_k)=(2p_{cont}-1)^{k-1}$.
- *Conditional on the current state of the limit order book, the distribution of the n-th subsequent price change X_n is:*

$$p_n(x,y):=\mathbb{P}[X_n=\delta|q^a_0=x,\ q^b_0=y]=\frac{1+(2p_{cont}-1)^{n-1}(2p_1(x,y)-1)}{2},$$

Proof. Let, for $(x,y)\in\mathbb{N}^2$, and for all $n\geq 2$, $p_n(x,y)$ the probability that $X_n=\delta$, conditioned on $q^a_0=x$ and $q^b_0=y$. To simplify, we note p_n for $p_n(x,y)$. p_n is characterized by the following recurrence relation:

$$\begin{pmatrix} p_n \\ 1-p_n \end{pmatrix}=\begin{pmatrix} p_{cont} & 1-p_{cont} \\ 1-p_{cont} & p_{cont} \end{pmatrix}\begin{pmatrix} p_{n-1} \\ 1-p_{n-1} \end{pmatrix},$$

hence

$$\begin{pmatrix} p_n \\ 1-p_n \end{pmatrix}=\begin{pmatrix} p_{cont} & 1-p_{cont} \\ 1-p_{cont} & p_{cont} \end{pmatrix}^{n-1}\begin{pmatrix} p_1 \\ 1-p_1 \end{pmatrix}.$$

The eigenvalues of this matrix are 1 and $2p_{cont}-1$:

$$\begin{pmatrix} p_{cont} & 1-p_{cont} \\ 1-p_{cont} & p_{cont} \end{pmatrix}=\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 2p_{cont}-1 \end{pmatrix}\begin{pmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{pmatrix}.$$

Therefore

$$p_n=\frac{1+(2p_{cont}-1)^{n-1}(2p_1-1)}{2}.$$

Moreover for all $n\geq 2$,

$$Cov(X_1,X_n)=p_1p_n+(1-p_n)(1-p_1)-p_1(1-p_n)-p_n(1-p_1)$$

$$Cov(X_1,X_n)=(1+2p_np_1-p_n-p_1)$$

$$Cov(X_1,X_n)=(2p_{cont}-1)^{n-1}.$$

REMARK 1 (NEGATIVE AUTOCORRELATION OF PRICE CHANGES AT FIRST LAG). It is empirically observed that high frequency price movements have a negative autocorrelation at the first lag Cont (2001). In our model $Cov(X_k,X_{k+1})<0$ if and only if $p_{cont}<1/2$, which happens when

$$\sum_{i=1}^{\infty}\sum_{j\geq i}f(i,j)>1/2$$

where f is the joint distribution of queue sizes after a price increase. This condition is verified on all high-frequency data sets we have examined. For example, for CitiGroup stock we find

$$\sum_{i=1}^{\infty}\sum_{j\geq i}f(i,j)>0.7$$

This asymmetry condition on f corresponds to the fact that, after an upward price move, the new bid queue is generally smaller than the ask queue since the ask queue corresponds to the limit order previously sitting at second best ask level, while the bid queue results from the accumulation of orders over the very short period since the last price move. Under this condition, high frequency increments of the price are negatively correlated: an increase in the price is more likely to be followed by a decrease in the price.

REMARK 2. The sequence of price increments $(X_1, X_2, ...)$ is uncorrelated if and only if $p_{cont} = 1/2$ which happens when

$$\sum_{i=1}^{\infty} \sum_{j \geq i} f(i,j) = 1/2.$$

4. Diffusion limit of the price process

Assume $\lambda + \theta \leq \mu$ and that the joint distribution f of the queue sizes after a price move satisfies:

$$m(f) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} ijf(i,j) < \infty \tag{9}$$

As discussed in Section 3.3, the high frequency dynamics of the price is described by a piecewise constant stochastic process $s_t = Z_{N_t}$ where

$$Z_n = X_1 + ... + X_n \quad \textit{and} \quad N_t = \sup\{k;\ \tau_1 + ... + \tau_k \leq t\}$$

is the number of price moves during $[0, t]$.

However, over time scales much larger than the interval between individual order book events, prices are observed to have diffusive dynamics and modeled as such. To establish the link between the high frequency dynamics and the diffusive behavior at longer time scales, we shall consider a time scale $t_n = t\zeta(n)$ over which the average number of order book events is of order n and exhibit conditions under which the rescaled price process

$$(s_t^n := \frac{s_{t_n}}{\sqrt{n}}, t \geq 0)_{n \geq 1}$$

verifies a functional central limit theorem i.e. converges in distribution to a non-degenerate process $(p_t, t \geq 0)$ as $n \to \infty$. The choice of the time scale $t_n = t\zeta(n)$ cannot be arbitrary: it is imposed by the distributional properties of the durations which, as observed in Section 3.1, are heavy tailed. More precisely, $\zeta(n)$ is chosen such that

$$\frac{\tau_1 + ... + \tau_n}{\zeta(n)}$$

has a well-defined limit.

We will show that the limit p is then a diffusion process which describes the dynamics of the price at lower frequencies. In particular, we will compute the volatility of this *diffusion limit* p and relate it to the properties of the order flow.

In the following $\mathcal{D}$ denotes the space of right continuous paths $\omega : [0, \infty) \to \mathbb{R}^2$ with left limits, equipped with the Skorokhod topology J_1, and $\Rightarrow$ will designate weak convergence on $(\mathcal{D}, J_1)$ (see Billingsley (1968) for definitions).

4.1. Balanced order book

We first consider the case of a *balanced* order flow for which the intensity of market orders and cancelations is equal to the intensity of limit orders. The study of high-frequency quote data indicates that this is an empirically relevant case for many liquid stocks.

Theorem 1 *If $\lambda = \mu + \theta$,*

$$\left(\frac{s_{n\log nt}}{\sqrt{n}}, t \geq 0\right) \overset{n\to\infty}{\Rightarrow} \left(\delta\sqrt{\frac{\pi\lambda}{m(f)}}W_t, t \geq 0\right)$$

where δ is the tick size, $m(f)$ is given by (9) and W is a standard Brownian motion.

Denote by $\tau_0 = 1/\lambda$ the typical time scale separating order book events. Typically τ_0 is of the order of milliseconds. In plain terms, Theorem 1 states that, observed over a time scale $\tau_2 >> \tau_0$ (say, 10 minutes), the price has a diffusive behavior with a diffusion coefficient given by

$$\sigma = \delta\sqrt{\frac{n\pi\lambda}{m(f)}} \tag{10}$$

where δ is the tick size and $n \ln n \quad \tau_0 = \tau_2$. In particular, this formula shows that the microstructure of order flow affects price volatility through the ratio $\lambda/m(f)$ where λ is the rate of arrival of limit orders and $m(f)$, given by (9), is a measure of market depth: in fact, our model predicts a proportionality between the variance of price increments and this ratio. This is an empirically testable prediction: Figure 5 compares, for stocks in the Dow Jones index, the standard deviation of 10-minute price increments with $\sqrt{\lambda/m(f)}$. We observe that, indeed, stocks with a higher value of the ratio $\lambda/m(f)$ have a higher variance, and standard deviation of price increments increases roughly proportionally to $\sqrt{\lambda/m(f)}$.

Remark 4.1 *The short range correlations present in the increments of s_t vanish in the diffusion limit. This is consistent with empirical observations: price changes exhibit autocorrelations at the first few lags at high frequency but returns at lower frequency (say, hourly or daily) do not exhibit any significant autocorrelation.*

Proof. For all $t \geq 0$ and $n \geq 1$, let $t_n = n\log nt$ and

$$\frac{s_{n\log nt}}{\sqrt{n}} = \frac{Z(t\pi\lambda/m(f))\delta}{\sqrt{n}} + \left(\frac{Z(N_{t_n})\delta}{\sqrt{n}} - \frac{Z(t\pi\lambda/m(f))\delta}{\sqrt{n}}\right) \tag{11}$$

According to Proposition 5, the sequence of processes $(\frac{Z(t\pi\lambda n/m(f))}{\sqrt{n}}, t \geq 0)$ converges in $(\mathcal{D}, J_1)$ to a Brownian motion with volatility $\delta\sqrt{\frac{\pi\lambda}{m(f)}}$. Let $\rho : (1,\infty) \mapsto (1,\infty)$ be a function satisfying:

$$\rho(t)\log(\rho(t)) = t$$

Since $\rho(t) \sim_{t\to\infty} \frac{t}{\log(t)}$,

$$N_{t_n} \underset{n\to\infty}{\sim} \rho(\frac{\pi\lambda t\zeta(n)}{m(f)}) \sim \frac{t\pi\lambda\zeta(n)}{m(f)\log(\zeta(n))}, \tag{12}$$

$$N_{t_n} \underset{n\to\infty}{\sim} \frac{t\pi\lambda}{m(f)}.$$

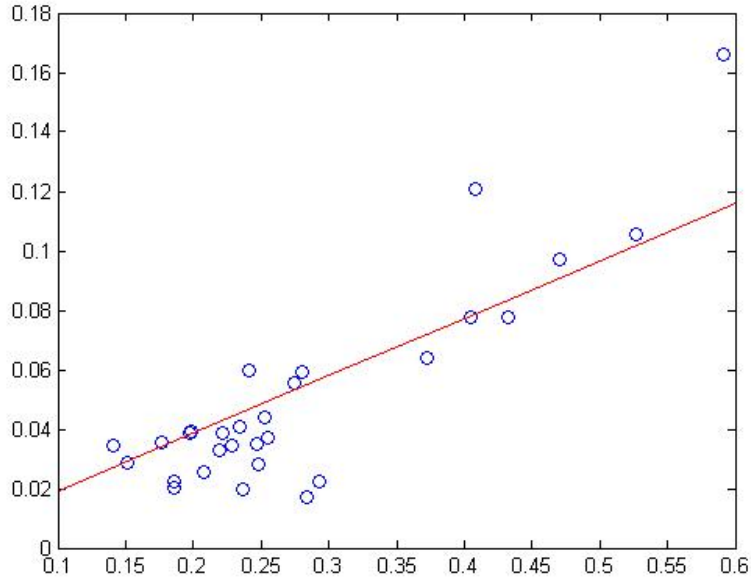

Figure 5 $\sqrt{\lambda/m(f)}$, estimated from tick-by-tick order flow (vertical axis) vs standard deviation of 10-minute price increments (horizontal axis) for stocks in the Dow Jones Index, estimated from high frequency data on June 26, 2008. Each point represents one stock. Red line indicates the best linear approximation.

Therefore for all $t \geq 0$,

$$\left(\frac{Z(N_{t_n})\delta}{\sqrt{n}} - \frac{Z(t\pi\lambda/m(f))\delta}{\sqrt{n}}\right) \overset{n\to\infty}{\Rightarrow} 0 \tag{13}$$

Therefore all finite dimensional distributions of the sequence of processes $\left(\frac{Z(N_{t_n})\delta}{\sqrt{n}} - \frac{Z(t\pi\lambda/m(f))\delta}{\sqrt{n}}\right)_{t\geq 0}$ converge to a point mass at zero . Since this sequence of processes is tight on $(\mathcal{D}, J_1)$, it converges weakly to zero on $(\mathcal{D}, J_1)$ (see Whitt (2002)). Finally,

$$\left(\frac{s_{n\log nt}}{\sqrt{n}}, t \geq 0\right) \overset{n\to\infty}{\Rightarrow} \delta\sqrt{\frac{\pi\lambda}{m(f)}}W,$$

4.2. Case when market orders and cancelations dominate

We now consider the case in which the flow of market orders and cancellations dominates that of limit orders: $\lambda < \theta + \mu$. In this case, price changes are more frequent since the order queues are depleted at a faster rate than they are replenished by market orders. We also obtain a diffusion limit though with a different scaling:

Theorem 2 *Let $\lambda < \theta + \mu$ and f a probability distribution on $\mathbb{N}^2$ which satisfies*

$$m(\lambda, \theta+\mu, f) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} m(\lambda, \theta+\mu, i, j) f(i,j) < \infty,$$

where for all $(x, y) \in (\mathbb{N}^*)^2$,

$$m(\lambda, \theta+\mu, x, y) = \int_0^\infty dt \int_t^\infty \psi_{x,\lambda,\mu+\theta}(u)du \int_t^\infty \psi_{y,\lambda,\mu+\theta}(u)du$$

where $\psi_{x,\lambda,\mu+\theta}$ is given by (4). Then

$$\left(\frac{s_{nt}}{\sqrt{n}}, t \geq 0\right) \overset{n\to\infty}{\Rightarrow} \left(\sqrt{\frac{1}{m(\lambda,\theta+\mu,f)}}\delta W_t, t \geq 0\right)$$

where W is a standard Brownian motion.

Proof. The sequence $(\tau_2, \tau_3, ...)$ is a sequence of i.i.d random variables with finite mean equal to $m(\lambda, \theta+\mu, f)$. We apply the law of large numbers:

$$\frac{\tau_1 + \tau_2 + ... + \tau_n}{n} \overset{n\to\infty}{\to} m(\lambda, \theta+\mu, f).$$

Therefore,

$$\forall t \geq 0, \qquad N_t^n \overset{n\to\infty}{\sim} [\frac{tn}{m(\lambda,\theta+\mu,f)}].$$

The rest of the proof follows the lines of the proof of theorem 1.

Variance of price change at intermediate frequency Theorem 2 leads to an expression of the variance of the price at a time scale $\tau >> \tau_0$, where $\tau_0 (\sim \text{ms})$ is the average interval between order book events:

$$\sigma^2 = \frac{\tau}{\tau_0} \frac{\pi\lambda}{m(f,\lambda,\theta+\mu)} \delta^2 \tag{14}$$

Note that while the left hand side of this equation is the variance of price changes (over a time scale τ_2), the right hand side only involves the tick size and quantities which relate to the statistical properties of the order flow.

4.3. Conclusion

We have exhibited a simple model of a limit order market in which order book events are described in terms of a Markovian queueing system. The analytical tractability of our model allows to compute various quantities of interest such as

- the distribution of the duration until the next price change,
- the distribution of price changes, and
- the diffusion limit of the price process and its volatility.

in terms of parameters describing the order flow. These results provide some insight into the relation between price dynamics and order flow in a limit order market. Moreover, the connection with two-dimensional queueing systems allows to use the rich analytical theory developed for these systems (see Cohen and Boxma (1983)) to compute many other quantities. We hope to pursue further some of these analytical ramifications in future work.

A relevant question is to examine which of the above results are robust to departures from the model assumptions and whether the intuitions conveyed by our model remain valid in a more general context where one or more of these assumptions are dropped. This issue is further studied in a companion paper Cont and de Larrard (2010) where we explore a more general dynamic model relaxing some of the assumptions above.

5. Proofs of technical lemmas

Lemma 3 *Let $p_{cont} = \mathbb{P}[X_2 = \delta | X_1 = \delta | q_0^a = x, q_0^b = y] = \mathbb{P}[X_2 = -\delta | X_1 = -\delta | q_0^a = x, q_0^b = y]$ be the probability of two successive increase (or decrease) in the price, then*

- *For all sequence of integers $(n_1, n_2, ..., n_p)$ and for all $k \geq 1$,*

$$\mathbb{E}[X_{n_1}^2 ... X_{n_p}^2 X_k] = \delta^{2p+1}(2p_1 - 1)(2p_{cont} - 1)^{k-1},$$

- *for all* $p \geq 1$,

$$\lim_{n\to\infty} \mathbb{E}[\frac{(X_1+X_2+...+X_n)^{2p}}{n^p}] = \frac{\delta^{2p}(2p)!}{2^p p!},$$

- *for all* $p \geq 1$,

$$\lim_{n\to\infty} \mathbb{E}[\frac{(X_1+X_2+...+X_n)^{2p+1}}{n^{p+1/2}}] = 0.$$

Proof. Since for all $i \geq 1$, $X_i^2 = 1$, $\mathbb{E}[X_{n_1}^2...X_{n_p}^2 X_k] = \mathbb{E}[X_k]\delta^{2p}$, the following relation

$$\mathbb{E}[X_{n_1}^2...X_{n_p}^2 X_k] = \delta^{2p+1}(2p_1-1)(2p_{cont}-1)^{k-1}$$

is valid for all sequence $(n_1, n_2, ..., n_p)$ and for all $k \geq 1$. For all $p \geq 0$,

$$\mathbb{E}[(X_1+...+X_m)^{2p}] = \sum_{k_1+...+k_m=2p} \frac{(2p)!}{k_1!...k_m!}\mathbb{E}[X_1^{k_1}...X_m^{k_m}];$$

for $k_1+....+k_m = 2p$, $\mathbb{E}[X_{k_1}^{k_1}...X_{k_m}^{k_m}]$ is maximal (equal to one) for $k_1 = k_2 = ... = k_p = 2$. The number of such p-uplets is equal to C_n^p, therefore as $n \to \infty$,

$$\mathbb{E}[(X_1+...+X_n)^{2p}] \sim \frac{n!}{p!(n-p)!}\frac{(2p)!}{p!}.$$

Moreover, since $\dfrac{n!}{(n-p)!n^p} \to 1$ as n goes to infinity,

$$\lim_{n\to\infty} \frac{\mathbb{E}[(X_1+...+X_n)^{2p}]}{n^p} = \frac{(2p)!}{2^p p!},$$

We prove that

$$\lim_{n\to\infty} \frac{\mathbb{E}[(X_1+...+X_n)^{2p+1}]}{n^{p+1/2}} = 0.$$

using the same method.

Lemma 4 *There is* $K > 0$ *such that for all* $n \geq 1$,

$$\mathbb{E}[(X_1+...+X_n)^4] \leq Kn^2.$$

Proof of lemma:

Let $(n_1, ..., n_k)$ be p-uplet of integers, then the random variable $X_{n_1}...X_{n_k}/\delta^k$ is a binomial random variable with parameter

$$\frac{1+(2p_1-1)^k(2p_{cont}-1)^{n_1+...+n_k-k}}{2};$$

so

$$\mathbb{E}[X_{n_1}...X_{n_k}] = \delta^k(2p_1-1)^k(2p_{cont}-1)^{n_1+...+n_k-k}$$

. We will use this relation in the following computations. For all $n \geq 1$, the multinomial formula yields

$$\mathbb{E}[(X_1+...+X_n)^4] \leq I_1+I_2+I_3+I_4,$$

where

$$I_1 = X_1^4 + ... + X_n^4 = n\delta, \qquad I_2 = 6\sum_{i=1}^{n}\sum_{j=i}^{n}\mathbb{E}[X_i^2 X_j^2] \leq 6n^2\delta^2,$$

$$I_3 = 12\sum_{i=1}^{n}\sum_{j=i}^{n}\sum_{l=j}^{n}\mathbb{E}[X_i^2 X_j X_k] \leq 12\delta^4(2p_1-1)^2 n\sum_{j=1}^{\infty}(2p_c-1)^{i-1}\sum_{k=1}^{\infty}(2p_c-1)^{j-1} \leq \frac{3\delta^4 n(2p_1-1)^2}{(1-p_{cont})^2},$$

$$I_4 = 24\sum_{i=1}^{n}\sum_{j=i}^{n}\sum_{l=j}^{n}\sum_{r=l}^{n}\mathbb{E}[X_i X_j X_k X_r] \leq 24\delta^4(2p_1-1)^4\left(\sum_{i=1}^{\infty}(2p_c-1)^i\right)^4 \leq \frac{3\delta^4(2p_1-1)^4}{2(p_c-1)^2}.$$

Therefore there exists $K>0$ such that for all $n \geq 1$,

$$\mathbb{E}[(X_1 + ... + X_n)^4] \leq Kn^2.$$

Proposition 4. *For any $0<p_1<1$, $0<p_{cont}<1$ and for any sequence $(t_1,...,t_k)$,*

$$\left(\frac{S_{[nt_1]}}{\sqrt{n}},...,\frac{S_{[nt_k]}}{\sqrt{n}}\right) \overset{n\to\infty}{\Rightarrow} (\delta\mathcal{N}(0,t_1),...,\delta\mathcal{N}(0,t_k)).$$

Proof. First we will prove that

$$\frac{X_1 + ... + X_n}{\sqrt{n}} \overset{n\to\infty}{\Rightarrow} \delta\mathcal{N}(0,1).$$

Since the sequence $(X_1, X_2, ...)$ is neither stationary nor composed of independent random variables, the central limit theorem cannot be directly applied. Let, for $n \geq 1$,

$$Z_n = \frac{X_1 + ... + X_n}{\sqrt{n}},$$

and let $\Psi_n(k)E[\exp(ikZ_n)]$. Since the moments of Z_n are finite, we have:

$$\Psi_n(k) = 1 + \sum_{m=1}^{\infty}\frac{(-ik)^m}{m!}\mu_n^m, \qquad \text{where} \qquad \mu_n^m = \mathbb{E}[\frac{(X_1+X_2+...+X_n)^m}{n^{m/2}}],$$

is the m-th moment of Z_n. To prove the convergence of Z_n toward a Gaussian distribution, it is sufficient to prove that for all $k \in \mathbb{R}$,

$$\lim_{n\to\infty}\Psi_n(k) = e^{-\frac{\delta^2 k^2}{2}} = \sum_{m=0}^{\infty}\frac{(-1)^m k^{2m}\delta^{2m}}{2^m m!},$$

therefore it is enough to prove that for all $p \geq 0$,

$$\lim_{n\to\infty}\mu_n^{2p} = \frac{\delta^{2p}(2p)!}{2^p p!}, \qquad \text{and} \qquad \lim_{n\to\infty}\mu_n^{2p+1} = 0,$$

which was proven in lemma 3. Now, let $(t_1,...,t_k)$ an sequence of real numbers. We assume without loss of generality that this sequence is non-decreasing. We will prove that

$$\left(\frac{S_{[nt_1]}}{\sqrt{n}},...,\frac{S_{[nt_k]}}{\sqrt{n}}\right) \overset{n\to\infty}{\Rightarrow} (\delta\mathcal{N}(0,t_1),...,\delta\mathcal{N}(0,t_k)).$$

Using the Cramer-Wold device, it is enough to prove that for all sequence of real numbers $(a_1, ..., a_k)$,

$$\frac{a_1 S_{[nt_1]} + a_2 S_{[nt_2]} + ... + a_k S_{[nt_k]}}{\sqrt{n}} \overset{n\to\infty}{\Rightarrow} \delta\mathcal{N}(0, a_1^2 t_1 + ... + a_k^2 t_k).$$

Let $(a_1, ..., a_k)$ be a sequence of real numbers and let for all $n \geq 1$,

$$U_n = \frac{a_1 S_{[nt_1]} + a_2 S_{[nt_2]} + ... + a_k S_{[nt_k]}}{\sqrt{n}}.$$

We will prove that is the Fourier transform of U_n converges to the Laplace transform of $\delta\mathcal{N}(0, a_1^2 t_1 + ... + a_k^2 t_k)$. Let Ψ_n be the Laplace transform of U_n:

$$\forall k \geq 1, \Psi_n(k) = 1 + \sum_{m=1}^{\infty} \frac{(-ik)^m}{m!} \nu_n^m,$$

where

$$\nu_n^m = \mathbb{E}[U_n^m].$$

Let for all $i \geq 1$,

$$Y_i := a_j X_i, \quad for \quad [nt_{j-1}] < i \leq [nt_j].$$

We use the same method as the one described in lemma 3:

$$\nu_n^{2p} = \frac{1}{n^p} \sum_{k_1+...+k_m=2p} \frac{\delta^{2p}(2p)!}{k_1!...k_m!} \mathbb{E}[Y_{k_1}^{k_1} ... Y_{k_m}^{k_m}];$$

The only term that do not vanish when n goes to infinity, are the terms corresponding to $k_1 = k_2 = ... = k_p = 2$, and when we group all the terms we obtain from the multinomial formula:

$$\nu_n^{2p} \to_{n\to\infty} \frac{(a_1^2 t_1 + ... + a_k^2 t_k)^p (2p)!}{2^p p!}.$$

Using the same method we have

$$\nu_n^{2p+1} \overset{n\to\infty}{\Rightarrow} 0.$$

Therefore

$$\frac{a_1 S_{[nt_1]} + a_2 S_{[nt_2]} + ... + a_k S_{[nt_k]}}{\sqrt{n}} \overset{n\to\infty}{\Rightarrow} \delta\mathcal{N}(0, a_1^2 t_1 + ... + a_k^2 t_k).$$

Since this holds true for any sequence $(a_1, ..., a_k)$,

$$\left(\frac{S_{[nt_1]}}{\sqrt{n}}, ..., \frac{S_{[nt_k]}}{\sqrt{n}}\right) \overset{n\to\infty}{\Rightarrow} (\delta\mathcal{N}(0, t_1), ..., \delta\mathcal{N}(0, t_k)).$$

Proposition 5. *The sequence of processes $(\xi^n)_{n\geq 1}$ defined as:*

$$\xi_t^n = \frac{1}{\sqrt{n}} \sum_{k=1}^{j} X_k + \frac{nt - j}{\sqrt{n}} X_{j+1}, \quad \frac{j}{n} \leq t < \frac{j+1}{n}$$

converges in distribution to a Brownian motion B on $\mathcal{C}(\mathbb{R}_+, \mathbb{R})$:

$$\xi^n \overset{n\to\infty}{\Rightarrow} B \quad on \quad \mathcal{C}(\mathbb{R}_+, \mathbb{R})$$

Proof. We proved in Proposition 4 the convergence of all finite dimensional distributions of ξ^n to those of a standard Brownian motion. In order to prove the theorem, we need to prove that the sequence of processes $(\xi^n, n \geq 1)$ is tight on $\mathcal{C}(\mathbb{R}_+, \mathbb{R})$ According to the Lamperti criteria from Lamperti (1962), if there exists $M > 0$ such that for all $(s,t) \in \mathbb{R}_+^2$,

$$\mathbb{E}[|\xi_n(t) - \xi_n(s)|^4] \leq M|t-s|^2,$$

then the sequence is tight on $\mathcal{C}(\mathbb{R}_+, \mathbb{R})$.

First case: if $\frac{j}{n} \leq s \leq t \leq \frac{j+1}{n}$, then

$$|\xi_n(t) - \xi_n(s)| = \frac{1}{\sqrt{n}} n(t-s) X_{j+1}$$

$$|\xi_n(t) - \xi_n(s)| \leq \sqrt{n}|(t-s)X_{j+1}|.$$

Since $n(t-s) \leq 1$ and $\mathbb{E}[X_{j+1}^4] = 1$,

$$\mathbb{E}[|\xi_n(t) - \xi_n(s)|^4] \leq |t-s|^2$$

If $\frac{j-1}{n} \leq s \leq \frac{j}{n} \leq \frac{j+k}{n} \leq t \leq \frac{j+k+1}{n}$, then

$$|\xi_n(t) - \xi_n(s)| \leq |\xi_n(t) - \xi_n(\frac{j+k}{n})| + |\xi_n(\frac{j+k}{n} - \xi_n(\frac{j}{n}))| + |\xi_n(\frac{j}{n}) - \xi_n(s)|.$$

By Jensen's inequality,

$$\mathbb{E}[|\xi_n(t) - \xi_n(s)|^4] \leq 27 \left(\mathbb{E}[|\xi_n(t) - \xi_n(\frac{j+k}{n})|^4] + \mathbb{E}[|\xi_n(\frac{j+k}{n}) - \xi_n(\frac{j}{n}))|^4] + \mathbb{E}[|\xi_n(\frac{j}{n}) - \xi_n(s)|^4] \right).$$

The first term and the last term can be treated as in the previous case. For the term in the middle,

$$\mathbb{E}[|\xi_n(\frac{j+k}{n}) - \xi_n \frac{j}{n}|^4] \leq (\sqrt{n})^4 \mathbb{E}[(X_{j+1} + ... + X_{j+k})^4] \leq (\frac{1}{\sqrt{n}})^4 K k^2 \leq K|t-s|^2$$

because $|t-s| \geq \frac{j+k}{n} - \frac{j}{n} = \frac{k}{n}$ and $\mathbb{E}[(X_1 + ... + X_n)^4] \leq Kn^2$. Therefore the sequence $(\xi_n)_{n \geq 0}$ is tight. Finally the sequence $(\xi^n, n \geq 1)$ converges weakly to the Brownian motion.

Lemma 5 *If $\lambda = \theta + \mu$, and f is a probability distribution on $\mathbb{N}^2$ such that*

$$m(f) = \sum_{i=1}^{\infty} \sum_{j=i}^{\infty} ijf(i,j) < \infty,$$

then

$$\frac{\tau_1 + \tau_2 + ... + \tau_k}{k \log k} \overset{k \to \infty}{\Rightarrow} \frac{m(f)}{\pi \lambda}.$$

Proof. The sequence $(\tau_2, \tau_3, ...)$ is a sequence of i.i.d, regularly varying random variables, with tail index equal to 1. Let $\mathcal{L}(s)$ the Laplace transform of τ_2 and $F(t) = \mathbb{P}[\tau_2 > t]$:

$$\mathcal{L}(s) = 1 - s \int_0^{\infty} \exp(-st) F(t) dt.$$

From the expression of the conditional distribution of τ, we obtain that $F(t) \sim m(f)/\pi\lambda t$ as $t \to \infty$. Since

$$\frac{1}{\log(n)} \int_0^n \exp(-st/n\log(n))F(t)dt \overset{n\to\infty}{\to} \frac{m(f)}{\pi\lambda}$$

and, using integration by parts,

$$\frac{1}{\log(n)} \int_n^\infty \exp(-st/n\log(n))F(t)dt \overset{n\to\infty}{\to} 0.$$

Therefore

$$\frac{s}{n\log(n)} \int_0^\infty \exp(-\frac{st}{n\log n})F(t)dt = \frac{m(f)}{\pi\lambda}\frac{s}{n} + o(\frac{1}{n});$$

This implies that

$$(\mathcal{L}(\frac{s}{n\log n}))^n = (1 + \frac{m(f)}{\pi\lambda} + o(\frac{1}{n}))^n \to \exp(-\frac{m(f)}{\pi\lambda});$$

and

$$\frac{\tau_1 + \tau_2 + ... + \tau_n}{n\log n} \overset{n\to\infty}{\Rightarrow} \frac{m(f)}{\pi\lambda}$$

Acknowledgments

The authors thank Peter Carr, Xin Guo, Charles Lehalle, Pete Kyle, Arseniy Kukanov, Costis Maglaras and seminar participants at the European Meeting of Statisticians (Athens, August 2010), Morgan Stanley, INFORMS 2010, the SIAM Conference on Financial Engineering and the Conference on Market microstructure (Paris, 2010) for helpful comments and discussions.

References

Alfonsi, A., A. Schied, A. Schulz. 2010. Optimal execution strategies in limit order books with general shape function. *Quantitative Finance* **10** 143–157.

Biais, Bruno, Pierre Hillion, Chester Spatt. 1995. An empirical analysis of the order flow and order book in the paris bourse. *Journal of Finance* **50**(5) 1655–1689.

Billingsley, P. 1968. *Convergence of Probability Measures*. Wiley Series in Probability and Statistics.

Bouchaud, Jean-Philippe, Doyne Farmer, Fabrizio Lillo. 2008. How markets slowly digest changes in supply and demand. T. Hens, K. Schenk-Hoppe, eds., *Handbook of Financial Markets: Dynamics and Evolution*. Elsevier: Academic Press.

Cohen, Jacob Willem, O. J. Boxma. 1983. *Boundary value problems in queueing system analysis*, *North-Holland Mathematics Studies*, vol. 79. North-Holland Publishing Co., Amsterdam.

Cont, Rama. 2001. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance* **1**(2) 223–236.

Cont, Rama, Adrien de Larrard. 2010. Linking volatility with order flow: heavy traffic approximations and diffusion limits of order book dynamics. Working paper.

Cont, Rama, Arseniy Kukanov, Sasha Stoikov. 2010a. The price impact of order book events. Working Paper http://ssrn.com/abstract=1712822, SSRN.

Cont, Rama, Sasha Stoikov, Rishi Talreja. 2010b. A stochastic model for order book dynamics. *Operations Research* **58** 549–563.

Engle, Robert F., Asger Lunde. 2003. Trades and Quotes: A Bivariate Point Process. *Journal of Financial Econometrics* **1**(2) 159–188.

Farmer, J. Doyne, László Gillemot, Fabrizio Lillo, Szabolcs Mike, Anindya Sen. 2004. What really causes large price changes? *Quantitative Finance* **4** 383–397.

Feller, William. 1971. *An Introduction to Probability Theory and Its Applications, Vol. 2*. John Wiley and Sons.

Gourieroux, Christian, Joanna Jasiak, Gaelle Le Fol. 1999. Intra-day market activity. *Journal of Financial Markets* **2**(3) 193–226.

Harris, Lawrence E., Venkatesh Panchapagesan. 2005. The information content of the limit order book: evidence from NYSE specialist trading decisions. *Journal of Financial Markets* **8**(1) 25–67.

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

Lamperti, J. 1962. On convergence of stochastic processes. *American Mathematical Society* **104** 430–435.

Lawler, Gregory F., Vlada Limic. 2010. *Random Walk: A Modern Introduction*, vol. 123. Cambridge Studies in Advanced Mathematics.

Parlour, Christine. 1998. Price dynamics in limit order markets. *Review of Financial Studies* **11**(4) 789–816.

Predoiu, S., G. Shaikhet, S. Shreve. 2010. Optimal execution of a general one-sided limit-order book. Working paper.

Rosu, Ioanid. 2009. A dynamic model of the limit order book. *Review of Financial Studies* **22** 4601–4641.

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

Whitt, Ward. 2002. *Stochastic Process Limits*. Springer Verlag.