Order book dynamics in liquid markets: limit theorems and diffusion approximations

Rama Cont, Adrien de Larrard

Order book dynamics in liquid markets: limit theorems and diffusion approximations

Rama CONT & Adrien de LARRARD
Laboratoire de Probabilités et Modèles Aléatoires
CNRS - Université Pierre et Marie Curie (Paris VI)

August 2011. Revised : Feb 2012

We propose a model for the dynamics of a limit order book in a liquid market where buy and sell orders are submitted at high frequency. We derive a functional central limit theorem for the joint dynamics of the bid and ask queues and show that, when the frequency of order arrivals is large, the intraday dynamics of the limit order book may be approximated by a Markovian jump-diffusion process in the positive orthant, whose characteristics are explicitly described in terms of the statistical properties of the underlying order flow. This result allows to obtain tractable analytical approximations for various quantities of interest, such as the probability of a price increase or the distribution of the duration until the next price move, conditional on the state of the order book. Our results allow for a wide range of distributional assumptions and temporal dependence in the order flow and apply to a wide class of stochastic models proposed for order book dynamics, including models based on Poisson point processes, self-exciting point processes and models of the ACD-GARCH family.

Key words: limit order book, queueing systems, heavy traffic limit, functional central limit theorem, diffusion limit, high-frequency data, market microstructure, point process, limit order market

Contents

1. Introduction

An increasing proportion of financial transactions -in stocks, futures and other contracts- take place in electronic markets where participants may submit limit orders (for buying or selling), market orders and order cancelations which are then centralized in a *limit order book* and executed according to precise time and price priority rules. The limit order book represents, at each point in time, the outstanding orders which are awaiting execution: it consists in queues at different price levels where these orders are arranged according to time of arrival. A new limit buy (resp. sell) order of size x increases the size of the bid (resp. ask) queue by x. Market orders are executed against limit orders at the best available price: a market order decreases of size x the corresponding queue size by x. Limit orders placed at the best available price are executed against market orders.

The availability of high-frequency data on limit order books has generated a lot of interest in statistical modeling of order book dynamics, motivated either by high-frequency trading applications or simply a better understanding of intraday price dynamics (see Cont (2011) for a recent survey). The challenge here is to develop statistical models which capture salient features of the data while allowing for some analytical and computational tractability.

Given the discrete nature of order submissions and precise priority rules for their execution, is quite natural to model a limit order book as a queueing system; early work in this direction dates back to Mendelson (1982). More recently, Cont, Stoikov and Talreja Cont et al. (2010b) have studied a Markovian queueing model of a limit order book, in which arrivals of market orders and limit orders at each price level are modeled as independent Poisson processes. Cont and de Larrard (2010) used this Markovian queueing approach to compute useful quantities (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) and relate the volatility of the price with statistical properties of the order flow.

However, the results obtained in such Markovian models rely on the fact that time intervals between orders are independent and exponentially distributed, orders are of the same size and that the order flow at the bid is independent from the order flow at the ask. Empirical studies on high-frequency data show these assumptions to be incorrect (Hasbrouck (2007), Bouchaud et al. (2002, 2008), Andersen et al. (2010)). Figure 1 compares the quantiles of the duration between order book events for CitiGroup stock on June 26, 2008 to those of an exponential distribution with the same mean, showing that the empirical distribution of durations is far from being exponential. Figure 9 shows the autocorrelation function of the inverse durations: the persistent positive value of this autocorrelation shows that durations may not be assumed to be independent. Finally, as shown in Figure 2 which displays the (positive or negative) changes in queue size induced by successive orders for CitiGroup shares, there is considerable heterogeneity in sizes and clustering in the timing of orders.

Other, more complex, statistical models for order book dynamics have been developed to take these properties into account (see Section 2.3). However, only models based on Poisson point processes such as Cont et al. (2010b), Cont and de Larrard (2010) have offered so far the analytical tractability necessary when it comes to studying quantities of interest such as durations or transition probabilities of the price, conditional on the state of the order book. It is therefore of interest to know whether the conclusions based on Markovian models are robust to a departure from these simplifying assumptions and, if not, how they must be modified in the presence of other distributional features and dependence in durations and order sizes.

The goal of this work is to show that it is indeed possible to restore analytical tractability without imposing restrictive assumptions on the order arrival process, by exploiting the *separation of time scales* involved in the problem. The existence of widely different time scales, from milliseconds to minutes, makes it possible to obtain meaningful results from an asymptotic analysis of order book dynamics using a *diffusion approximation* of the limit order book. We argue that this diffusion approximation provides relevant and computationally tractable approximations of the quantities of interest in liquid markets where order arrivals are frequent.

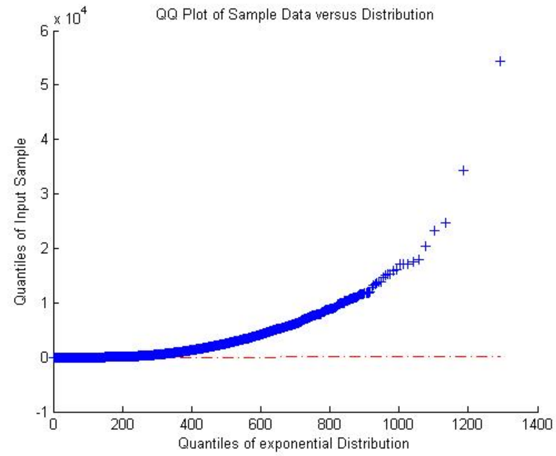

Figure 1 Quantiles of inter-event durations compared with quantiles of an exponential distribution with the same mean (Citigroup, June 2008). The dotted line represents the benchmark case where the observations are exponentially distributed, which is clearly not the case here.

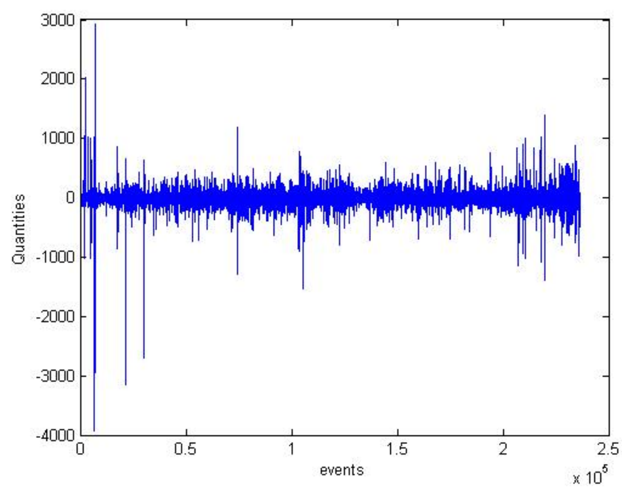

Figure 2 Number of shares per event for events affecting the ask. The stock is Citigroup on the 26th of June 2008

As shown in Table 1, most applications involve the behavior of prices over time scales an order of magnitude larger than the typical inter-event duration: for example, in optimal trade execution the benchmark is the Volume weighted average price (VWAP) computed over a period which may range from 10 minutes to a day: over such time scales much of the microstructural details of the market are averaged out. Second, as noted in Table 2, in liquid equity markets the number of events affecting the state of the order book over such time scales is quite large, of the order of hundreds or thousands. The typical duration τ_L (resp. τ_M) between limit orders (resp. market orders and cancelations) is typically $0.001 - 0.01 \ll 1$ (in seconds). These observations show that it is relevant to consider *heavy-traffic limits* in which the rate of arrival of orders is large for studying the dynamics of order books in liquid markets.

Regime	Time scale	Issues
Ultra-high frequency (UHF)	$\sim 10^{-3} - 0.1$ s	Microstructure, Latency
High Frequency (HF)	$\sim 1 - 100$ s	Trade execution
"Daily"	$\sim 10^{3} - 10^{4}$ s	Trading strategies, Option hedging

Table 1 A hierarchy of time scales.

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

Table 2 Average number of orders in 10 seconds and number of price changes (June 26th, 2008).

In this limit, the complex dynamics of the discrete queueing system is approximated by a simpler system with a continuous state space, which can be either described by a system of ordinary differential equations (in the 'fluid limit', where random fluctuations in queue size vanish) or a system of stochastic differential equations (in the 'diffusion limit' where random fluctuations dominate) (Iglehart and Whitt (1970), Harrison and Nguyen (1993), Whitt (2002)). Intuitively, the fluid limit corresponds to the regime of law of large numbers, where random fluctuations average out and the limit is described by average queue size, whereas the diffusion limit corresponds to the regime of the central limit theorem, where fluctuations in queue size are asymptotically Gaussian. When order sizes or durations fail to have finite moments of first or second order, other scaling limits may intervene, involving Lévy processes (see Whitt (2002)) or fractional Brownian motion Araman and Glynn (2011). As shown by Dai and Nguyen (1994), there are also cases where such a 'heavy traffic limit' may fail to exist. The relevance of each of these asymptotic regimes is, of course, not a matter of 'taste' but an empirical question which depends on the behavior of high-frequency order flow in these markets.

Using empirical data on US stocks, we argue that for most liquid stocks, while the rate of arrival of market orders and limit orders is large, the *imbalance* between limit orders, which increase queue size, and market orders and cancels, which decrease queue size, is an order of magnitude smaller: over, say, a 10 minute interval, one observes an imbalance ranging from 1 to 10 % of order flow. In other words, over a time scale of several minutes, a large number N of events occur, but the bid/ask imbalance accumulating over the same interval is of order $\sqrt{N} \ll N$. In this regime, random fluctuations in queue sizes cannot be ignored and it is relevant to consider the *diffusion limit* of the limit order book.

In this paper we study the behavior of a limit order book in this diffusion limit: we prove a functional central limit theorem for the joint dynamics of the bid and ask queues when the intensity of orders becomes large, and use it to derive an analytically tractable jump-diffusion approximation. More precisely, we show that under a wide range of assumptions, which are shown to be plausible for empirical data on liquid US stocks, the intraday dynamics of the limit order book behaves like as a planar Brownian motion in the interior of the positive orthant, and jumps to the interior of the orthant at each hitting time of the boundary.

This jump-diffusion approximation allows various quantities of interest to be computed analytically: we obtain analytical expressions for various quantities such as the probability that the price will increase at the next price change, and the distribution of the duration between price changes, conditional on the state of the order book.

Our results extend previous analysis of heavy traffic limits for such auction processes (Kruk (2003), Bayraktar et al. (2006), Cont and de Larrard (2010)) to a setting which is relevant and useful for quantitative modeling of limit order books and provide a foundation for recently proposed diffusion models for order book dynamics Avellaneda et al. (2011).

Outline. The paper is organized as follows. Section 2 describes a general framework for the dynamics of a limit order book; various examples of models studied in the literature are shown to fall within this modeling framework (Section 2.3). Section 3 reviews some statistical properties of high frequency order flow in limit order markets: these properties highlight the complex nature of the order flow and motivate the statistical assumptions used to derive the diffusion limit. Section 4 contains our main result: Theorem 2 shows that, in a limit order market where orders arrive at high frequency, the bid and ask queues behaves like a Markov process in the positive quadrant which diffuses inside the quadrant and jumps to the interior each time it hits the boundary. We provide a complete description of this process, and use it to derive, in Section 4.3, a simple jump-diffusion approximation for the joint dynamics of bid and ask queues, which is easier to study and simulate than the initial queueing system.

In particular, we show that in this asymptotic regime the price process is characterized as a piecewise constant process whose transition times correspond to hitting times of the axes by a two dimensional Brownian motion in the positive orthant (Proposition 1). This result allows to study analytically various quantities of interest, such as the distribution of the duration between price moves and the probability of an increase in the price: this is discussed in Section 5.

2. A model for the dynamics of a limit order book

2.1. Reduced-form 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)). All electronic trading venues also allow to place limit orders *pegged* to the best available price (National Best Bid Offer, or NBBO); market makers used these *pegged* orders to liquidate their inventories. 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 of the consolidated order book are more easily obtainable (from records on trades and quotes) than information on deeper levels of the order book, 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 3 summarizes this representation.

If the stock is traded in several venues, the quantities q^b and q^a represent the best bids and offers in the *consolidated* order book, obtained by aggregating over all (visible) trading venues. At every time t, q_t^b (resp. q_t^a) corresponds to all visible orders available at the bid price s_t^b (resp. s_t^a) across all exchanges.

The state 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.

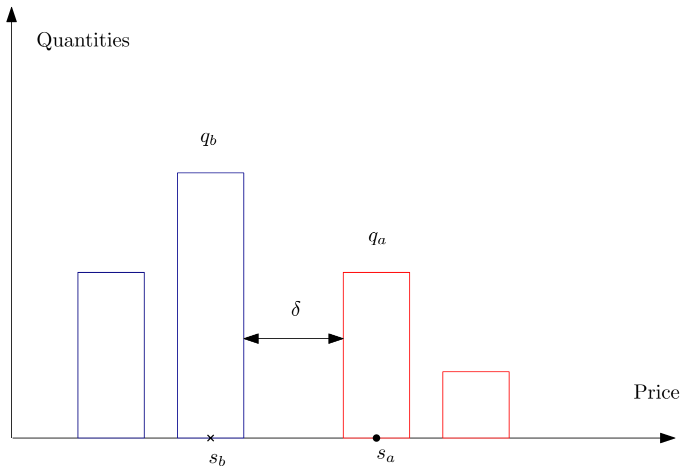

Figure 3 Simplified representation of a limit order book.

sell) order decreases the corresponding queue size by x. Cancelation 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 cancelations have the same effect on the queue sizes (q_t^b, q_t^a).

The bid and ask prices are multiples of the tick size δ. When either the bid or ask queue is depleted by market orders and cancelations, the price moves up or down to the next level of the order book. The price processes s_t^b, s_t^a are thus piecewise constant processes whose transitions correspond to hitting times of the axes $\{(0,y), y > 0\} \cup \{(x,0), x > 0\}$ by the process $q_t = (q_t^b, q_t^a)$.

If the order book contains no 'gaps' (empty levels), these price increments are equal to one tick:

- when the bid queue is depleted, the (bid) price decreases by one tick.
- when the ask queue is depleted, the (ask) 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 dynamics. We will ignore this feature in what follows but it is not hard to generalize our results to include it.

The quantity $s_t^a - s_t^b$ is the *bid-ask spread*, which may be one or several ticks. As shown in Table 3, for liquid stocks the bid-ask spread is equal to one tick for more than 98% of observations.

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

When either the bid or ask queue is depleted, the bid-ask spread widens immediately to more than one tick. Once the spread has increased, a flow of limit sell (resp. buy) orders quickly fills the gap and the spread reduces again to one tick. When a limit order is placed inside the spread, all the limit orders pegged to the NBBO price move in less than a millisecond to the price level corresponding to this new order. Once this happens, both the bid price and the ask price have increased (resp. decreased) by one tick.

The histograms in Figure 4 show that this 'closing' of the spread takes place very quickly: as shown in Figure 4 (left) the lifetime of a spread larger than one tick is of the order of a couple of milliseconds, which is negligible compared to the lifetime of a spread equal to one tick (Figure 4 , right). In our model we assume that the second step occurs *infinitely fast*: once the bid-ask spread widens, it returns immediately to one tick. For the example of Dow Jones stocks (Figure 4), this is a reasonable assumption since the widening of the spread lasts only a few milliseconds. This simply means that we are not trying to describe/model how the orders flow inside the bid-ask spread at the millisecond time scale and, when we describe the state of the order book *after* a price change we have in mind the state of the order book *after* the bid-ask spread has returned to one tick.

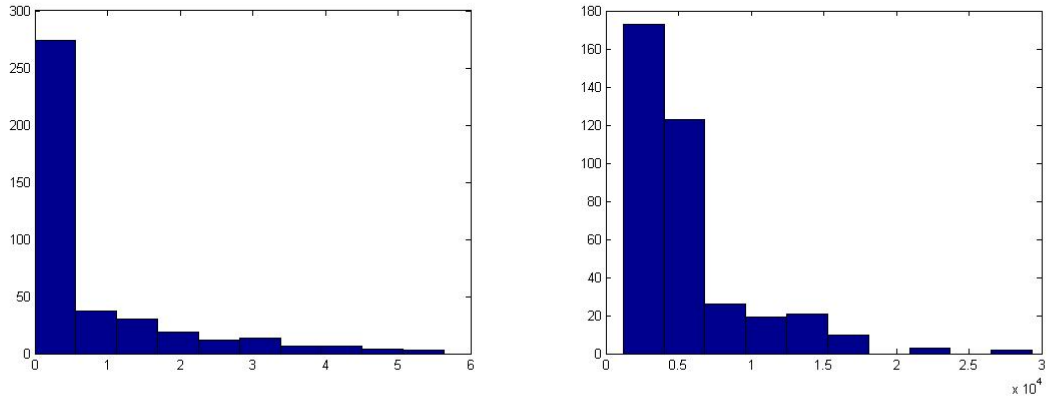

Figure 4 Left: Average lifetime, in milliseconds of a spread larger than one tick for Dow Jones stocks. Right: Average lifetime, in milliseconds of a spread equal to one tick.

Under this assumption, each time *one* of the queues is depleted, *both* the bid queue and the ask queue move to a new position and the bid-ask spread remains equal to one tick after the price change. Thus, under our assumptions the bid-ask 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.

Once either the bid or the ask queue are depleted, the bid and ask queues assume new values. Instead of keeping track of arrival, cancelation and execution of orders at *all* price levels (as in Cont et al. (2010b), Smith et al. (2003)), we treat the queue sizes after a price change as a stationary sequence of random variables whose distribution represents the depth of the order book in a statistical sense. More specifically, we model the size of the bid and ask queues after a price increase by a stationary sequence $(R_k)_{k\geq 1}$ of random variables with values in $\mathbb{N}^2$. Similarly, the size of the bid and ask queues after a price decrease is modeled by a stationary sequence $(\tilde{R}_k)_{k\geq 1}$ of random variables with values in $\mathbb{N}^2$. The sequences $(R_k)_{k\geq 1}$ and $(\tilde{R}_k)_{k\geq 1}$ 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.

The variables R_k (resp. $\tilde{R}_k$) have a common distribution which represents the *depth* of the order book after a price increase (resp. decrease): Figure 5 shows the (joint) empirical distribution of bid and ask queue sizes after a price move for Citigroup stock on June 26th 2008.

The simplest specification could be to take $(R_k)_{k\geq 1}$, $(\tilde{R}_k)_{k\geq 1}$ to be IID sequences; this approach, used in Cont and de Larrard (2010), turns out to be good enough for many purposes. But this IID assumption is not necessary; in the next section we will see more general specifications which allow for serial dependence.

In summary, state of the limit order book is thus described by a continuous-time process (s_t^b, q_t^b, q_t^a) which takes values in the discrete state space $\delta\mathbb{Z} \times \mathbb{N}^2$, with piecewise constant sample paths whose transitions correspond to the order book events. Denoting by $(t_i^a, i \geq 1)$ (resp. t_i^b) the event times at the ask (resp. the bid), V_i^a (resp. V_i^b) the corresponding change in ask (resp. bid) queue size, and $k(t)$ the number of price changes in $[0, t]$, the above assumptions translate into the following dynamics for (s_t^b, q_t^b, q_t^a):

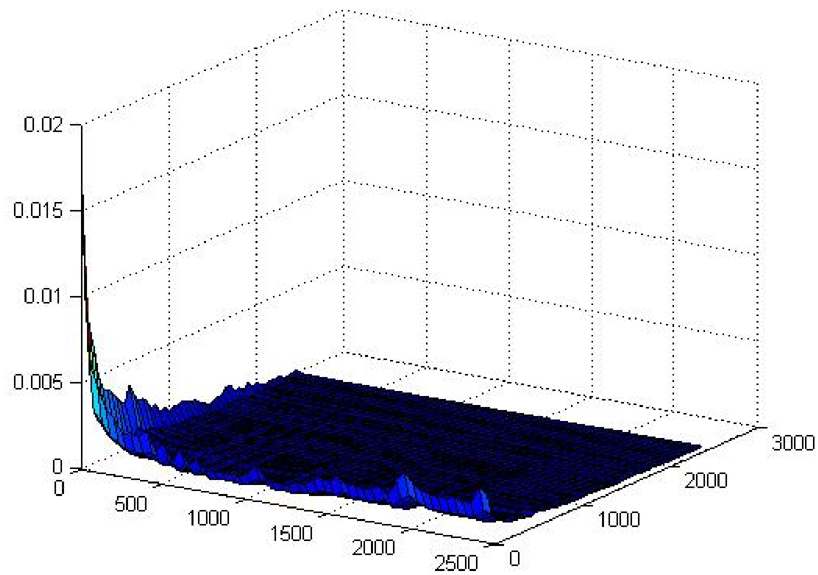

Figure 5 Joint density of bid and ask queue sizes after a price move (Citigroup, June 26th 2008).

- If an order or cancelation of size V_i^a arrives on the ask side at $t = t_i^a$,
  —if $q_{t-}^a + V_i^a \geq 0$, the order can be satisfied without changing the price;
  —if $q_{t-}^a + V_i^a < 0$, the ask queue is depleted, the price increases by one 'tick' of size δ, and the queue sizes take new values $R_{k(t)} = (R_{k(t)}^b, R_{k(t)}^a)$,

$$(s_t^b, q_t^b, q_t^a) = (s_{t-}^b, q_{t-}^b, q_{t-}^a + V_i^a) 1_{\{q_{t-}^a \geq -V_i^a\}} + (s_{t-}^b + \delta, R_{k(t)}^b, R_{k(t)}^a) 1_{\{q_{t-}^a < -V_i^a\}}, \tag{1}$$

- If an order or cancelation of size V_i^b arrives on the bid side at $t = t_i^b$,
  —if $q_{t-}^b + V_i^b \geq 0$, the order can be satisfied without changing the price;
  —if $q_{t-}^b + V_i^b < 0$, the bid queue gets depleted, the price decreases by one 'tick' of size δ and the queue sizes take new values $\tilde{R}_{k(t)} = (\tilde{R}_{k(t)}^b, \tilde{R}_{k(t)}^a)$:

$$(s_t^b, q_t^b, q_t^a) = (s_{t-}^b, q_{t-}^b + V_i^b, q_{t-}^a) 1_{\{q_{t-}^b \geq -V_i^b\}} + (s_{t-}^b - \delta, \tilde{R}_{k(t)}^b, \tilde{R}_{k(t)}^a) 1_{\{q_{t-}^b < -V_i^b\}}. \tag{2}$$

2.2. The limit order book as a 'regulated' process in the orthant

As in the case of reflected processes arising in queueing networks, the process $q_t = (q_t^b, q_t^a)$ may be constructed from the **net order flow process**

$$x_t = (x_t^b, x_t^a) = \left(\sum_{i=1}^{N_t^b} V_i^b, \sum_{i=1}^{N_t^a} V_i^a \right)$$

where N_t^b (resp. N_t^a) is the number of events (i.e. orders or cancelations) occurring at the bid (resp. the ask) during $[0, t]$. $x_t = (x_t^b, x_t^a)$ is analogous to the 'net input' process in queuing systems Whitt (2002): x_t^b (resp. x_t^a) represents the cumulative sum of all orders and cancelations at the bid (resp. the ask) between 0 and t.

$q = (q_t^b, q_t^a)_{t\geq 0}$ which takes values in the positive orthant, may be constructed from x by reinitializing its value to a a new position inside the positive orthant according to the rules (1)–(2) each time one of the queues is depleted: every time $(q_t)_{t\geq 0}$ attempts to exit the positive orthant, it jumps to a a new position inside the orthant, taken from the sequence $(R_n, \tilde{R}_n)$.

This construction may be done path by path, as follows:

DEFINITION 1. Let $\omega \in D([0,\infty),\mathbb{R}^2)$ be a right-continuous function with left limits (i.e. a cadlag function), $R=(R_n)_{n\geq 1}$ and $\tilde{R}=(\tilde{R}_n)_{n\geq 1}$ two sequences with values in $\mathbb{R}_+^2$. There exists a unique cadlag function $\Psi(\omega,R,\tilde{R}) \in D([0,\infty),\mathbb{R}_+^2)$ such that

- For $t<\tau_1$, let $\Psi(\omega,R,\tilde{R})(t)=\omega(t)$ where

$$\tau_1 = \inf\{t \geq 0,\ \omega(t).(1,0)<0 \text{ or } \omega(t).(0,1)<0\}.$$

is the first exit time of ω from the positive orthant.

- $\Psi(\omega,R,\tilde{R})(\tau_1) = R_1$ if $\Psi(\omega,R,\tilde{R})(\tau_1-).(0,1) < 0$, and $\Psi(\omega,R,\tilde{R})(\tau_1) = \tilde{R}_1$ if $\Psi(\omega,R,\tilde{R})(\tau_1-).(1,0)<0$.
- For $k\geq 1$,

$\Psi(\omega,R,\tilde{R})(t+\tau_k)=\Psi(\omega,R,\tilde{R})(\tau_k)+\omega(t+\tau_k)-\omega(\tau_k)$ for $0\leq t<\tau_{k+1}-\tau_k$, where

$$\tau_{k+1} = \inf\{t \geq \tau_{k-1},\ \Psi(\omega,R,\tilde{R})(\tau_k)+\omega(t+\tau_k)-\omega(\tau_k) \notin \mathbb{R}_+^2\}$$

is the first exit time of $(\Psi(\omega,R,\tilde{R})(t), t\geq\tau_k)$ from the positive orthant.

- $\Psi(\omega,R,\tilde{R})(\tau_k)=R_k$ if $\Psi(\omega,R,\tilde{R})(\tau_k-).(0,1)<\Psi(\omega,R,\tilde{R})(\tau_{k-1}).(1,0)$ and $\Psi(\omega,R,\tilde{R})(\tau_k)=\tilde{R}_k$ otherwise.

The path $\Psi(\omega,R,\tilde{R})$ is obtained by "regulating" the path ω with the sequences $(R,\tilde{R})$: in between two exit times, the increments of $\Psi(\omega,R,\tilde{R})$ follow those of ω and each time the process attempts to exit the positive orthant by crossing the x-axis (resp. the y-axis), it jumps to a a new position inside the orthant, taken from the sequence $(R_n)_{n\geq 1}$ (resp. from the sequence $(\tilde{R}_n)_{n\geq 1}$).

Unlike the more familiar case of a continuous reflection at the boundary, which arises in heavy-traffic limits of multiclass queueing systems (see Harrison (1978), Harrison and Nguyen (1993), Whitt (2002), Ramanan and Reiman (2003) for examples), this construction introduces a discontinuity by pushing the process into the interior of the positive orthant each time it attempts to exit from the axes.

To study the continuity properties of this map, we endow $D([0,\infty),\mathbb{R}^2)$ with Skorokhod's J_1 topology Billingsley (1968), Lindvall (1973) and the set $(\mathbb{R}_+^2)^{\mathbb{N}}$ with the topology induced by 'cylindrical' semi-norms, defined as follows: for a sequence $(R^n)_{n\geq 1}$ in $(\mathbb{R}_+^2)^{\mathbb{N}}$

$$R^n \overset{n\to\infty}{\to} R \in (\mathbb{R}_+^2)^{\mathbb{N}} \quad \Longleftrightarrow \Big(\forall k\geq 1, \quad \sup\{|R_1^n-R_1|,...,|R_k^n-R_k|)\overset{n\to\infty}{\to} 0\Big).$$

$D([0,\infty),\mathbb{R}^2)\times(\mathbb{R}_+^2)^{\mathbb{N}}\times(\mathbb{R}_+^2)^{\mathbb{N}}$ is then endowed with the corresponding product topology.

THEOREM 1. *Let $R=(R_n)_{n\geq 1}, \tilde{R}=(\tilde{R}_n)_{n\geq 1}$ be sequences in $]0,\infty[\times]0,\infty[$ which do not have any accumulation point on the axes. If $\omega\in C^0([0,\infty),\mathbb{R}^2)$ is such that*

$$(0,0)\notin\Psi(\omega,R,\tilde{R})([0,\infty)\ ). \tag{3}$$

Then the map

$$\Psi : D([0,\infty),\mathbb{R}^2)\times(\mathbb{R}_+^2)^{\mathbb{N}}\times(\mathbb{R}_+^2)^{\mathbb{N}} \to D([0,\infty),\mathbb{R}_+^2) \tag{4}$$

is continuous at $(\omega,R,\tilde{R})$.

Proof: See Section 6.2 in the Appendix.

This construction may be applied to any cadlag stochastic process: given a cadlag process X with values in $\mathbb{R}^2$ and (random) sequences $R=(R_n)_{n\geq 1}$ and $\tilde{R}=(\tilde{R}_n)_{n\geq 1}$ with values in $\mathbb{R}_+^2$, the process $\Psi(X,R,\tilde{R})$ is a cadlag process with values in $\mathbb{R}_+^2$.

It is easy to see that the order book process $q_t=(q_t^b,q_t^a)$ may be constructed by this procedure:

Lemma 1. *The queue size process $q = (q_t^b, q_t^a)_{tgeq0}$ is related to the net order flow by*

$$q = (q^b, q^a) = \Psi(x, R, \tilde{R})$$

where

- $x_t = (x_t^b, x_t^a) = \left(\sum_{i=1}^{N_t^b} V_i^b, \sum_{i=1}^{N_t^a} V_i^a\right)$ *is the net order flow at the bid and the ask,*
- $R = (R_n)_{n\geq 1}$ *is the sequence of queue sizes after a price increase, and*
- $\tilde{R} = (\tilde{R_n})_{n\geq 1}$ *is the sequence of queue sizes after a price decrease.*

One can thus build a statistical model for the limit order book by specifying the joint law of x and of the regulating sequences $(R, \tilde{R})$. This approach simplifies the study of the (asymptotic) properties of $q_t = (q_t^b, q_t^a)$.

Example 1 (IID reinitializations). The simplest case is the case where the queue length after each price change is independent from the history of the order book, as in Cont and de Larrard (2010). $R = (R_n)_{n\geq 1}$ and $\tilde{R} = (\tilde{R_n})_{n\geq 1}$ are then IID sequences with values in $]0, \infty[^2$. Figure 5 shows an example of such a distribution for a liquidly traded stock (NYSE: CitiGroup).

The law of the process $Q = \Psi(x, R, \tilde{R})$ is then entirely determined by the law of the net order flow x and the distributions of R_n, $\tilde{R_n}$: it can be constructed from the concatenation of the laws of $(x_t, \tau_k \leq t < \tau_{k+1})$ for $k \geq 0$ (where we define $\tau_0 := 0$).

Example 2 (Pegged limit orders). Most electronic trading platforms allow to place limit orders which are *pegged* to the best quote: if the best quote moves to a new price level, a pegged limit order moves along with it to the new price level. The presence of pegged orders leads to positive autocorrelation and dependence in the queue size before/after a price change. The queue size after a price change may be modeled as

- $q_{\tau_n} = R_n = (\epsilon_n^b + \beta q_{\tau_n-}^b, \epsilon_n^a)$ if the price has increased, and
- $q_{\tau_n} = \tilde{R_n} = (\tilde{\epsilon}_n^b, \tilde{\epsilon}_n^a + \tilde{\beta} q_{\tau_n-}^a)$ if the price has decreased

where $\epsilon_n = (\epsilon_n^b, \epsilon_n^a), \tilde{\epsilon}_n = (\tilde{\epsilon}_n^b, \tilde{\epsilon}_n^a)$ are IID sequences. Empirically, one observes a correlation of $\sim 10\% - 20\%$ between the queue lengths before and after a price change, which suggests an order magnitude for the fraction of pegged orders.

As in the previous example, the law of of the process $q = \Psi(x, R, \tilde{R})$ is determined by the law of the net order flow x, the coefficients $\beta, \tilde{\beta}$ and the distributions of ϵ,$\tilde{\epsilon}$: it can be constructed from the concatenation of the laws of $(x_t, \tau_k \leq t < \tau_{k+1})$ for $k \geq 0$.

More generally, one could consider other extensions where the queue size after a price move may depend in a (nonlinear) way on the queue size before the price move and a random term ϵ_n representing the inflow of new orders after the n-th price change:

$$q_{\tau_n} = g(q_{\tau_n -}, \epsilon_n). \tag{5}$$

The results given below hold for this general specification although the examples 1 and 2 above are sufficiently general for most applications.

2.3. Examples

The framework described in Section 2.1 allows a wide class of specifications for the order flow process, and contains as special cases various models proposed in the literature. Each model involves a specification for the (random) sequences $(t_i^a, t_i^b, V_i^a, V_i^b)_{i\geq 1}$, $R = (R_n)_{n\geq 1}$ and $\tilde{R} = (\tilde{R_n})_{n\geq 1}$ or, equivalently, $(T_i^a, T_i^b, V_i^a, V_i^b)_{i\geq 1}$, $R = (R_n)_{n\geq 1}$ and $\tilde{R} = (\tilde{R_n})_{n\geq 1}$ where $T_i^a = t_{i+1}^a - t_i^a$ (resp. $T_i^b = t_{i+1}^b - t_i^b$) are the durations between order book events on the ask (resp. the bid) side.

2.3.1. Models based on Poisson point processes Cont and de Larrard (2010) study a stylized model of a limit order market in which market orders, limit orders and cancelations arrive at independent and exponential times with corresponding rates μ, λ and θ, the process $q=(q^b,q^a)$ becomes a Markov process. If we assume additionally that all orders have the same size, the dynamics of the reduced limit order book is described by:

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

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

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

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

- All these sequences are independent.

It is readily verified that this model is a special case of the framework of Section 2.1: $(q_t)_{t\geq 0}$ may be constructed as in Definition 1, where the unconstrained process x_t is now a compound Poisson process.

2.3.2. Self-exciting point processes Empirical studies of order durations highlight the dependence in the sequence of order durations. This feature, which is not captured in models based on Poisson processes, may be adequately represented by a multidimensional *self-exciting* point process Andersen et al. (2010), Hautsch (2004), in which the arrival rate $\lambda_i(t)$ of an order of type i is represented as a stochastic process whose value depends on the recent history of the order flow: each new order increases the rate of arrival for subsequent orders of the same type (self-exciting property) and may also affect the rate of arrival of other order types (mutually exciting property):

$$\lambda_i(t)=\theta_i+\sum_{j=1}^{J}\delta_{ij}\int_0^t e^{-\kappa_i(t-s)}dN_j(s)$$

Here δ_{ij} measures the impact of events of type j on the rate of arrival of subsequent events of type i: as each event of type j occurs, λ_i increases by δ_{ij}. In between events, $\lambda_i(t)$ decays exponentially at rate κ_i. Maximum likelihood estimation of this model on TAQ data Andersen et al. (2010) shows evidence of self-exciting and mutually exciting features in order flow: the coefficients δ_{ij} are all significantly different from zero and positive, with $\delta_{ii}>\delta_{ij}$ for $j\neq i$.

2.3.3. Autoregressive conditional durations Models based on Poisson process fail to capture serial dependence in the sequence of durations, which manifests itself in the form of clustering of order book events. One approach for incorporating serial dependence in event durations is to represent the duration T_i between transactions $i-1$ and i as

$$T_i=\psi_i\epsilon_i,$$

where $(\epsilon_i)_{i\geq 1}$ is a sequence of independent positive random variables with common distribution and $\mathbb{E}[\epsilon_i]=1$ and the *conditional duration* $\psi_i=\mathbb{E}[T_i|\psi_{i-j},T_{i-j},j\geq 1]$ is modeled as a function of past history of the process:

$$\psi_i=G(\psi_{i-1},\psi_{i-2},...,..;T_{i-1},T_{i-2},...,..).$$

Engle and Russell's Autoregressive Conditional Duration model Engle and Russell (1998) propose an ARMA(p,q) representation for G:

$$\psi_i = a_0 + \sum_{i=1}^{p} a_k \psi_{i-k} + \sum_{i=1}^{q} b_q T_{j-k}$$

where $(a_0, ..., a_p)$ and $(b_1, ..., b_q)$ are positive constants. The ACD-GARCH model Ghysels and Jasiak (1998) combine this model with a GARCH model for the returns. Engle (2000) proposes a GARCH-type model with random durations where the volatility of a price change may depend on the previous durations. Variants and extensions are discussed in Hautsch (2004). Such models, like ARMA or GARCH models defined on fixed time intervals, have likelihood functions which are numerically computable. Although these references focus on transaction data, the framework can be adapted to model the durations $(T_i^a, i \geq 1)$ and $(T_i^b, i \geq 1)$ between order book events with the ACD framework (Hautsch 2004).

2.3.4. A limit order market with patient and impatient agents Another way of specifying a stochastic model for the order flow in a limit order market is to use an 'agent-based' formulation where agent types are characterized in terms of the statistical properties of the order flow they generate. Consider for example a market with three types of traders:

- *impatient* traders who only submit market orders:
- *patient* traders who use *only limit orders*: this is the case for example of traders who place stop loss orders or engage in strategies such as mean-reversion arbitrage or pairs trading which are only profitable with limit orders.
- other traders who use *both limit and market orders*; we will assume these traders submit a proportion γ of their orders as limit orders and $(1-\gamma)$ as market orders, where $0 < \gamma < 1$.

Denote by m (resp. l) the proportion of orders generated by impatient (resp. patient) traders:

$$\begin{aligned} \forall i \geq 1, \quad \mathbb{P}[i-\text{th trader uses only market orders}] &= m, \\ \mathbb{P}[i-\text{th trader uses only limit orders}] &= l, \\ \mathbb{P}[i\text{th trader uses both limit and market orders}] &= 1-l-m. \end{aligned}$$

Assume that the sequence $(T_i, i \geq 1)$ of duration between consecutive orders is a stationary ergodic sequence of random variables with $\mathbb{E}[T_i] < \infty$, that each trader has an equal chance of being a buyer or a seller and that the type of trader (buyer or seller) is independent from the past:

$$\mathbb{P}[i-\text{th trader is a buyer}] = \mathbb{P}[i-\text{th trader is a seller}] = \frac{1}{2}$$

Trader i generates an order of size V_i, where $(V_i, i \geq 1)$ is an IID sequence with:

$$\begin{aligned} \mathbb{P}[(V_i^b, V_i^a) = (V_i, 0)] = \mathbb{P}[(V_i^b, V_i^a) = (0, V_i)] &= \frac{m}{2}, \\ \mathbb{P}[(V_i^b, V_i^a) = (-V_i, 0)] = \mathbb{P}[(V_i^b, V_i^a) = (0, -V_i)] &= \frac{l}{2}, \\ \mathbb{P}[(V_i^b, V_i^a) = (\gamma V_i, -(1-\gamma)V_i)] = \mathbb{P}[(V_i^b, V_i^a) = (-(1-\gamma)V_i, \gamma V_i)] &= \frac{1-l-m}{2}. \end{aligned}$$

3. Statistical properties of high-frequency order flow

As described in Section 2.1, the sequence of order book events –the *order flow*– is characterized by the sequences $(T_i^a, i \geq 1)$ and $(T_i^b, i \geq 1)$ of durations between orders and the sequences of order sizes $(V_i^b, i \geq 1)$ and $(V_i^a, i \geq 1)$. In this section we illustrate the statistical properties of these sequences using high-frequency quotes and trades for liquid US stocks –CitiGroup, General Electric, General Motors– on June 26th, 2008.

3.1. Order sizes

Empirical studies Bouchaud et al. (2002, 2008), Gopikrishnan et al. (2000), Maslov and Mills (2001) have shown that order sizes are highly heterogeneous and exhibit heavy-tailed distributions, with Pareto-type tails:

$$\mathbb{P}(V_i^a \geq x) \sim C x^{-\beta}$$

with tail exponent $\beta > 0$ between 2 and 3, which corresponds to a series with finite variance but infinite moments of order ≥ 3. The tail exponent $\beta > 0$ is difficult to estimate precisely, but the Hill estimator Resnick (2006) can be used to measure the heaviness of the tails. Table 4 gives the Hill estimator of the tail coefficient of order sizes for our samples. This estimator is larger than 2 for both the bid and the ask; this means that the sequence of order sizes have a finite moment of order two.

	Bid side	Ask side
Citigroup	[0.42, 0.46]	[0.29, 0.32]
General Electric	[0.42, 0.45]	[0.41, 0.46]
General Motors	[0.36, 0.42]	[0.44, 0.51]

Table 4 95-percent confidence interval of the Hill estimator of the sequence of order sizes. When the Hill estimator is < 0.5, the estimated tail index is large than 2 and the distribution has finite variance.

The sequences of order sizes $(V_i^a, i \geq 1)$ and $(V_i^b, i \geq 1)$ exhibit insignificant autocorrelation, as observed on Figure 6. However, they are far from being independent: the series of squared order sizes $((V_i^b)^2, i \geq 1)$ and $((V_i^a)^2, i \geq 1)$ are positively autocorrelated, as shown in Figure 7.

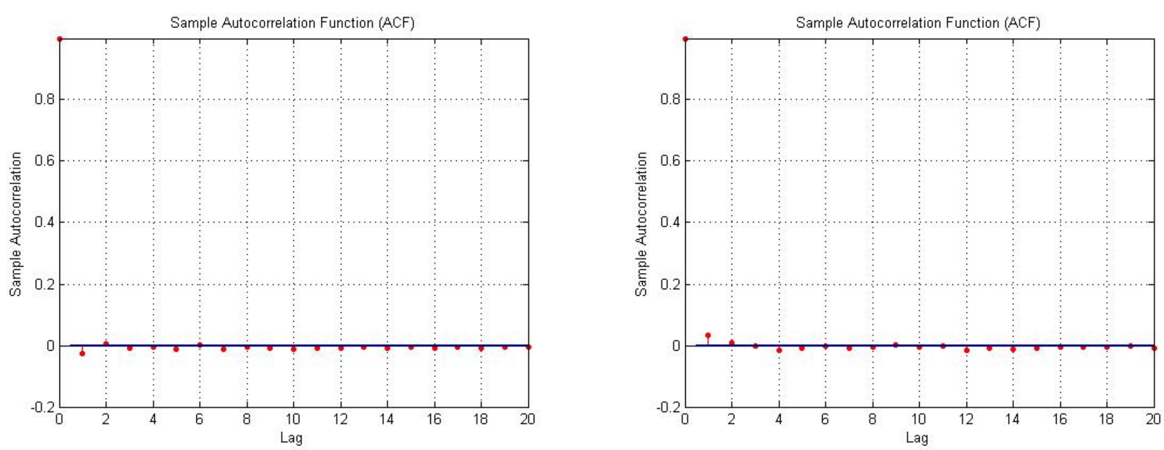

Figure 6 Autocorrelogram of the sequence of order sizes. Order coming at the ask on the left and at the bid on the right.

Finally, the sequences $(V_i^a, i \geq 1)$ and $(V_i^b, i \geq 1)$ may be negatively correlated. This stems from the fact that a buyer can simultaneously use market orders on the ask side (which correspond to negative values of V_i^a) and limit orders on the bid side (which correspond to positive values of V_i^b); the same argument holds for sellers (see Section 2.3.4).

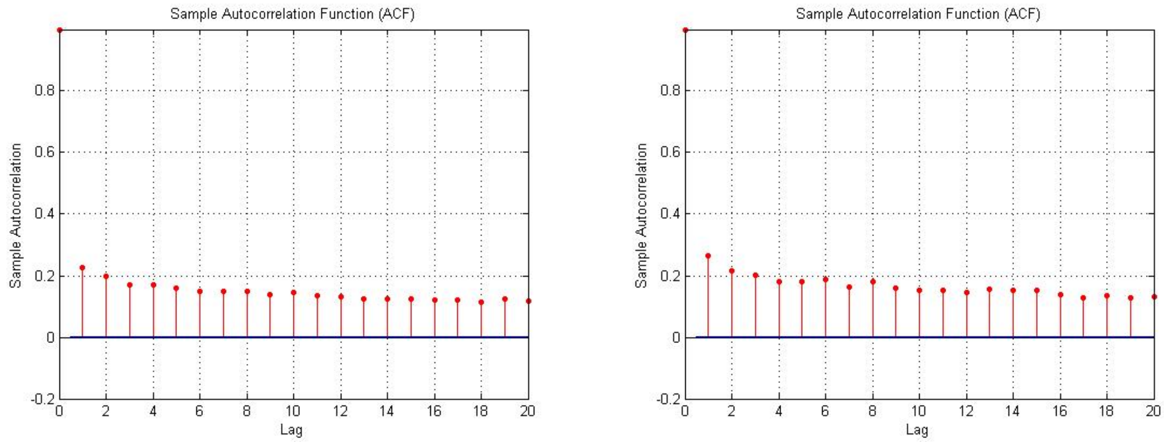

Figure 7 Autocorrelogram of the sequence of absolute order sizes. Order coming at the ask on the left and at the bid on the right.

These properties of the sequence $(V_i^a, V_i^b)_{i\geq 1}$ may be modeled using a bivariate ARCH process:

$$V_i^b = \sigma_i^b z_i^b \qquad V_i^a = \sigma_i^a z_i^a$$
$$(\sigma_i^b)^2 = \alpha_0^b + \alpha_1^b (V_{t-1}^b)^2, \qquad (\sigma_i^a)^2 = \alpha_0^a + \alpha_1^a (V_{t-1}^a)^2, \qquad \text{where} \qquad (z_i^b, z_i^a)_{i\geq 1} \overset{IID}{\sim} N\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

and $(\alpha_0^b, \alpha_1^b, \alpha_0^a, \alpha_1^a)$ are positive coefficients satisfying

$$0 < \alpha_0^b + \alpha_1^b < 1, \quad \text{and} \quad 0 < \alpha_0^a + \alpha_1^a < 1. \tag{6}$$

As shown by Bougerol and Picard (1992), under the assumption (6), the sequence of order sizes $(V_i^b, V_i^a)_{i\geq 1}$ is then a well defined, stationary sequence of random variables with finite second-order moments, satisfying the properties enumerated above.

3.2. Durations

The timing of order book events is describe by the sequence of durations $(T_i^b, i \geq 1)$ at the bid and $(T_i^a, i \geq 1)$ at the ask. These sequences have zero autocorrelation (see Figure 8) but are not independence sequences: for example, as shown in Figure 9, the sequence of inverse durations $(1/T_i^b, i \geq 1)$ and $(1/T_i^b, i \geq 1)$ has significant autocorrelations.

Figure 10 represents the empirical distribution functions $\mathbb{P}[T^a > u]$ and $\mathbb{P}[T^b > u]$ in logarithmic scale. Both empirical distributions exhibit thin, exponential-type tails (which implies in particular that T^a and T^b have finite expectation).

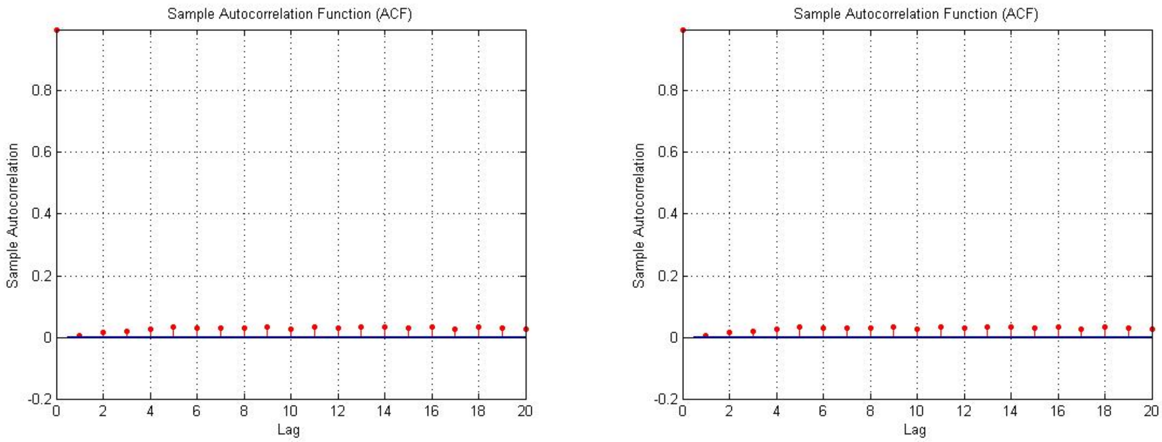

Figure 8 Autocorrelogram of the sequence of durations for events at the ask (left) and the bid (right).

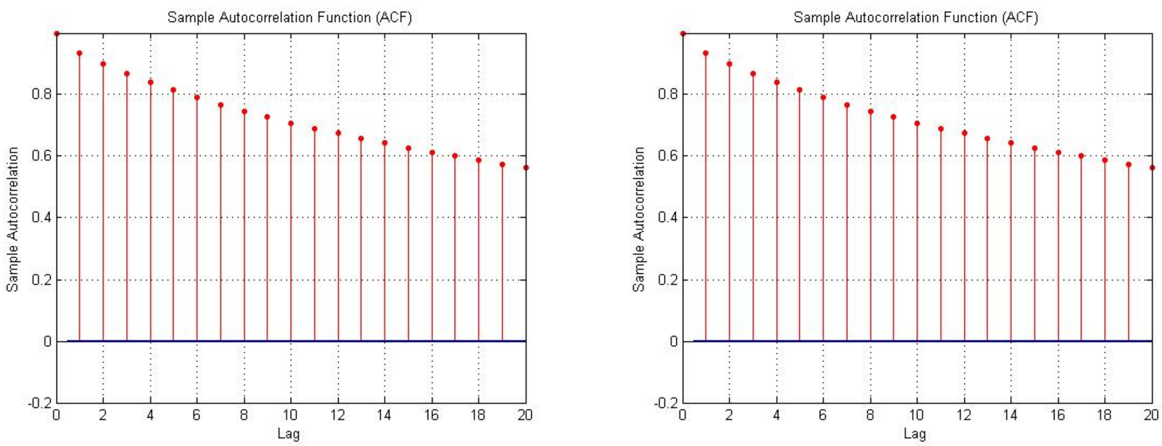

Figure 9 Autocorrelogram of the sequence of inverse durations for events at the ask (left) and the bid (right).

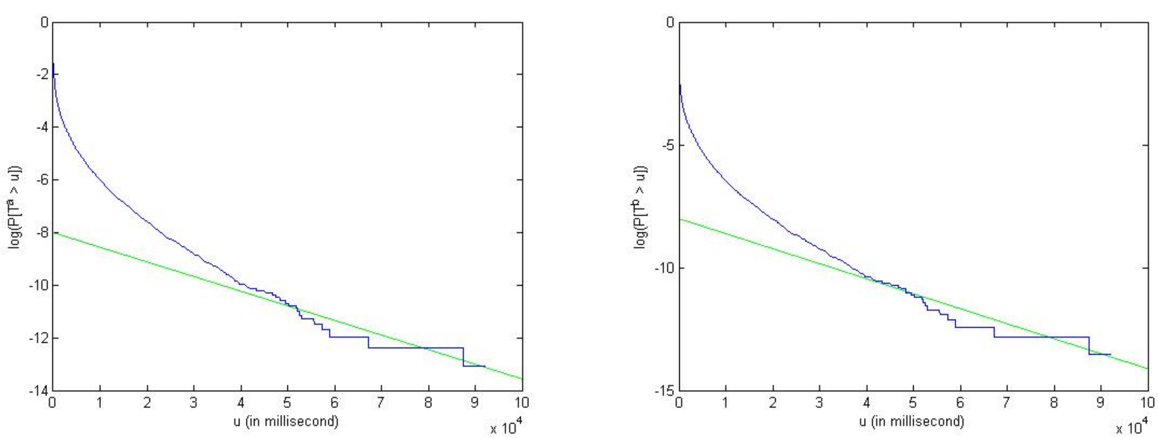

Figure 10 Logarithm of the empirical distribution function of durations for events at the ask (left) and the bid (right).

4. Heavy Traffic limit

At very high frequency, the limit order book is described by a two-dimensional piecewise constant process $q_t = (q_t^b, q_t^a)_{t\geq 0}$, whose evolution is determined by the flow of orders. The complex nature of this order flow –heterogeneity and serial dependence in order sizes, dependence between orders coming at the ask and at the bid– described in section 3, makes it difficult to describe q_t in an analytically tractable manner which would allow the quantities of interest to be computed either in closed form or numerically in real time applications. However, if one is interested in the evolution of the order book over time scales much larger than the interval between individual order book events, the (coarse-grained) dynamics of the queue sizes may be described in terms of a simpler process Q, called the *heavy traffic approximation* of q. In this limit, the complex dynamics of the discrete queueing system is approximated by a simpler system with a continuous state space, which can be either described by a system of ordinary differential equations (in the 'fluid limit', where random fluctuations in queue sizes vanish) or a system of stochastic differential equations (in the 'diffusion limit' where random fluctuations dominate). This idea has been widely used in queueing theory to obtain useful analytical insights into the dynamics of queueing systems Harrison and Nguyen (1993), Iglehart and Whitt (1970), Whitt (2002).

We argue that the heavy traffic limit is highly relevant for the study of limit order books in liquid markets, and that the correct scaling limit for the liquid stocks examined in our data sets is the "diffusion" limit. This *heavy traffic* limit is then derived in Theorem 2 and described in Section 4.3.

4.1. Fluid limit or diffusion limit?

Let $(V_i^{n,a}, i \geq 1)$ the sequence of order sizes, whose properties depend on the index n. One way of viewing the heavy traffic limit is to view the limit order book at a lower time resolution, by grouping together events in batches of size n. Since the inter-event durations are finite, this is equivalent to rescaling time by n. The impact, on the net order flow, of a batch of n events at the ask is

$$\frac{V_1^{n,a} + V_2^{n,a} + V_3^{n,a} + \ldots + V_n^{n,a}}{\sqrt{n}} = \frac{(V_1^{n,a} - \overline{V^{n,a}}) + (V_2^{n,a} - \overline{V^{n,a}}) + \ldots + (V_n^{n,a} - \overline{V^{n,a}})}{\sqrt{n}} + \sqrt{n} \quad \overline{V^{n,a}},$$

where $\overline{V^{n,a}} = \mathbb{E}[V_1^{n,a}]$. Under appropriate assumptions (see next section), this sum behaves approximately as a Gaussian random variable for large n:

$$\frac{V_1^{n,a} + V_2^{n,a} + V_3^{n,a} + \ldots + V_n^{n,a}}{\sqrt{n}} \sim N(\sqrt{n}\ \overline{V^{n,a}}, \mathrm{Var}(V_1^{n,a})) \quad \text{as} \quad n \to \infty. \tag{7}$$

Two regimes are possible, depending on the behavior of the ratio $\frac{\sqrt{n}\ \overline{V^{n,a}}}{\sqrt{\mathrm{Var}(V_1^{n,a})}}$ as n grows:

- If $\frac{\sqrt{n}\ \overline{V^{n,a}}}{\sqrt{\mathrm{Var}(V_i^{n,a})}} \to \infty$ as $n \to \infty$, the correct approximation is given by the *fluid limit*, which describes the (deterministic) behavior of the average queue size.
- If $\lim_{n\to\infty} \frac{\sqrt{n}\ \overline{V^{n,a}}}{\sqrt{\mathrm{Var}(V_i^{n,a})}} < \infty$, the rescaled queue sizes behave like a diffusion process.

The fluid limit corresponds to the regime of law of large numbers, where random fluctuations average out and the limit is described by average queue size, whereas the diffusion limit corresponds to the regime of the (functional) central limit theorem, where fluctuations in queue size are asymptotically Gaussian.

Figure 11 displays the histogram of the ratio $\frac{\sqrt{n}\ \overline{V^{n,a}}}{\sqrt{\mathrm{Var}(V_i^{n,a})}}$ for stocks in the Dow Jones index, where for each stock n is chosen to represent the average number of order book events in a 10 second interval (typically $n \sim 100 - 1000$). This ratio is shown to be rather small at such intraday time scales, showing that the diffusion approximation, rather than the fluid limit, is the relevant approximation to use here.

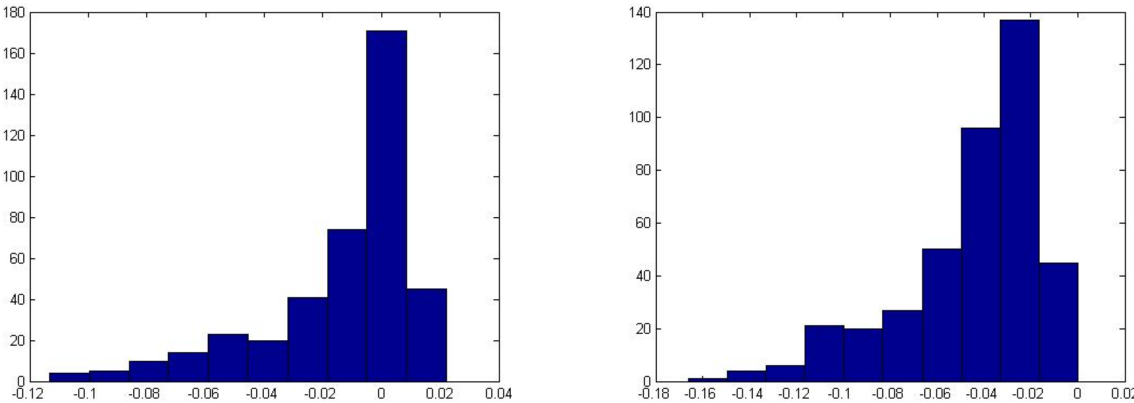

Figure 11 Empirical distribution of the ratio $\sqrt{n}/\overline{V^{n,a}}\sqrt{\mathrm{Var}(V_1^{n,a})}$ showing the relative importance of average change vs fluctuations in queue size, for stocks in the Dow-Jones index during June 08 (see Section 4.1). Low values of the ratio indicate that intraday changes in bid/ask queue size are dominated by fluctuations, rather than the average motion of the queue. Left: bid side. Right: ask side.

Bid and ask queue sizes (q_t^b, q_t^a) exhibit a diffusion-type behavior at such intraday time scales. Figure 12 shows the path of the net order flow process

$$x_t = (q_0^b, q_0^a) + \left(\sum_{i=1}^{N_t^b} V_i^b, \sum_{i=1}^{N_t^a} V_i^a \right) \tag{8}$$

sampled every second for CitiGroup stocks on a typical trading day. In this example, for which the average time between consecutive orders is $\lambda^{-1} \simeq 13\ ms \ll 1$ second, we observe that the process X behaves like a diffusion in the orthant with negative drift: the randomness of queue sizes does not average out at this time scale.

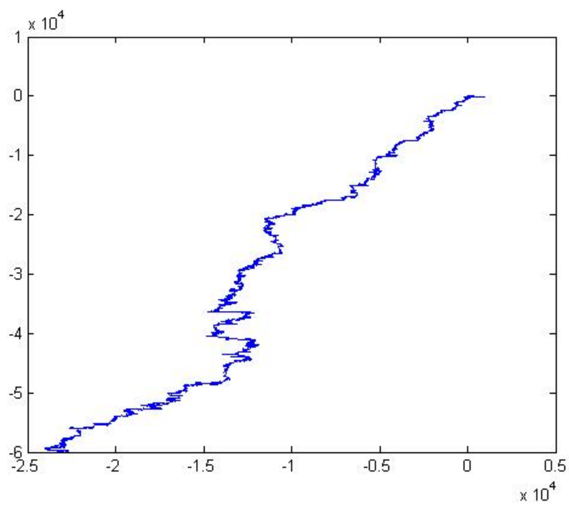

Figure 12 Evolution of the net order flow $X_t = (X_t^b, X_t^a)$ given by Eq. (8) for CitiGroup shares over one trading day (June 26, 2008). The starting point is taken to be (0,0) at the marke open. Note the irregular, diffusive feature of the path of X and its negative drift.

We will now show that this is a general result: under mild assumptions on the order flow process, we will show that the (rescaled) queue size process

$$(Q_t^n)_{t\geq 0} := (\frac{q_{nt}^{n,b}}{\sqrt{n}}, \frac{q_{nt}^{n,a}}{\sqrt{n}})_{t\geq 0} \tag{9}$$

converges in distribution to a Markov process $(Q_t)_{t\geq 0}$ in the positive orthant, whose features we will now describe in terms of the statistical properties of the order flow.

4.2. A functional central limit theorem for the limit order book

Consider now a sequence $q^n = (q_t^n)_{t\geq 0}$ of processes, where q^n represents the dynamics of the bid and ask queues in the limit order book at a time resolution corresponding to n events (see discussion above). The dynamics of q^n is characterized by the sequence of order sizes $(V_i^{n,b}, V_i^{n,a})_{i\geq 1}$, durations $(T_i^{n,b}, T_i^{n,a})_{i\geq 1}$ between orders and the fact that, at each price change

- $q_{\tau_k}^n = R_k^n = g(q_{\tau_k -}^n, \epsilon_k^n)$ if the price has increased, and
- $q_{\tau_k}^n = \tilde{R}_k^n = g(q_{\tau_k -}^n, \tilde{\epsilon}_k^n)$ if the price has decreased,

where $(\epsilon_k^n, k\geq 1)$ is an IID sequence with distribution f_n, and $(\tilde{\epsilon}_k^n, k\geq 1)$ is an IID sequence with distribution $\tilde{f}_n$. Note that this specification includes Examples 1 and 2 as special cases.

We make the following assumptions, which allow for an analytical study of the heavy traffic limit and are sufficiently general to accommodate high frequency data sets of trades and quotes such as the ones described in Section 3:

Assumption 1. *$(T_i^{n,a}, T_i^{n,b})_{i\geq 1}$ is a stationary array of positive random variables whose common distribution has a continuous density and satisfies*

$$\lim_{n\to\infty} \frac{T_1^{n,a} + T_2^{n,a} + \ldots + T_n^{n,a}}{n} = \frac{1}{\lambda^a} < \infty, \qquad \lim_{n\to\infty} \frac{T_1^{n,b} + T_2^{n,b} + \ldots + T_n^{n,b}}{n} = \frac{1}{\lambda^b} < \infty.$$

λ^a (resp. λ^b) represents the arrival rate of orders at the ask (resp. the bid).

Assumption 2. *$(V_i^{n,a}, V_i^{n,b})_{i\geq 1}$ is a stationary, uniformly mixing array of random variables satisfying*

$$\sqrt{n}\mathbb{E}[V_1^{n,a}] \overset{n\to\infty}{\to} \overline{V^a}, \qquad \sqrt{n}\mathbb{E}[V_1^{n,b}] \overset{n\to\infty}{\to} \overline{V^b}, \tag{10}$$

$$\lim_{n\to\infty} \mathbb{E}[(V_i^{n,a} - \overline{V^a})^2] + 2\sum_{i=2}^{\infty} \mathrm{cov}(V_1^{n,a}, V_i^{n,a}) = v_a^2 < \infty, \qquad \text{and}$$

$$\lim_{n\to\infty} \mathbb{E}[(V_i^{n,b} - \overline{V^b})^2] + 2\sum_{i=2}^{\infty} \mathrm{cov}(V_1^{n,b}, V_i^{n,b}) = v_b^2 < \infty.$$

The assumption of uniform mixing (Billingsley 1968, Ch. 4) implies that the partial sums of order sizes verify a central limit theorem, but allows for various types of serial dependence in order sizes. The scaling assumptions on the first two moments corresponds to the properties of the empirical data discussed in Section 4.1. Under Assumption 2, one can define

$$\rho := \lim_{n\to\infty} \frac{1}{v_a v_b} \left(2\max(\lambda^a, \lambda^b)\mathrm{cov}(V_1^{n,a}, V_1^{n,b}) + 2\sum_{i=1}^{\infty} \lambda^a \mathrm{cov}(V_1^{n,a}, V_i^{n,b}) + \lambda^b \mathrm{cov}(V_1^{n,b}, V_i^{n,a}) \right). \tag{11}$$

$\rho \in (-1,1)$ may be interpreted as a measure of 'correlation' between event sizes at the bid and event sizes at the ask.

These assumptions hold for the examples of Section 2.3. In the case of the Hawkes model, Assumption 1 was shown to hold in Bacry et al. (2010). Also, these assumptions are quite plausible for high frequency quotes for liquid US stocks since, as argued in Section 3:

• The tail index of order sizes is larger than two, so the sequences $(V_i^b, i \geq 1)$ and $(V_i^a, i \geq 1)$ have a finite second moment.

• The sequence of order sizes is uncorrelated i.e. has statistically insignificant autocorrelation. Therefore the sum of autocorrelations of order sizes is finite (zero, in fact).

• The sequence of inter-event durations has a finite empirical mean and is not autocorrelated.

These empirical observations support the plausibility of Assumptions 1 and 2 for the data sets examined.

Assumption 2 has an intuitive interpretation: if orders are grouped in batches of n orders, then Assumption 2 amounts to stating that the variance of batch sizes should scale linear with n. This assumption can be checked empirically, using a variance ratio test for example: Figure 13 shows that this linear relation is indeed verifies for the data sets examined in Section 3.

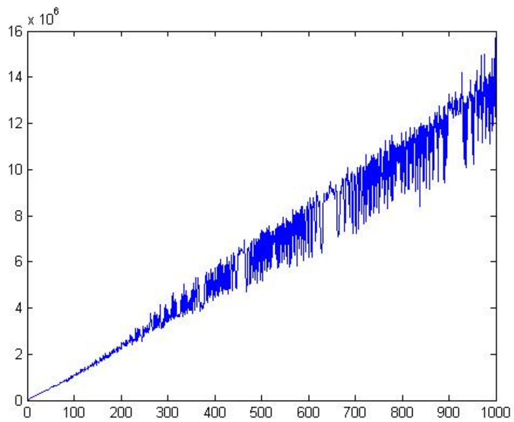

Figure 13 Variance of batch sizes of n orders, for General Electric shares, on June 26th, 2008. Left: ask side. Right: bid side.

The following scaling assumption states that, when grouping orders in batches of n orders, a good proportion of batches should have a size $O(\sqrt{n})$ (otherwise their impact will vanish in the limit when n becomes large):

Assumption 3. *There exist probability distributions $F, \tilde{F}$ on the interior $(0,\infty)\times(0,\infty)$ of the positive orthant, such that*

$$n f_n(\sqrt{n}\ .) \overset{n\to\infty}{\Rightarrow} F \quad \text{and} \quad n\tilde{f}_n(\sqrt{n}\ .) \overset{n\to\infty}{\Rightarrow} \tilde{F}.$$

Assumption 4. $g \in C^2(\mathbb{R}_+^2 \times \mathbb{R}_+^2,]0,\infty[^2)$ *and*

$$\exists \alpha > 0, \forall (x,y) \in \mathbb{R}_+^2 \times \mathbb{R}_+^2, \qquad \|g(x,y)\| \geq \ \alpha \min(y_1, y_2).$$

Finally, we add the following condition for the initial value of the queue sizes:

$$\left(\frac{q_0^{n,b}}{\sqrt{n}}, \frac{q_0^{n,a}}{\sqrt{n}}\right) \overset{n\to\infty}{\to} (x_0, y_0) \in]0,\infty[\times]0,\infty[\tag{12}$$

The following theorem, whose proof is given in the Appendix, describes the joint dynamics of the bid and ask queues in this heavy traffic limit:

Theorem 2 **(Heavy traffic limit)**. *Under Assumptions 3, 1, 2 and 4, the rescaled process*

$$(Q_t^n, t \geq 0) = \left(\frac{q_{nt}^n}{\sqrt{n}}, t \geq 0\right)$$

converges weakly, on the Skorokhod space $(D(]0,\infty[,\mathbb{R}_+^2), J_1)$,

$$Q^n \overset{n\to\infty}{\Rightarrow} Q$$

to a Markov process $(Q_t)_{t\geq 0}$ *with values in* $\mathbb{R}_+^2$, *initial value* $Q_0 = (x_0, y_0)$ *given in* (12) *and infinitesimal generator* $\mathcal{G}$ *given, for* $x > 0, y > 0$, *by*

$$\mathcal{G}h(x,y) = \lambda^a \overline{V^a} \frac{\partial h}{\partial y} + \lambda^b \overline{V^b} \frac{\partial h}{\partial x} + \frac{\lambda^a v_a^2}{2} \frac{\partial^2 h}{\partial y^2} + \frac{\lambda^b v_b^2}{2} \frac{\partial^2 h}{\partial x^2} + \rho\sqrt{\lambda^a \lambda^b} v_a v_b \frac{\partial^2 h}{\partial x \partial y} \tag{13}$$

$$\mathcal{G}h(x,0) = \int_{]0,\infty[^2} \mathcal{G}h(g((x,0),(u,v)))F(du,dv), \quad \mathcal{G}h(0,y) = \int_{]0,\infty[^2} \mathcal{G}h(g((0,y),(u,v)))\tilde{F}(du,dv),$$

and whose domain is the set $dom(\mathcal{G})$ *of functions* $h \in C^2(]0,\infty[\times]0,\infty[,\mathbb{R}) \cap C^0(\mathbb{R}_+^2,\mathbb{R})$ *verifying the Wentzell boundary conditions*

$$\forall x > 0, \qquad h(x,0) = \int_{\mathbb{R}_+^2} h(g((x,0),(u,v)))F(du,dv), \tag{14}$$

$$\forall y > 0, \qquad h(0,y) = \int_{\mathbb{R}_+^2} h(g((0,y),(u,v)))\tilde{F}(du,dv).$$

We outline here the main steps of the proof. The technical details are given in the Appendix. Define the counting processes

$$N_t^{a,n} = \sup\{k \geq 0,\ T_1^{a,n} + ... + T_k^{a,n} \leq t\} \quad \text{and} \quad N_t^{b,n} = \sup\{k \geq 0,\ T_1^{b,n} + ... + T_k^{b,n} \leq t\} \tag{15}$$

which correspond to the number of events at the ask (resp. the bid), and the net order flow

$$X_t^n = \left(\sum_{i=1}^{N_{nt}^{b,n}} \frac{V_i^{b,n}}{\sqrt{n}}, \sum_{i=1}^{N_{nt}^{a,n}} \frac{V_i^{a,n}}{\sqrt{n}}\right)$$

Then, as shown in Proposition 3 (see Appendix), X^n converges in distribution on $(D([0,\infty[,\mathbb{R}^2), J_1)$ to a two-dimensional Brownian motion with drift

$$(X_t^n)_{t\geq 0} \overset{n\to\infty}{\Rightarrow} \left(Z_t + t(\lambda^b \overline{V^b}, \lambda^a \overline{V^a})\right)_{t\geq 0}$$

where Z is a planar Brownian motion with covariance matrix

$$\begin{pmatrix} \lambda^b v_b^2 & \rho\sqrt{\lambda^a\lambda^b} v_a v_b \\ \rho\sqrt{\lambda^a\lambda^b} v_a v_b & \lambda^a v_a^2 \end{pmatrix}.$$

Under assumption 3, using the Skorokhod representation theorem, there exist IID sequences $((\epsilon_k^n, n \geq 1), (\tilde{\epsilon}_k^n, n \geq 1), \epsilon_k, \tilde{\epsilon}_k)_{k\geq 1}$ and a copy X of the process

$$\left((x_0, y_0) + Z_t + t(\lambda^b \overline{V^b}, \lambda^a \overline{V^a})\right)_{t\geq 0}$$

on some probability space $(\Omega_0, \mathcal{B}, \mathbb{Q})$ such that $\epsilon_k^n \sim f_n, \tilde{\epsilon}_k^n \sim \tilde{f}_n, \epsilon_k \sim F, \tilde{\epsilon}_k \sim \tilde{F}$ and

$$\mathbb{Q}\left(X^n \overset{n\to\infty}{\to} X \quad ; \forall k \geq 1, \frac{\epsilon_k^n}{\sqrt{n}} \overset{n\to\infty}{\to} \epsilon_k, \quad \frac{\tilde{\epsilon}_k^n}{\sqrt{n}} \overset{n\to\infty}{\to} \tilde{\epsilon}_k\right) = 1.$$

Using the notations of Appendix 6.2, denote by

- $\tau_1^n = \tau(X^n)$ the first exit time of X^n from the positive orthant $\mathbb{R}_+^2$ and
- τ_k^n the first exit time of $\Psi_{k-1}(X^n, Q^n_{\tau_1^n}, ..., Q^n_{\tau_{k-1}^n})$ from $\mathbb{R}_+^2$.

We can now construct the process Q by an induction procedure. Let $\tau_1 = \tau(X)$ be the first exit time of X from the orthant. Let $Q_t = X_t$ for $t < \tau_1$ and, by continuity of the first-passage time map and the last-evaluation map at a first passage time (Whitt 2002, Sec. 13.6.3),

$$(\tau_1^n, Q^n_{\tau_1^n -}) \overset{n\to\infty}{\to} (\tau_1, Q_{\tau_1 -}) \ \mathbb{Q}-a.s.$$

We now set

$$Q_{\tau_1} = g(X_{\tau_1 -}, \epsilon_1) 1_{X_{\tau_1}.(0,1)<0} + g(X_{\tau_1 -}, \tilde{\epsilon}_1) 1_{X_{\tau_1}.(1,0)<0}.$$

Since X is a Brownian motion,

$$\liminf_{r\downarrow 0} (X_{\tau_1 + r} - X_{\tau_1}).(1,0) < 0$$

therefore $\mathbb{P}\left(1_{X_{\tau_1}.(1,0)<0} = 1_{X_{\tau_1}.(1,0)\leq 0}\right) = 1$ so we can also write

$$Q_{\tau_1} = g(Q_{\tau_1 -}, \epsilon_1) 1_{X_{\tau_1}.(0,1)\leq 0} + g(Q_{\tau_1 -}, \tilde{\epsilon}_1) 1_{X_{\tau_1}.(1,0)\leq 0}.$$

X is a continuous process and the probability thats its path crosses the origin is zero, so by Lemma 2, X lies with probability 1 in the continuity set of the map $G: \omega \to 1_{\omega_{\tau(\omega)}.(0,1)<0}$. So using the continuity of $g(.,.)$, we can apply the continuous mapping theorem (Billingsley 1968, Theorem 5.1), to conclude that

$$Q^n_{\tau_1^n} \overset{n\to\infty}{\to} Q_{\tau_1} \quad \mathbb{Q}-a.s.$$

Let us now assume that we have defined Q on $[0, \tau_{k-1}]$ and shown that

$$(\tau_1^n, .., \tau_{k-1}^n, Q^n_{\tau_1^n}, ..., Q^n_{\tau_{k-1}^n}) \overset{n\to\infty}{\to} (\tau_1, .., \tau_{k-1}, Q_{\tau_1}, ..., Q_{\tau_{k-1}}) \quad \mathbb{Q}-a.s.$$

Since $\mathbb{Q}((0,0) \notin \Psi_k(X, Q_{\tau_1}, ..., Q_{\tau_{k-1}})([0,\infty))\ ) = 1$, Lemma 4 implies that $(X, Q_{\tau_1}, ..., Q_{\tau_{k-1}})$ lies with probability 1 in the continuity set of Ψ_k, so by the continuous mapping theorem

$$\Psi_k(X^n, Q^n_{\tau_1^n}, ..., Q^n_{\tau_{k-1}^n}) \overset{n\to\infty}{\to} \Psi_k(X, Q_{\tau_1}, ..., Q_{\tau_{k-1}}) \quad \mathbb{Q}-a.s.$$

Define now τ_k as the first exit time of $\Psi_k(X, Q_{\tau_1}, ..., Q^n_{\tau_{k-1}})$ from the positive orthant $[0,\infty[\times[0,\infty[$. As before, by continuity of the first-passage time map and the last-value map at a first passage time (Whitt 2002, Sec. 13.6.3),

$$(\tau_k^n, Q^n_{\tau_k^n -}) \overset{n\to\infty}{\to} (\tau_k, Q_{\tau_k -}) \quad \mathbb{Q}-a.s.$$

We can now extend the definition of Q to $[0, \tau_k]$ by setting

$$Q_t = \Psi_k(X, Q_{\tau_1}, ..., Q_{\tau_k})(t) \qquad \text{for} \quad t < \tau_k, \qquad \text{and}$$

$$Q_{\tau_k} = g(Q_{\tau_k -}, \epsilon_k) 1_{\Psi_k(X, Q_{\tau_1}, ..., Q_{\tau_{k-1}}).(0,1)<0} + g(Q_{\tau_k -}, \tilde{\epsilon}_k) 1_{\Psi_k(X, Q_{\tau_1}, ..., Q_{\tau_{k-1}}).(1,0)<0}$$

As above, using the continuity properties of Ψ_k from Lemma 4 we conclude that $Q^n_{\tau_k^n} \to Q_{\tau_k}$ a.s. and using the Brownian property of X we can show that

$$Q_{\tau_k} = g(Q_{\tau_k -}, \epsilon_k) 1_{\Psi_k(X, Q_{\tau_1}, ..., Q_{\tau_{k-1}}).(0,1)\leq 0} + g(Q_{\tau_k -}, \tilde{\epsilon}_k) 1_{\Psi_k(X, Q_{\tau_1}, ..., Q_{\tau_{k-1}}).(1,0)\leq 0} \quad a.s.$$

So finally, we have shown that

$$\forall k \geq 1, \qquad (\tau_1^n, .., \tau_k^n, Q^n_{\tau_1^n}, ..., Q^n_{\tau_k^n}) \overset{n\to\infty}{\to} (\tau_1, .., \tau_k, Q_{\tau_1}, ..., Q_{\tau_k}) \qquad \mathbb{Q}-a.s.$$

We can now construct the sequences $R, \tilde{R}$ by setting

- $R_k = Q_{\tau_k}$ if $\Psi_k(X, Q_{\tau_1}, ..., Q_{\tau_{k-1}})(\tau_k-).(0,1) < 0$,
- $\tilde{R}_k = Q_{\tau_k}$ if $\Psi_k(X, Q_{\tau_1}, ..., Q_{\tau_{k-1}})(\tau_k-).(1,0) < 0$.

Then $Q = \Psi(X, R, \tilde{R})$ where Ψ is the map defined in Definition 1. Let us now show that $(X, R, \tilde{R})$ lies with probability 1 in the J_1-continuity set of Ψ, in order to apply the continuous mapping theorem. X is a continuous process whose paths lie in $C^0([0,\infty), \mathbb{R}^2 - \{(0,0)\})$ almost surely. Since F and $\tilde{F}$ have zero mass on the axes, with probability 1 the sequences $(\epsilon_k)_{k\geq 1}, (\tilde{\epsilon}_k)_{k\geq 1}$ do not have any accumulation point on the axes. Assumption 4 then implies that the sequences $(R_k)_{k\geq 1}, (\tilde{R}_k)_{k\geq 1}$ do not have any accumulation point on the axes. From the definition of Ψ (Definition 1), Q jumps at each hitting time of the axes and, in between two jumps, its increments follow those of the planar Brownian motion X. Since $F, \tilde{F}$ have no mass at the origin and planar Brownian paths have a zero probability of hitting isolated points, with probability 1 the graph of $Q = \Psi(X, R, \tilde{R})$ does not hit the origin :

$$\mathbb{Q}\Big(\ (0,0) \notin \Psi(X, R, \tilde{R})([0,\infty)\ )\ \Big) = 1. \tag{16}$$

So the triplet $(X, R, \tilde{R})$ satisfies the conditions of Theorem 1 almost-surely i.e. Ψ is continuous at $(X, R, \tilde{R})$ with probability 1. We can therefore apply the continuous mapping theorem (Billingsley 1968, Theorem 5.1) and conclude that

$$Q^n = (X^n, R^n, \tilde{R}^n) \overset{n\to\infty}{\Rightarrow} Q = \Psi(X, R, \tilde{R}).$$

The process $Q = \Psi(X, R, \tilde{R})$ can be explicitly constructed from the planar Brownian motion X and the sequences $R, \tilde{R}$: Q follows the increments of X and is reinitialized to R_n or $\tilde{R}_n$ at each hitting time of the axes. Lemma 5 in Appendix 6.4 uses this description to show that Q is a Markov process whose infinitesimal generator is given by (13)- (14).

Remark 1 (Lévy process limits). The diffusion approximation inside the orthant fails when order sizes do not have a finite second moment. For example, if the sequence (V_i^a, V_i^b) is regularly varying with tail exponent $\alpha \in (0,2)$ (see Resnick (2006) for definitions), the heavy-traffic approximation Q is a pure-jump process in the positive orthant, constructed by applying the map Ψ to a two-dimensional α-stable Lévy process L:

$$Q = \Psi(L, R, \tilde{R}),$$

i.e. by re-initializing it according to (5) at each attempted exit from the positive orthant. We do not further develop this case here, but it may be of interest for the study of illiquid limit order markets, or those where order flow is dominated by large block trades.

4.3. Jump-diffusion approximation for order book dynamics

Theorem 2 implies that, when examined over time scales much larger than the interval between order book events, the queue sizes q^b and q^a are well described by a Markovian jump-diffusion process $(Q_t)_{t\geq 0}$ in the positive orthant $\mathbb{R}_+^2$ which behaves like a a planar Brownian motion with drift vector

$$(\lambda^b \overline{V^b}, \lambda^a \overline{V^a}) \tag{17}$$

and covariance matrix

$$\begin{pmatrix} \lambda^b v_b^2 & \rho\sqrt{\lambda^a\lambda^b} v_a v_b \\ \rho\sqrt{\lambda^a\lambda^b} v_a v_b & \lambda^a v_a^2 \end{pmatrix}. \tag{18}$$

in the interior $]0,\infty[^2$ of the orthant and, at each hitting time τ_k of the axes, jumps to a new position

- $Q_{\tau_k} = R_k = g(Q_{\tau_k -}, \epsilon_k)$ whenever $Q^a_{\tau_{k-}} = 0$,

- $Q_{\tau_k} = \tilde{R}_k = g(Q_{\tau_k-}, \tilde{\epsilon}_k)$ whenever $Q^b_{\tau_{k-}} = 0$,

where the ϵ_k are IID with distribution F and the $\tilde{\epsilon}_k$ are IID with distribution $\tilde{F}$. We note that similar processes in the orthant were studied byBaccelli and Fayolle (1987) with queueing applications in mind, but not in the context of heavy traffic limits.

This process is analytically and computationally tractable and allows various quantities related to intraday price behavior to be computed (see next section).

If $\gamma_0 = (\mathbb{E}[T^a_1] + \mathbb{E}[T^b_1])/2$ is the average time between order book events, ($\gamma_0 \leq 100$ milliseconds), and $\gamma_1 \gg \gamma_0$ (typically, $\gamma_1 \sim$ 10-100 seconds) then Theorem 2 leads to an approximation for the distributional properties of the queue dynamics in terms of Q_t:

$$q_t \simeq^d \sqrt{N} \quad Q_{t/N} \qquad \text{where} \qquad N = \frac{\gamma_1}{\gamma_0}$$

So, under Assumptions 1, 2, 3 and 4 the queue sizes $(q^b_t, q^a_t)_{t\geq 0}$ can be approximated at the time scale γ_1 by a Markov process which

- behaves like a two-dimensional Brownian motion with drift (μ_b, μ_a) and covariance matrix Λ on $\{x > 0\} \cap \{y > 0\}$ with

$$\mu_a = \sqrt{N}\lambda^a\overline{V^a}, \qquad \mu_b = \sqrt{N}\lambda^b\overline{V^b}, \qquad \Lambda = N\begin{pmatrix} \lambda^b v_b^2 & \rho\sqrt{\lambda^a\lambda^b}v_a v_b \\ \rho\sqrt{\lambda^a\lambda^b}v_a v_b & \lambda^a v_a^2 \end{pmatrix} \tag{19}$$

and,

- jumps to a new value $g(q_{t-}, \sqrt{N}\epsilon_k)$ if $q^a_{t-} = 0$,
- jumps to a new value $g(q_{t-}, \sqrt{N}\tilde{\epsilon}_k)$ if $q^b_{t-} = 0$,

where $\epsilon_k \sim F$, $\tilde{\epsilon}_k \sim \tilde{F}$ are IID.

This gives a rigorous justification for modeling the queue sizes by a diffusion process at such intraday time scales, as proposed in Avellaneda et al. (2011). The parameters involved in this approximation are straightforward to estimate from empirical data: they involve estimating first and second moments of durations and order sizes.

Example 3. Set for instance $\gamma_1 = 30$ seconds and $\gamma_0 = (\mathbb{E}[T^a] + \mathbb{E}[T^b])/2$. The following table shows the parameters (19) estimated from high frequency records or order book events for three liquid US stocks.

	Std deviation of Bid queue	Std deviation of Ask queue	μ_b	μ_a	ρ
Citigroup	6256	4457	-1033	-2467	0.07
General Electric	2156	2928	-334	-1291	0.03
General Motors	578	399	+78	-96	- 0.04

Table 5 Parameters for the heavy-traffic approximation of bid / ask queues over a 30-second time scale. The unit is a number of orders per period of 30 seconds.

In particular we observe that the order of magnitude of the standard deviation of queue lengths is an order of magnitude larger than their expected change.

Example 4. Theorem 2 may also be used to derive jump-diffusion approximations for the limit order book in theoretical models such as the ones presented in Section 2.3. Let us illustrate this in the case of the heterogeneous trader model of Section 2.3.4.

Let $(T_i, i \geq 1)$ be the sequence of duration between consecutive orders. We assume that this sequence is a sequence of stationary random variables with $\mathbb{E}[T_1] < \infty$. We also assume that every trader has an equal chance of being a buyer or a seller and that the type of trader (buyer or seller) is independent from the past:

$$\mathbb{P}[i-\text{th trader is a buyer}] = \mathbb{P}[i-\text{th trader is a seller}] = \frac{1}{2}$$

Finally the sequence of number of orders $(V_i, i \geq 1)$ is a stationary sequence of orders traded by the i-th trader with the property that $\mathbb{E}[V_1^2] < \infty$.

This order flow given by $(T_i, i \geq 1)$, $(V_i, i \geq 1)$, and the sequence of type (buyers or sellers, using limit orders, market orders or both) generates a sequence of durations $(T_i^a, i \geq 1)$, $(T_i^b; i \geq 1)$ and order sizes $(V_i^a, i \geq 1)$ and $(V_i^b, i \geq 1)$ which satisfy assumptions 1 and 2.

The sequence of durations $(T_i^a, i \geq 1)$ and $(T_i^b, i \geq 1)$ are two stationary sequences of random variables with finite mean:

$$\forall i \geq 0, \quad T_i = T_i^a = T_i^b. \text{ therefore } \mathbb{E}[T_i] = \mathbb{E}[T_i^a] = \mathbb{E}[T_i^b] < \infty.$$

The sequence of order sizes $((V_i^b, V_i^a), i \geq 1)$ is a sequences of IID random variables with

$$\mathbb{P}[(V_i^b, V_i^a) = (V_i, 0)] = \mathbb{P}[(V_i^b, V_i^a) = (0, V_i)] = \frac{m}{2}, \tag{20}$$

$$\mathbb{P}[(V_i^b, V_i^a) = (-V_i, 0)] = \mathbb{P}[(V_i^b, V_i^a) = (0, -V_i)] = \frac{l}{2}, \tag{21}$$

$$\mathbb{P}[(V_i^b, V_i^a) = (\gamma V_i, -(1-\gamma)V_i)] = \mathbb{P}[(V_i^b, V_i^a) = (-(1-\gamma)V_i, \gamma V_i)] = \frac{1-l-m}{2}. \tag{22}$$

Theorem 2 then shows that (Q^b, Q^a) is a Markov process which behaves like a two-dimensional Brownian motion with drift (μ_b, μ_a) and covariance matrix Λ inside the positive orthant $\{x > 0\} \cap \{y > 0\}$ where:

$$\mu_b = \mu_a = \frac{\overline{V}}{2\mathbb{E}[T_1]}\left(2m + 2\gamma(1-l-m) - 1\right), \quad \Lambda = v^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \quad \textit{where} \tag{23}$$

$$v^2 = \frac{\mathbb{E}[T_1]\mathbb{E}[V_1^2]}{4}\left(m + l + \frac{\gamma^2 + (1-\gamma)^2}{2}(1-l-m)\right) \quad \textit{and} \quad \rho = -\frac{(1-l-m)^2\gamma(1-\gamma)}{1 + (1-l-m)(\gamma^2 - \gamma - 1/2)} < 0. \tag{24}$$

Figure 14 displays the value of the correlation ρ in different scenarios as a function of γ and the proportion $1-(l+m)$ of traders submitting orders of both types.

5. Price dynamics

5.1. Price dynamics in the heavy traffic limit

Denote by $(s_t^n, t \geq 0)$ the (bid) price process corresponding to the limit order book process $(q_t^n)_{t\geq 0}$. As explained in Section 2, s^n is a piecewise constant stochastic process which

- increases by one tick at each event $(t_i^{a,n}, V_i^{a,n})$ at the ask for which $q^{a,n}(t_i^{a,n}) + V_i^{a,n} < 0$,
- decreases by one tick at each event $(t_i^{b,n}, V_i^{b,n})$ at the bid for which $q^{b,n}(t_i^{b,n}) + V_i^{b,n} < 0$.

Due to the complex dependence structure in the sequence of order durations and sizes, properties of the process s^n are not easy to study, even in simple models such as those given in Section 2.3. The following result shows that the price process converges to a simpler process in the heavy traffic limit, which is entirely characterized by hitting times of the two dimensional Markov process Q:

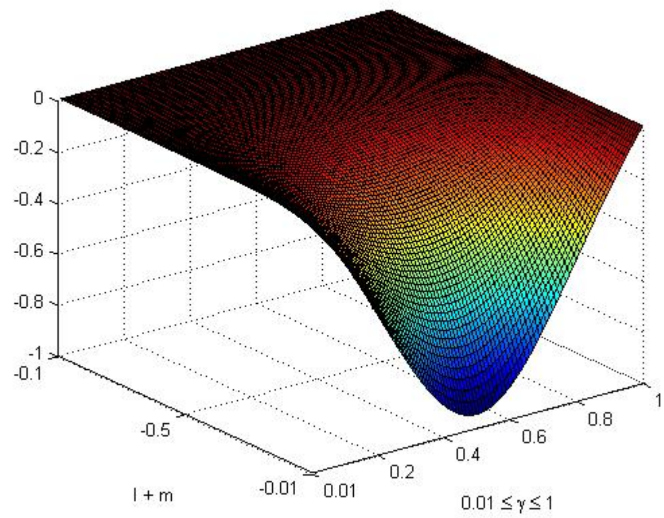

Figure 14 Correlation ρ between bid and ask queue sizes for different scenario. 1 - $(l+m)$ represents the proportion of traders using both market and limit orders, γ the proportion of limit orders and $(1-\gamma)$ the proportion of market orders.

PROPOSITION 1. *Under the assumptions of Theorem 2,*

$$(s^n_{nt}, t \geq 0) \overset{n\to\infty}{\Rightarrow} S, \quad on \quad (D([0,\infty[,\mathbb{R}), M_1), \qquad \text{where}$$

$$S_t = \delta \left(\sum_{0\leq s\leq t} \mathbf{1}_{Q^a_{s-}=0} - \sum_{0\leq s\leq t} \mathbf{1}_{Q^b_{s-}=0} \right). \tag{25}$$

S is a piecewise constant cadlag process which
- *increases by one tick at t if $Q^a_{t-}=0$ and*
- *decreases by one tick at t if $Q^b_{t-}=0$.*

We refer the reader to Whitt (2002) or Whitt (1980) for a description of the M_1 topology. The price process s^n (rescaled in time) can be expressed as

$$s^n_{nt} = \sum_{\tau^n_k \leq t} 1_{\Psi_{k-1}(X^n, Q^n_{\tau^n_1},\ldots,Q^n_{\tau^n_{k-1}})(\tau^n_k).(0,1)<0} - 1_{\Psi_{k-1}(X^n, Q^n_{\tau^n_1},\ldots,Q^n_{\tau^n_{k-1}})(\tau^n_k).(1,0)<0}.$$

where $\tau^n_k, Q^n_{\tau^n_k}$ are defined in the proof of Theorem 2, where it was shown that

$$\forall k \geq 1, \quad (X^n, \tau^n_1, .., \tau^n_k, Q^n_{\tau^n_1}, ..., Q^n_{\tau^n_k}) \overset{n\to\infty}{\Rightarrow} (X, \tau_1, .., \tau_k, Q_{\tau_1}, ..., Q_{\tau_k}).$$

As shown in the proof of Theorem 2, $(X, Q_{\tau_1}, ..., Q_{\tau_k,...})$ lies, with probability 1, in the set of continuity points of Ψ_k for each $k \geq 1$ so

$$\Psi_k(X^n, Q^n_{\tau^n_1}, ..., Q^n_{\tau^n_k}) \overset{n\to\infty}{\Rightarrow} \Psi_k(X, Q_{\tau_1}, ..., Q_{\tau_k}).$$

Applying Lemma 2 and the continuous mapping theorem (Billingsley 1968, Theorem 5.1) then shows that

$$1_{\Psi_{k-1}(X^n, Q^n_{\tau^n_1},\ldots,Q^n_{\tau^n_{k-1}})(\tau^n_k-).(0,1)<0} \overset{n\to\infty}{\Rightarrow} 1_{\Psi_{k-1}(X, Q_{\tau_1},\ldots,Q_{\tau_{k-1}})(\tau_k-).(0,1)<0}$$

The sequences of processes $\sum_{\tau_k^n \le t} 1_{\Psi_{k-1}(X^n, Q^n_{\tau_1^n}, \dots, Q^n_{\tau^n_{k-1}})(\tau_k^n).(0,1)<0}$ and $\sum_{\tau_k^n \le t} 1_{\Psi_{k-1}(X^n, Q^n_{\tau_1^n}, \dots, Q^n_{\tau^n_{k-1}})(\tau_k^n).(1,0)<0}$ belong to $D_\uparrow([0,\infty[,\mathbb{R}_+)$, the set of increasing cadlag trajectories. The convergence for the M_1 topology of sequences in $D_\uparrow$ reduces to the convergence on a dense subset including zeros. Therefore

$$\sum_{\tau_k^n \le t} 1_{\Psi_{k-1}(X^n, Q^n_{\tau_1^n}, \dots, Q^n_{\tau^n_{k-1}})(\tau_k^n).(0,1)<0} \Rightarrow \sum_{\tau_k \le t} 1_{\Psi_{k-1}(X, Q_{\tau_1}, \dots, Q_{\tau_{k-1}})(\tau_k).(0,1)<0}, \quad \text{and}$$

$$\sum_{\tau_k^n \le t} 1_{\Psi_{k-1}(X^n, Q^n_{\tau_1^n}, \dots, Q^n_{\tau^n_{k-1}})(\tau_k^n).(1,0)\le 0} \Rightarrow \sum_{\tau_k \le t} 1_{\Psi_{k-1}(X, Q_{\tau_1}, \dots, Q_{\tau_{k-1}})(\tau_k).(1,0)\le 0}.$$

On the other hand, since the set of discontinuities of $\sum_{\tau_k \le t} 1_{\Psi_k(X, Q_{\tau_1}, \dots, Q_{\tau_{k-1}})(\tau_k).(0,1)<0}$ and $\sum_{\tau_k \le t} 1_{\Psi_k(X, Q_{\tau_1}, \dots, Q_{\tau_{k-1}})(\tau_k).(1,0)<0}$ have an intersection which is almost surely void, one can apply (Whitt 1980, Theorem 4.1) and (Whitt 2002, Theorem 12.7.1) and

$$(s^n_{nt}, t \ge 0) \overset{n\to\infty}{\Rightarrow} S. \quad \text{on} \quad (D([0,\infty[,\mathbb{R}_+), M_1).$$

S is thus the difference between the occupation time of the y axis and the occupation time of the x axis by the Markov process Q. In particular, this result shows that, in a market where order arrivals are frequent, distributional properties of the price process s^n may be approximated using the distributional properties of its limit S. We will now use this result to obtain some analytical results on the distribution of durations between price changes and the transition probabilities of the price.

5.2. Duration between price moves

Starting from an initial order book configuration $Q_0 = (x, y)$,

- the next price increase occurs at the first hitting time of the x-axis by $(Q_{t-})_{t\ge 0}$:

$$\tau_a = \inf\{t \ge 0, Q^a_{t-} = 0\}$$

- the next price decrease occurs at the first hitting time of the y-axis by $(Q_{t-})_{t\ge 0}$:

$$\tau_b = \inf\{t \ge 0, \quad Q^b_{t-} = 0\}.$$

The *duration* τ until the next price changes is then given by

$$\tau = \tau_a \wedge \tau_b,$$

which has the same law as the first exit time from the positive orthant of a two-dimensional Brownian motion with drift. Using the results of Metzler (2010), Lipton (2001), Zhou (2001) we obtain the following result which relates the distribution of this *duration* to the state of the order book and the statistical features of the order flow process, for a balanced order flow where $\overline{V^a} = \overline{V^b} = 0$.

Proposition 2 **(Conditional distribution of duration between price changes)**. *In the case of balanced order flow where $\overline{V^a} = \overline{V^b} = 0$ the distribution of the duration τ until the next price change, conditional on the current state of the bid and ask queues, is given by*

$$\mathbb{P}[\tau > t | Q_0^b = x, Q_0^a = y] = \sqrt{\frac{2U}{\pi t}} e^{-\dfrac{U}{4t}} \sum_{n=0}^{\infty} \frac{1}{(2n+1)} \sin \frac{(2n+1)\pi\theta_0}{\alpha} (I_{(\nu_n-1)/2}(\frac{U}{4t}) + I_{(\nu_n+1)/2}(\frac{U}{4t})),$$

where $\nu_n = (2n+1)\pi/\alpha$, I_n is the nth Bessel function,

$$U = \frac{(\frac{x}{\lambda_a v_a^2})^2 + (\frac{y}{\lambda_b v_b^2})^2 - 2\rho\frac{xy}{\lambda_a \lambda_b v_a^2 v_b^2}}{(1-\rho)}, \qquad \text{and}$$

$$\alpha = \begin{cases} \pi + \tan^{-1}(-\frac{\sqrt{1-\rho^2}}{\rho}) & \rho > 0 \\ \frac{\pi}{2} & \rho = 0 \\ \tan^{-1}(-\frac{\sqrt{1-\rho^2}}{\rho}) & \rho < 0 \end{cases} \text{ and } \quad \theta_0 = \begin{cases} \pi + \tan^{-1}(-\frac{y\sqrt{1-\rho^2}}{x-\rho y}) & x < \rho y \\ \frac{\pi}{2} & x = \rho y \\ \tan^{-1}(-\frac{y\sqrt{1-\rho^2}}{x-\rho y}) & x > \rho y \end{cases} \tag{26}$$

In particular, τ is regularly varying with tail index $\frac{\pi}{2\alpha}$.

When $\overline{V^a} = \overline{V^b} = 0$, the process Q behaves like a two-dimensional Brownian motion Z with covariance matrix given by (18) up to the first hitting time of the axes, so the distribution of the duration τ has the same law as the first exit time of Z from the orthant:

$$\tau \overset{d}{=} \inf\{t \geq 0, Q^a_{t-} < 0 \quad \text{or } Q^b_{t-} < 0\}$$

Using the results of Iyengar (1985), corrected by Metzler (2010) for the distribution of the first exit time of a two-dimensional Brownian motion from the orthant we obtain the result.

A result of Spitzer (1958) then shows that

$$\mathbb{E}[\tau^\beta | Q^b_0 = x, Q^a_0 = y] = \int_0^\infty t^{\beta - a}\mathbb{P}[\tau > t | Q^b_0 = x, Q^a_0 = y]dt < \infty$$

if and only if $\beta < \pi/2\alpha$, where α is defined in (26). Therefore the tail index of τ is $\frac{\pi}{2\alpha}$. This result does not depend on the initial state (x, y).

- If $\rho = 0$, the two components of the Brownian Motion are independent and τ is a regularly-varying random variable with tail index 1. This random variable does not have a moment of order one.
- If $\rho < 0$, $\frac{\pi}{2\alpha} > 1$ and τ has a finite moment of order one. In practice, $\rho \approx -0.7$; this means that if $\mu_a = 0$ and $\mu_b = 0$, the tail index of τ is around 2.
- When $\rho > 0$, $\frac{\pi}{2\alpha} < 1$. The tail of τ is very heavy; τ does not have a finite moment of order one.

For all high frequency data sets examined, the estimates for μ_a, μ_b are negative (see Section 4.3); the durations then have finite moments of all orders.

Remark 2. Using the results of (Lipton 2001, Eq.(12.87)) (see also Zhou (2001)) on the first exit time of a two-dimensional Brownian motion with drift, one can generalize the above results to the case where $(\overline{V^b}, \overline{V^a}) \neq (0,0)$: we obtain in that case

$$\mathbb{P}[\tau > t | Q^b_0 = x, Q^a_0 = y] = \frac{2e^{a_1 x + a_2 y + a_t t - r_0^2/2t}}{\alpha t}\sum_{n=1}^{\infty}\sin\left(\frac{n\pi\theta_0}{\alpha}\right)\int_0^\alpha \sin\left(\frac{n\pi}{\alpha}\right)g_n(\theta)d\theta \tag{27}$$

where θ_0, α are defined as above, $r_0 = \sqrt{U}$ and

$$g_n(\theta) = \int_0^\infty re^{-r^2/2t}e^{d_1 r\sin(\theta-\alpha) - d_2 r\cos(\theta-\alpha)}I_{n\pi/\alpha}(\frac{rr_0}{t})dr,$$

$$d_1 = \left(a_1\sqrt{\lambda^a}v_a + \rho a_2 v_b\sqrt{\lambda^b}\right), \quad d_2 = \left(\rho a_1\sqrt{\lambda^a}v_a + a_2 v_b\sqrt{\lambda^b}\right) \tag{28}$$

$$a_1 = -\frac{\mu_a\sqrt{\lambda^b}v_b + \mu_b\rho v_a\sqrt{\lambda^a}}{(1-\rho^2)\sigma_a^2\lambda^a\sqrt{\lambda^b}v_b}, \qquad a_2 = -\frac{\rho\mu_a\sqrt{\lambda^b}v_b + \mu_b v_a\sqrt{\lambda^a}}{(1-\rho^2)\sigma_b^2\lambda^b\sqrt{\lambda^a}v_a}, \tag{29}$$

and
$$a_t = \left(a_1\frac{\lambda^a v_a^2}{2} + a_2\frac{\lambda^b v_b^2}{2} + 2\rho a_1\sqrt{\lambda^a\lambda^b}v_a a_2 v_b\right) - a_1\mu_a - a_2\mu_b. \tag{30}$$

5.3. Probability of a price increase

A useful quantity for short-term prediction of intraday price moves is the probability $p_1^{up}(x,y)$ that the price will increase at the next move given x orders at the bid and y orders at the ask; in our setting this is equal to the probability that the ask queue gets depleted before the bid queue.

In the heavy traffic limit, this quantity may be represented as the probability that the two-dimensional process $(Q_t, t \geq 0)$, starting from an initial position (x,y), hits the horizontal axis before hitting the vertical axis:

$$p_1^{up}(x,y) = \mathbb{P}[\tau_a < \tau_b | (Q_0^b, Q_0^a) = (x,y)].$$

Since this quantity only involves the process Q up to its first hitting time of the boundary of the orthant, it may be equivalently computed by replacing Q by a two-dimensional Brownian motion with drift and covariance given by (17)–(18).

However, when $\overline{V^a} = \overline{V^b} = 0$, one has a simple analytical solution which only depends on the size x of the bid queue, the size y of the ask queue and the correlation ρ between their increments:

Theorem 3. *Assume $\overline{V^a} + \overline{V^b} \leq 0$. Then $p_1^{up} : \mathbb{R}_+^2 \to [0,1]$ is the unique bounded solution of the Dirichlet problem*

$$\frac{\lambda^a v_a^2}{2}\frac{\partial^2 p_1^{up}}{\partial y^2} + \frac{\lambda^b v_b^2}{2}\frac{\partial^2 p_1^{up}}{\partial x^2} + 2\rho\sqrt{\lambda^a\lambda^b}\sigma^a\sigma^b\frac{\partial^2 p_1^{up}}{\partial x\partial y} + \lambda^a\overline{V^a}\frac{\partial p_1^{up}}{\partial y} + \lambda^b\overline{V^b}\frac{\partial p_1^{up}}{\partial x} = 0 \quad \text{for} \quad x>0, \quad y>0 \tag{31}$$

with the boundary conditions

$$\forall x > 0, \quad p_1^{up}(x,0) = 1 \quad \text{and} \quad \forall y > 0, \quad p_1^{up}(0,y) = 0. \tag{32}$$

When $\overline{V^a} = \overline{V^b} = 0$, $p_1^{\text{up}}(x,y)$ is given by

$$p_1^{\text{up}}(x,y) = \frac{1}{2} - \frac{\arctan\left(\sqrt{\frac{1+\rho}{1-\rho}}\,\frac{\frac{y}{\sqrt{\lambda_a}v_a} - \frac{x}{\sqrt{\lambda_b}v_b}}{\frac{y}{\sqrt{\lambda_a}v_a} + \frac{x}{\sqrt{\lambda_b}v_b}}\right)}{2\arctan\left(\sqrt{\frac{1+\rho}{1-\rho}}\right)}, \tag{33}$$

where $\lambda^a, \lambda^b, v_a$ and v_b are defined in Assumptions 1 and 2.

Using the results of Yoshida and Miyamoto (1999), the Dirichlet problem (31)–(32) has a unique positive bounded solution $u \in C^2(]0,\infty[^2, \mathbb{R}_+) \cap C_b^0(\mathbb{R}_+^2, \mathbb{R}_+)$. Application of Ito's formula to $M_t = u(Q_t^b, Q_t^a)$ then shows that the process M^τ stopped at τ is a martingale, and conditioning with respect to $(Q_0^b, Q_0^a) = (x,y)$ gives $u(x,y) = p^{\text{up}}(x,y)$

Assume now $\overline{V^a} = \overline{V^b} = 0$. Using a change of variable $x \mapsto x\ \sqrt{\lambda^b}v_b$ and $y \mapsto y\ \sqrt{\lambda^a}v_a$, one only needs to consider the case where $\sqrt{\lambda^b}v_b = \sqrt{\lambda^a}v_a$.

Up to the first hitting time of the axes, $(Q_t, t \geq 0)$ is identical in law to $Q = AB$ where

$$A = \begin{pmatrix} \cos(\beta) & \sin(\beta) \\ \sin(\beta) & \cos(\beta) \end{pmatrix},$$

with β satisfying $\rho = \sin(2\beta)$, $\beta \leq \pi/4$ and B a standard planar Brownian Motion with identity covariance. Using polar coordinates $(x,y) = (r\cos\theta, r\sin\theta)$ we have

$$\phi(r,\theta) := p_1^{up}(r\ A^{-1}(\cos(\theta),\sin(\theta)) = p_1^{up}\left(\frac{r}{\cos^2(\beta)-\sin^2(\beta)}(\cos(\beta+\theta),\sin(\theta-\beta))\right)$$

is a solution of the Dirichlet problem

$$\frac{1}{r}\frac{\partial}{\partial r}(r\frac{\partial\phi}{\partial r}) + \frac{1}{r^2}\frac{\partial^2\phi}{\partial\theta^2} = 0 \tag{34}$$

in the cone $C = \{(r,\theta), \qquad r>0, \theta \in]-\beta, \frac{\pi}{2}-\beta[\}$, with the boundary conditions

$$\forall r>0, \qquad \phi(r,-\beta) = 1 \qquad \phi(r,\frac{\pi}{2}-\beta) = 0 \tag{35}$$

A positive bounded solution, which in this case does not depend on r, is given by

$$\phi(r,\theta) = \frac{1}{\pi/2 + \arcsin\rho}(-\theta + \pi/2 + \arcsin(\rho)/2),$$

where ρ is the correlation coefficient between the bid and ask queues. By (Yoshida and Miyamoto 1999, Theorem 3.2), the Dirichlet problem (34)–(35) has a unique bounded solution, so finally

$$p_1^{up}(x,y) = \frac{1}{\pi/2 + \arcsin\rho}\left(\pi/2 + \arcsin(\rho)/2 - \arctan(\frac{\sin(\arctan(y/x)-\beta)}{\cos(\beta+\arctan(y/x))})\right).$$

Remark 3. When $\sqrt{\lambda^a}v_a = \sqrt{\lambda^b}v_b$, the probability $p_1^{up}(x,y)$ only depends on the ration y/x and on the correlation ρ

$$p_1^{\mathrm{up}}(x,y) = \frac{1}{2} - \frac{\arctan(\sqrt{\frac{1+\rho}{1-\rho}}\frac{y-x}{y+x})}{2\arctan(\sqrt{\frac{1+\rho}{1-\rho}})}, \tag{36}$$

and when $\rho = 0$ (which is the case for some empirical examples, see Section 4.3),

$$p_1^{up}(x,y) = \frac{2}{\pi}\arctan(\frac{y}{x}).$$

Figure 15 displays the dependence of the uptick probability p_1^{up} on the bid-ask imbalance variable $\theta = \arctan(y/x)$ for different values of ρ.

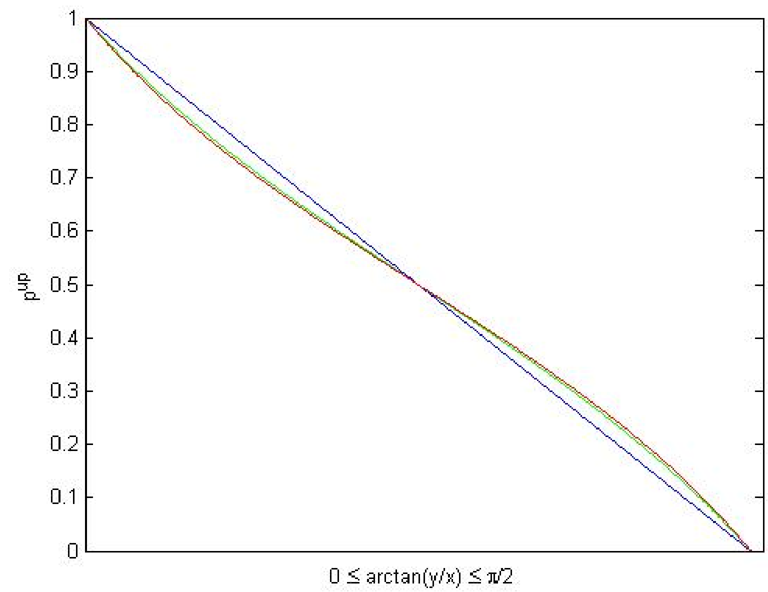

Figure 15 p_1^{up} as a function of the bid-ask imbalance variable $\theta = \arctan(y/x)$ for $\rho = 0$ (blue line), $\rho = -0.7$ (green line) and $\rho = -0.9$ (red line).

6. Appendix: Technical Proofs

6.1. A J_1-continuity property

LEMMA 2. *Let $\tau : D([0,\infty),\mathbb{R}^2) \mapsto [0,\infty[$ be the first exit time from the positive orthant. The map*

$$G : (D([0,\infty),\mathbb{R}^2), J_1) \to \mathbb{R} \qquad \omega \to 1_{\omega(\tau(\omega)).(0,1)<0}. \tag{37}$$

is continuous on the set $\{\omega \in C([0,\infty),\mathbb{R}^2\backslash\{(0,0)\}), \quad \tau(\omega)<\infty\}$.

When $\tau(\omega)<\infty$, $G(\omega)=1$ indicates that ω first exits the orthant by crossing the x-axis. To prove this property, first note that

$$C([0,\infty),\mathbb{R}^2\backslash\{(0,0)\}) = \bigcup_{n\geq 1} C_0([0,\infty),\mathbb{R}^2\backslash B(0,1/n)).$$

Let $\omega_0 \in C([0,\infty),\mathbb{R}^2\backslash\{(0,0)\})$. There exists $n\in\mathbb{N}$ such that $\omega_0 \notin B(0,1/n)$. Let $\epsilon>0$ such that $\epsilon+\eta_{\omega_0}(\epsilon)+\eta_{\omega_0\circ\lambda}(\epsilon)<1/n$, where η_ω is the modulus of continuity of ω. Let $\omega' \in D([0,\infty),\mathbb{R}^2)$ with $d_{J_1}(\omega_0,\omega')\leq\epsilon$. There exists $\lambda:[0,T]\to[0,T]$ increasing such that

$$||\omega_0\circ\lambda-\omega||_\infty \leq \epsilon \quad \text{and} \quad ||\lambda-e||_\infty\leq\epsilon.$$

Without loss of generality, one can also assume, by J_1-continuity of τ at ω_0, that

$$|\tau(\omega_0)-\tau(\omega)|\leq\epsilon.$$

Now, we will show that $|\omega_0(\tau(\omega_0))-\omega'(\tau(\omega'))|\leq\epsilon+\eta_{\omega_0}(\epsilon)+\eta_{\omega_0\circ\lambda}(\epsilon)$:

$$|\omega_0(\tau(\omega_0))-\omega'(\tau(\omega'))| = |\omega_0(\tau(\omega_0))-\omega_0\circ\lambda(\tau(\omega'))+\omega_0\circ\lambda(\tau(\omega'))-\omega_0\circ\lambda(\tau(\omega_0))+\omega_0\circ\lambda(\tau(\omega_0))-\omega'(\tau(\omega'))|,$$

therefore

$$\begin{aligned}|\omega_0(\tau(\omega_0))-\omega'(\tau(\omega'))| &\leq ||\omega_0\circ\lambda-\omega'||_\infty+|\omega_0\circ\lambda(\tau(\omega'))-\omega_0\circ\lambda(\tau(\omega_0))|+|\omega_0\circ\lambda(\tau(\omega_0))-\omega_0(\tau(\omega_0))|\\ &\leq \epsilon+\eta_{\omega_0}(\epsilon)+\eta_{\omega_0\circ\lambda}(\epsilon).\end{aligned}$$

Since $\epsilon+\eta_{\omega_0}(\epsilon)+\eta_{\omega_0\circ\lambda}(\epsilon)<1/n$ and $\omega_0\notin B(0,1,n)$, $1_{\tau(\omega_0).(0,1)<0}=1_{\tau(\omega').(0,1)<0}$, which completes the proof of the continuity of the map G on the space $C([0,\infty),\mathbb{R}^2\backslash\{(0,0)\})$.

6.2. Continuity of Ψ: proof of Theorem 1

To study the continuity of the map Ψ, we endow $D([0,\infty),\mathbb{R}^2)$ with Skorokhod's J_1 topology (see Lindvall (1973), Whitt (1980)). Let Λ_T the set of continuous, increasing functions $\lambda:[0,T]\to[0,T]$ and e the identical function on $[0,T]$. Recall that the following metric

$$d_{J_1}(\omega_1,\omega_2)=\inf_{\lambda\in\Lambda}\left(||\omega_2\circ\lambda-\omega_1||_\infty+||\lambda-e||_\infty\right).$$

defined for $\omega_1,\omega_2\in D([0,T],\mathbb{R}^2)$, induces the J_1 topology on $D([0,T],\mathbb{R}^2)$, and $\omega_n\to\omega$ in $(D([0,\infty),\mathbb{R}^2),J_1)$ if for every continuity point T of ω, $\omega_n\to\omega$ in $(D([0,T],\mathbb{R}^2),J_1)$.

The set $\left(\mathbb{R}_+^2\right)^{\mathbb{N}}$ is endowed with the topology induced by 'cylindrical' semi-norms, defined as follows: for a sequence $(R^n)_{n\geq1}$ in $\left(\mathbb{R}_+^2\right)^{\mathbb{N}}$

$$R^n \overset{n\to\infty}{\to} R \quad \Longleftrightarrow \forall k\geq1, \quad \sup\{|R_1^n-R_1|,...,|R_k^n-R_k|)\overset{n\to\infty}{\to} 0.$$

$D([0,\infty),\mathbb{R}^2)\times(\mathbb{R}_+^2)^{\mathbb{N}}\times(\mathbb{R}_+^2)^{\mathbb{N}}$ is then endowed with the corresponding product topology. The goal of this section is to characterize the continuity set of the map

$$\Psi : D([0,\infty),\mathbb{R}^2)\times(\mathbb{R}_+^2)^{\mathbb{N}}\times(\mathbb{R}_+^2)^{\mathbb{N}} \mapsto D([0,\infty),\mathbb{R}_+^2)$$

introduced in Definition 1. Let us introduce $C([0,\infty),\mathbb{R}^2\backslash\{(0,0)\})$ be the space of continuous planar paths avoiding the origin:

$$C([0,\infty),\mathbb{R}^2\backslash\{(0,0)\}) = \bigcup_{n\geq 1} C_0([0,T],\mathbb{R}^2\backslash B(0,1/n)).$$

Lemma 3. *Let $\omega\in C([0,\infty),\mathbb{R}^2\backslash\{(0,0)\})$. Then the map*

$$\begin{aligned}\Psi_1 : D([0,\infty),\mathbb{R}^2)\times\mathbb{R}_+\times\mathbb{R}_+ &\to D([0,\infty),\mathbb{R}^2) \\ (\omega,R_1,\tilde{R}_1) &\mapsto \omega + 1_{[\tau(\omega),\infty)}\left(1_{\sigma_b(\omega)=\tau(\omega)}(R_1-\omega_{\tau(\omega)}) + 1_{\sigma_a(\omega)=\tau(\omega)}(\tilde{R}_1-\omega_{\tau(\omega)})\right),\end{aligned} \tag{38}$$

where

$$\sigma_b(\omega) = \inf\{t\geq 0,\ \omega_t.(0,1)<0\},\quad \sigma_a(\omega) = \inf\{t\geq 0,\ \omega_t.(1,0)<0\}\quad \textit{and}\quad \tau(\omega) = \sigma_b(\omega)\wedge\sigma_a(\omega).$$

is continuous at ω with respect to the following distance on $(D([0,\infty],\mathbb{R}^2)\times\mathbb{R}_+\times\mathbb{R}_+)$:

$$d((\omega,R_1,\tilde{R}_1),(\omega',R_1',\tilde{R}_1')) = d_{J_1}(\omega,\omega') + |R_1-R_1'| + |\tilde{R}_1-\tilde{R}_1|$$

Let $(\omega_0,R_1,\tilde{R}_1)\in C([0,\infty),\mathbb{R}^2\backslash\{(0,0)\})\times\mathbb{R}_+^2$, $(\omega',R_1',\tilde{R}_1')\in D(0,\mathbb{R}^2)\times\mathbb{R}_+^2$. Since $\omega_0\in C([0,\infty),\mathbb{R}^2\backslash\{(0,0)\})$, there exists $n>0$ such that $\omega_0\notin B(0,1/n)$. Let $0<\epsilon<1/n$ such that

$$d((\omega_0,R_1,\tilde{R}_1),(\omega',R_1',\tilde{R}_1'))<\epsilon.$$

Since $d_{J_1}(\omega_0,\omega')<\epsilon$, there exists $\lambda:[0,T]\to[0,T]$, non-decreasing such that:

$$||\lambda-e||_\infty<\epsilon,\quad \text{and}\quad ||\omega_0\circ\lambda-\omega||_\infty<\epsilon.$$

By continuity of τ for the J_1 topology Whitt (2002)[Theorem 13.6.4] at ω_0 (since ω_0 is continuous, the J_1 and M_1 topologies are identical at this point), one can also assume, without loss of generality, that

$$|\tau(\omega_0\circ\lambda)-\tau(\omega')|\leq\epsilon.$$

Moreover, since the graph of ω_0 does not intersect with $\overline{B(0,1/n)}$ and $\epsilon<1/n$, $1_{\tau(\omega_0)=\sigma_a(\omega_0)} = 1_{\tau(\omega')=\sigma_a(\omega')}$. Now define λ^ϵ by

$$\begin{aligned}\lambda^\epsilon : [0,T] &\to [0,T] \\ t &\mapsto \frac{\tau(\omega')}{\tau(\omega_0\circ\lambda)}\lambda_t.\end{aligned} \tag{39}$$

Then

$$\begin{aligned}||\lambda^\epsilon-e||_\infty = ||\frac{\tau(\omega)}{\tau(\omega_0\circ\lambda)}\lambda-e||_\infty &\leq ||\frac{\tau(\omega)}{\tau(\omega_0\circ\lambda)}\lambda-\frac{\tau(\omega)}{\tau(\omega_0\circ\lambda)}e||_\infty + ||\frac{\tau(\omega)}{\tau(\omega_0\circ\lambda)}e-e||_\infty \\ &\leq \epsilon\frac{\tau(\omega)}{\tau(\omega_0\circ\lambda)} + \frac{\epsilon}{\tau(\omega_0\circ\lambda)}.\end{aligned}$$

On the other hand

$$||\omega_0 \circ \lambda^\epsilon - \omega||_\infty = ||\omega_0 \circ \lambda^\epsilon - \omega_0 \circ \lambda + \omega_0 \circ \lambda - \omega||_\infty \leq ||\omega_0 \circ \lambda^\epsilon - \omega_0 \circ \lambda||_\infty + \epsilon \leq \eta_{\omega_0 \circ \lambda}(\epsilon) + \epsilon,$$

where $\eta_{\omega_0 \circ \lambda}$ is the modulus of continuity modulus of $\omega_0 \circ \lambda$. Therefore, since $1_{\tau(\omega_0 \circ \lambda^\epsilon)} = 1_{\tau(\omega')}$ by definition of λ^ϵ and

$$\begin{aligned}\Psi_1(\omega_0, R_1, \tilde{R}_1) \circ \lambda^\epsilon - \Psi_1(\omega', R_1', \tilde{R}_1')) &= \omega_0 \circ \lambda^\epsilon - \omega' \\ &+ 1_{\tau(\omega_0 \circ \lambda^\epsilon)} \left(1_{\tau(\omega') = \sigma_a}(R_1' - R_1) + 1_{\tau(\omega') = \sigma_b}(\tilde{R}_1' - \tilde{R}_1)\right).\end{aligned}$$

Thus λ^ϵ satisfies $||\lambda^\epsilon - e|| \leq \epsilon(\frac{\tau(\omega')+1}{\tau(\omega_0 \circ \lambda)})$ and

$$||\Psi_1(\omega_0, R_1, \tilde{R}_1) \circ \lambda^\epsilon - \Psi_1(\omega', R_1', \tilde{R}_1'))||_\infty \leq \eta_{\omega_0 \circ \lambda}(\epsilon) + \epsilon + 2\epsilon$$

which proves that $(\omega_0, R_1, \tilde{R}_1)$ is a continuity point for Ψ_1.

For $k \geq 2$, define recursively the maps

$$\begin{aligned}\Psi_k : D([0,\infty), \mathbb{R}^2) \times \mathbb{R}_+^{\mathbb{N}} \times \mathbb{R}_+^{\mathbb{N}} &\to D([0,\infty), \mathbb{R}^2) \\ (\omega, (R_i, \tilde{R}_i)_{i \geq 1}) &\mapsto \Psi_1(\Psi_{k-1}(\omega, (R_i, \tilde{R}_i)_{i=1..k-1}), R_k, \tilde{R}_k).\end{aligned} \tag{40}$$

To simplify notation we will denote the argument of Ψ_k as $(\omega, R, \tilde{R})(= (\omega, (R_i, \tilde{R}_i)_{i \geq 1})$ although it is easily observed from (40) that Ψ_k only depends on the first k elements $(R_i, \tilde{R}_i)_{i=1..k})$ of $R, \tilde{R}$.

LEMMA 4. *If $(\omega, R, \tilde{R}) \in C([0,\infty), \mathbb{R}^2 \backslash \{(0,0)\}) \times \mathbb{R}_+^{\mathbb{N}} \times \mathbb{R}_+^{\mathbb{N}}$ such that*

$$(0,0) \notin \Psi_k(\omega, R, \tilde{R})([0,\infty)\ ) \tag{41}$$

then Ψ_k is continuous at $(\omega, R, \tilde{R})$.

Let $(R_i, \tilde{R}_i)_{i \geq 1}$, $(R_i', \tilde{R}_i')_{i \geq 1}$, two sequences of random variables on $\mathbb{R}_+^2$ and define

$$\Omega_k(R, \tilde{R}) = \cap_{j=0}^{k} \Psi_j(C([0,\infty), \mathbb{R}^2 \backslash \{(0,0)\}), R, \tilde{R})$$

where we have set $\Psi_0 = Id$. Consider $\omega_0 \in \Omega_k(R, \tilde{R})$, and $\omega \in D([0,T], \mathbb{R}_+^2)$, such that:

$$d_{J_1}(\omega_0, \omega) + \sup_{i=1..k} |R_i - R_i'| + \sup_{i=1..k} |\tilde{R}_i - \tilde{R}_i'| \leq \epsilon.$$

An application of the triangle inequality yields

$$\begin{aligned}&d_{J_1}(\Psi_k(\omega_0, (R_i, \tilde{R}_i)), \Psi_k(\omega', (R_i', \tilde{R}_i'))) \\ \leq\ &d_{J_1}(\Psi_k(\omega_0, (R_i, \tilde{R}_i)), \Psi_k(\omega', (R_i, \tilde{R}_i))) + d_{J_1}(\Psi_k(\omega', (R_i, \tilde{R}_i)), \Psi_k(\omega', (R_i', \tilde{R}_i')))\end{aligned}$$

where the last term converges to zero when ϵ goes to zero by continuity of Ψ_1.

We can now prove Theorem 1.

Proof of Theorem 1. Since ω is continuous, the jumps of $\Psi(\omega, R, \tilde{R})$ correspond to the first exit times from the orthant of the paths $\Psi_k(\omega, R, \tilde{R})$. Therefore, if $(R_n)_{n \geq 1}, (\tilde{R}_n)_{n \geq 1}$ have no accumulation points on the axes, the paths $\Psi(\omega, R, \tilde{R})$ only has a finite number of discontinuities on $[0,T]$ for any $T > 0$. So, for any $T > 0$, there exists $k(T)$ such that $\Psi = \Psi_{k(T)}$. Then thanks to Lemma 4, Ψ is continuous on the set of continuous trajectories whose image has a finite number of discontinuities and does not contain the origin.

6.3. Functional central limit theorem for the net order flow

PROPOSITION 3. *Let $(T_i^{a,n},T_i^{b,n})_{i\geq 1}$ and $(V_i^{a,n},V_i^{b,n})_{i\geq 1}$ be stationary arrays of random variables which satisfy Assumptions 1 and 2. Let $(N_t^{a,n},t\geq 0)$ and $(N_t^{b,n},t\geq 0)$ be the counting processes defined in (15). Then*

$$\left(\sum_{i=1}^{N_{nt}^{a,n}}\frac{V_i^{a,n}}{\sqrt{n}},\sum_{i=1}^{N_{nt}^{b,n}}\frac{V_i^{b,n}}{\sqrt{n}}\right)_{t\geq 0}\overset{J_1}{\underset{n\to\infty}{\Rightarrow}}\left(\Sigma B_t+t(\lambda^a\overline{V^a},\lambda^b\overline{V^b})\right)_{t\geq 0} \tag{42}$$

where B is a standard planar Brownian motion and

$$\Sigma^t\Sigma=\begin{pmatrix}\lambda^a v_a^2 & \rho\sqrt{\lambda^a\lambda^b}v_a v_b\\ \rho\sqrt{\lambda^a\lambda^b}v_a v_b & \lambda^b v_b^2\end{pmatrix}, \tag{43}$$

Proof: First we will prove that the sequence of processes

$$\left(\sum_{i=1}^{[\lambda^a t]}\frac{V_i^{a,n}}{\sqrt{n}},\sum_{i=1}^{[\lambda^b t]}\frac{V_i^{b,n}}{\sqrt{n}}\right)_{t\geq 0}\overset{J_1}{\underset{n\to\infty}{\Rightarrow}}\left(\Sigma B_t+t(\lambda^a\overline{V^a},\lambda^b\overline{V^b})\right)_{t\geq 0}$$

weakly converges in the J_1 topology. Using the Cramer-Wold device, it is sufficient to prove that for $(\alpha,\beta)\in\mathbb{R}^2$,

$$\left(\alpha\sum_{i=1}^{[\lambda^a t]}\frac{V_i^{a,n}}{\sqrt{n}}+\beta\sum_{i=1}^{[\lambda^b t]}\frac{V_i^{b,n}}{\sqrt{n}}\right)_{t\geq 0}\underset{n\to\infty}{\Rightarrow}(\alpha\lambda^a\overline{V^a}+\beta\lambda^b\overline{V^b}))t+\sqrt{(\alpha^2\lambda^a v_a^2+\beta^2\lambda^b v_b^2+2\rho\alpha\beta v_a v_b\sqrt{\lambda^a\lambda^b})}B_t$$

If $\lambda^a\in\mathbb{Q}$ and $\lambda^b\in\mathbb{Q}$, it is possible to find λ such that $\lambda^a/\lambda\in\mathbb{N}$ and $\lambda^b/\lambda\in\mathbb{N}$. Let for all $(i,n)\in\mathbb{N}^2$,

$$W_i^n=\alpha\left(V_{(\lambda^a/\lambda)(i-1)+1}^{a,n}+V_2^{a,n}+...+V_{\lambda^a i/\lambda}^{a,n}\right)+\beta\left(V_{(\lambda^b/\lambda)(i-1)+1}^{b,n}+V_2^{b,n}+...+V_{\lambda^b i/\lambda}^{b,n}\right),$$

then for all $t>0$,

$$\alpha\sum_{i=1}^{[\lambda^a t]}\frac{V_i^{a,n}}{\sqrt{n}}+\beta\sum_{i=1}^{[\lambda^b t]}\frac{V_i^{b,n}}{\sqrt{n}}=\sum_{i=1}^{[\lambda t]}\frac{W_i^n}{\sqrt{n}}.$$

For all $n>0$, $(W_i^n,i\geq 1)$ is a sequence of stationary random variables. Therefore, thanks to (Jacod and Shiryaev 2003, Chap.VIII, Thm 2.29, p.426), and the fact that

$$\mathrm{var}(W_1^n)+2\sum_{i=2}^{\infty}cov(W_1^n,W_i^n)\overset{n\to\infty}{\to}\sigma^2, \tag{44}$$

the sequence of processes $\left(\sum_{i=1}^{[\lambda nt]}\frac{W_i^n}{\sqrt{n}},\ t\geq 0\right)_{n\geq 1}$ converges weakly to a Brownian motion with volatility $\sqrt{\lambda}\sigma$. If $(\lambda^a,\lambda^b)\notin\mathbb{Q}^2$, there exists $(\lambda_n^a,\lambda_n^b)_{n\geq 1}$ such that

$$\lambda_n^a,\lambda_n^b\in\mathbb{Q}\quad and\quad |\lambda_n^a-\lambda^a|\leq\frac{1}{n},\quad |\lambda_n^b-\lambda^b|\leq\frac{1}{n}.$$

As above, one can define an integer λ_n such that $\frac{\lambda_n^a}{\lambda_n}\in\mathbb{Q}$ and $\frac{\lambda_n^b}{\lambda_n}\in\mathbb{Q}$. Let for all $(i,n)\in\mathbb{N}^2$,

$$W_i^n=\alpha\left(V_{(\lambda_n^a/\lambda_n)(i-1)+1}^{a,n}+V_2^{a,n}+...+V_{\lambda_n^a i/\lambda_n}^{a,n}\right)+\beta\left(V_{(\lambda_n^b/\lambda_n)(i-1)+1}^{b,n}+V_2^{b,n}+...+V_{\lambda_n^b i/\lambda_n}^{b,n}\right),$$

One has for all $t > 0$,

$$\alpha \sum_{i=1}^{[\lambda^a t]} \frac{V_i^{a,n}}{\sqrt{n}} + \beta \sum_{i=1}^{[\lambda^b t]} \frac{V_i^{b,n}}{\sqrt{n}} = \sum_{i=1}^{[\lambda_n t]} \frac{W_i^n}{\sqrt{n}} + \alpha \sum_{i=1}^{[\lambda^a t - \lambda_n^a t]} \frac{V_i^{a,n}}{\sqrt{n}} + \beta \sum_{i=1}^{[\lambda^b t - \lambda_n^b t]} \frac{V_i^{b,n}}{\sqrt{n}}.$$

Moreover

$$\left(\alpha \sum_{i=1}^{[\lambda^a t - \lambda_n^a t]} \frac{V_i^{a,n}}{\sqrt{n}} + \beta \sum_{i=1}^{[\lambda^b t - \lambda_n^b t]} \frac{V_i^{b,n}}{\sqrt{n}} \right)_{t \geq 0} \Rightarrow^{J_1} 0,$$

therefore the convergence above holds even if λ^a or λ^b are not rationals. On one hand,

$$\begin{aligned}
\mathrm{var}(W_i^n) = &\quad \mathrm{var}\left(\alpha(V^{a,n}_{(\lambda_n^a/\lambda_n)(i-1)+1} + \ldots + V^{a,n}_{\lambda_n^a i/\lambda_n}) + \beta(V^{b,n}_{(\lambda_n^b/\lambda_n)(i-1)+1} + V_2^{b,n} + \ldots + V^{b,n}_{\lambda_n^b i/\lambda_n})\right) \\
= &\quad \alpha^2 \mathrm{var}\left(V^{a,n}_{(\lambda_n^a/\lambda_n)(i-1)+1} \ldots + V^{a,n}_{\lambda_n^a i/\lambda_n})\right) + \beta^2 \mathrm{var}\left(V^{b,n}_{(\lambda_n^b/\lambda_n)(i-1)+1} \ldots + V^{b,n}_{\lambda_n^b i/\lambda_n})\right) \\
+ &\quad 2\alpha\beta \mathrm{cov}\left(V^{a,n}_{(\lambda_n^a/\lambda_n)(i-1)+1} \ldots + V^{a,n}_{\lambda_n^a i/\lambda_n}, V^{b,n}_{(\lambda_n^b/\lambda_n)(i-1)+1} \ldots + V^{b,n}_{\lambda_n^b i/\lambda_n})\right).
\end{aligned}$$

On the other hand, for all $i \geq 2$,

$$\begin{aligned}
\mathrm{cov}(W_1^n, W_i^n) = &\quad \alpha^2 \mathrm{cov}\left(V_1^{a,n} + \ldots + V^{a,n}_{(\lambda_n^a/\lambda_n)}, V^{a,n}_{(\lambda_n^a/\lambda_n)(i-1)+1} + \ldots + V^{a,n}_{\lambda_n^a i/\lambda_n}\right) \\
+ &\quad \beta^2 \mathrm{cov}\left(V_1^{b,n} + \ldots + V^{b,n}_{(\lambda_n^b/\lambda_n)}, V^{b,n}_{(\lambda_n^b/\lambda_n)(i-1)+1} + \ldots + V^{b,n}_{\lambda_n^b i/\lambda_n}\right) \\
+ &\quad \alpha\beta \mathrm{cov}\left(V_1^{a,n} + \ldots + V^{a,n}_{(\lambda_n^a/\lambda_n)}, V^{b,n}_{(\lambda_n^b/\lambda_n)(i-1)+1} + \ldots + V^{b,n}_{\lambda_n^b i/\lambda_n}\right) \\
+ &\quad \alpha\beta \mathrm{cov}\left(V_1^{b,n} + \ldots + V^{b,n}_{(\lambda_n^b/\lambda_n)}, V^{a,n}_{(\lambda_n^a/\lambda_n)(i-1)+1} + \ldots + V^{a,n}_{\lambda_n^a i/\lambda_n}\right).
\end{aligned}$$

Therefore

$$\begin{aligned}
\mathrm{var}(W_1^n) + 2\sum_{i=2}^{\infty} \mathrm{cov}(W_1^n, W_i^n) = &\quad \mathrm{var}(V_1^{a,n}) \frac{\lambda_n^a}{\lambda_n} + 2\sum_{i=2}^{\infty} \mathrm{cov}(V_1^{a,n}, V_i^{a,n}) \frac{\lambda_n^a}{\lambda_n} \\
+ &\quad \mathrm{var}(V_1^{b,n}) \frac{\lambda_n^b}{\lambda_n} + 2\sum_{i=2}^{\infty} \mathrm{cov}(V_1^{b,n}, V_i^{b,n}) \frac{\lambda_n^b}{\lambda_n} \\
+ &\quad 2\alpha\beta \mathrm{cov}\left(V_1^{a,n} \ldots + V^{a,n}_{\lambda_n^a/\lambda_n}, V_1^{b,n} \ldots + V^{b,n}_{\lambda_n^b/\lambda_n}\right) \\
+ &\quad 2\alpha\beta \sum_{i=2}^{\infty} \mathrm{cov}\left(V_1^{a,n} \ldots + V^{a,n}_{\lambda_n^a/\lambda_n}, V^{b,n}_{(\lambda_n^b/\lambda_n)(i-1)+1} + \ldots + V^{b,n}_{\lambda_n^b i/\lambda_n}\right) \\
+ &\quad 2\alpha\beta \sum_{i=2}^{\infty} \mathrm{cov}\left(V_1^{b,n} \ldots + V^{b,n}_{\lambda_n^b/\lambda_n}, V^{a,n}_{(\lambda_n^a/\lambda_n)(i-1)+1} + \ldots + V^{a,n}_{\lambda_n^a i/\lambda_n}\right)
\end{aligned}$$

A simple calculation shows that

$$\begin{aligned}
&2\alpha\beta \mathrm{cov}\left(V_1^{a,n} \ldots + V^{a,n}_{\lambda_n^a/\lambda_n}, V_1^{b,n} \ldots + V^{b,n}_{\lambda_n^b/\lambda_n}\right) \\
&\qquad + \quad 2\alpha\beta \sum_{i=2}^{\infty} \mathrm{cov}\left(V_1^{a,n} \ldots + V^{a,n}_{\lambda_n^a/\lambda_n}, V^{b,n}_{(\lambda_n^b/\lambda_n)(i-1)+1} + \ldots + V^{b,n}_{\lambda_n^b i/\lambda_n}\right)
\end{aligned}$$

$$+ \quad 2\alpha\beta \sum_{i=2}^{\infty} \mathrm{cov}\left(V_1^{b,n} ... + V^{b,n}_{\lambda_n^b/\lambda_n}, V^{a,n}_{(\lambda_n^a/\lambda_n)(i-1)+1} + ... + V^{a,n}_{\lambda_n^a i/\lambda_n}\right)$$

$$= 2\alpha\beta \max(\frac{\lambda_n^a}{\lambda_n}, \frac{\lambda_n^b}{\lambda_n}) \mathrm{cov}(V_1^{a,n}, V_1^{b,n}) + \qquad 2\alpha\beta \sum_{i=2}^{\infty} \frac{\lambda_n^a}{\lambda_n} \mathrm{cov}(V_1^{a,n}, V_i^{b,n}) + \frac{\lambda_n^b}{\lambda_n} \mathrm{cov}(V_1^{b,n}, V_i^{a,n}).$$

Therefore

$$\lim_{n \mapsto \infty} \mathrm{var}(W_1^n) + 2\sum_{i=2}^{\infty} \mathrm{cov}(W_1^n, W_i^n) = \alpha \frac{\lambda^a}{\lambda} v_a^2 + \beta \frac{\lambda^b}{\lambda} v_b^2 + 2\rho\sqrt{\alpha\beta} \frac{\sqrt{\lambda^a \lambda^b}}{\lambda} v_a v_b,$$

where ρ is given in (11) and

$$\lim_{n \mapsto \infty} \mathbb{E}[W_i^n] = \alpha \frac{\lambda^a}{\lambda} \overline{V^a} + \beta \frac{\lambda^b}{\lambda} \overline{V^b},$$

which completes the proof of the convergence in (44). The law of large numbers for renewal processes implies that the following sequence of processes converges to zero in the J_1 topology Iglehart and Whitt (1971):

$$(N_{nt}^{a,n})_{t \ge 0} \overset{n \to \infty}{\Rightarrow} ([\lambda^a t])_{t \ge 0}, \quad \textit{and} \quad (N_{nt}^{b,n})_{t \ge 0} \overset{n \to \infty}{\Rightarrow} ([\lambda^b t])_{t \ge 0},$$

$$\left(\sum_{i=[\lambda^a t]}^{N_{nt}^{a,n}} \frac{V_i^{a,n}}{\sqrt{i}}, \sum_{i=[\lambda^b t]}^{N_{nt}^{b,n}} \frac{V_i^{b,n}}{\sqrt{i}} \right)_{t \ge 0} \Rightarrow 0 \qquad \text{in the } J_1 \text{ topology.}$$

6.4. Identification of the heavy traffic limit

Lemma 5. *The process Q is a Markov process with values in $\mathbb{R}_+^2$ and infinitesimal generator $(\mathcal{G}, \mathrm{dom}(\mathcal{G}))$ given by (13) and*

$$\mathrm{dom}(\mathcal{G}) = \{h \in C^2(]0,\infty[\times]0,\infty[,\mathbb{R}) \cap C^0(\mathbb{R}_+^2, \mathbb{R}), \qquad \forall x > 0, \qquad \forall y > 0,$$
$$h(x,0) = \int_{]0,\infty[^2} h(g((x,0),(u,v))) F(du,dv), \qquad h(0,y) = \int_{]0,\infty[^2} h(g((0,y),(u,v))) \tilde{F}(du,dv)\}$$

To identify the infinitesimal generator of the process, we note that $h \in \mathcal{C}^0(\mathbb{R}_+^2)$ is in the domain of the infinitesimal generator if for all $(x,y) \in \mathbb{R}_+^2$

$$\lim_{t \to 0} \frac{\mathbb{E}[h(Q_t) - h(Q_0) | Q_0 = (x,y)]}{t} < \infty.$$

For $x > 0$, and $y > 0$, a classical computation shows that if $h \in \mathcal{C}^2(]0,\infty[\times]0,\infty[)$,

$$\mathbb{E}[h(Q_t)|Q_0 = (x,y)] = h(x,y) + t\left(\lambda^a \overline{V^a} \frac{\partial h}{\partial x} + \lambda^b \overline{V^b} \frac{\partial h}{\partial y} + \frac{\lambda^a v_a^2}{2} \frac{\partial^2 h}{\partial x^2} + \frac{\lambda^b v_b^2}{2} \frac{\partial^2 h}{\partial y^2} + 2\rho\sqrt{\lambda^a \lambda^b} v_a v_b \frac{\partial^2 h}{\partial x \partial y}\right) + o(t),$$

which leads to equation (13). To examine whether the operator $\mathcal{G}$ is closable on $\mathbb{R}_+^2$ we note that, for $h \in C^2(]0,\infty[\times]0,\infty[) \cap C^0([0,\infty), \mathbb{R}^2)$ and $(x,y) \in \mathbb{R}_+^2$,

$$\begin{aligned}
\mathbb{E}[h(Q_t)|Q_0 = (x,0)] &= \int_{]0,\infty[^2} \mathbb{E}[h(Q_t)|Q_{0+} = g((x,0),(u,v))] F(du,dv) \\
&= \int_{]0,\infty[^2} \left(\mathbb{E}[h(Q_t)|Q_{0+} = g((x,0),(u,v))] - h(g((x,0),(u,v))) + h(g((x,0),(u,v)))\right) F(du,dv) \\
&= \int_{\mathbb{R}_+^2} \left(t\mathcal{G}h(g((x,0),(u,v))) + h(g((x,0),(u,v)))\right) F(du,dv) + o(t).
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[h(Q_t)|Q_0=(0,y)] &= \int_{]0,\infty[^2} \mathbb{E}[h(Q_t)|Q_{0+}=g((0,y),(u,v))]\tilde{F}(du,dv) \\
&= \int_{]0,\infty[^2} \left(\mathbb{E}[h(Q_t)|Q_{0+}=g((0,y),(u,v))] - h(g((0,y),(u,v))) + h(g((0,y),(u,v)))\right)\tilde{F}(du,dv) \\
&= \int_{]0,\infty[^2} \left(t\mathcal{G}h(g((0,y),(u,v))) + h(g((0,y),(u,v)))\right)\tilde{F}(du,dv) + o(t).
\end{aligned}$$

so as $t \to 0$, we have

$$\begin{aligned}
\frac{\mathbb{E}[h(Q_t)|Q_0=(x,0)] - h(x,0)}{t} &= \int_{]0,\infty[^2} \mathcal{G}h(g((x,0),(u,v)))F(du,dv) \\
&+ \frac{1}{t}\int_{]0,\infty[^2} \left(h(g((x,0),(u,v))) - h(x,0)\right)F(du,dv) + o(1).
\end{aligned}$$

$$\begin{aligned}
\frac{\mathbb{E}[h(Q_t)|Q_0=(0,y)] - h(0,y)}{t} &= \int_{]0,\infty[^2} \mathcal{G}h(g((0,y),(u,v)))\tilde{F}(du,dv) \\
&+ \frac{1}{t}\int_{]0,\infty[^2} \left(h(g((0,y),(u,v))) - h(0,y)\right)\tilde{F}(du,dv) + o(1).
\end{aligned}$$

Thus the limit $t \to 0$ is well defined only if h verifies, for $x > 0, y > 0$,

$$h(x,0) = \int_{]0,\infty[^2} h(g((x,0),(u,v)))F(du,dv), \qquad h(0,y) = \int_{]0,\infty[^2} h(g((0,y),(u,v)))\tilde{F}(du,dv), \quad (45)$$

This is a Wentzell boundary condition (Taira 1991) which corresponds to a jump to the interior whenever the process reaches the boundary of the quadrant. $\mathcal{G}$ is thus closable on the set

$$\mathrm{dom}(\mathcal{G}) = \{h \in \mathcal{C}^2(]0,\infty[\times]0,\infty[) \cap \mathcal{C}^0(\mathbb{R}_+^2), \qquad h \quad \text{verifies} \quad (14)\}$$

and, for $h \in \mathrm{dom}(\mathcal{G})$ we have

$$\mathcal{G}h(x,0) = \int_{]0,\infty[^2} \mathcal{G}h(g((x,0),(u,v)))F(du,dv), \qquad \mathcal{G}h(0,y) = \int_{]0,\infty[^2} \mathcal{G}h(g((0,y),(u,v)))\tilde{F}(du,dv).$$

The elliptic operator defined by the Laplacian on $(0,\infty)^2$ with Wentzell boundary conditions (45) thus admits a closure $(\mathcal{G}, \mathrm{dom}(\mathcal{G}))$ on $\mathbb{R}_+^2$ and verifies the assumptions of Galakhov and Skubachevskii (2001)[Theorem 3.1]. Galakhov and Skubachevskii (2001)[Theorem 3.1] then implies the existence of a $\mathbb{R}_+^2$–valued Feller process Q, unique in law, whose infinitesimal generator $(\mathcal{G}, \mathrm{dom}(\mathcal{G}))$. The limit process Q is thus a $\mathbb{R}_+^2$–valued Markov process associated with this semigroup.

References

Andersen, Allan, Rama Cont, Ekaterina Vinkovskaya. 2010. A point process model for the high-frequency dynamics of a limit order book. Financial Engineering Report 2010-08, Columbia University.

Araman, Victor, Peter Glynn. 2011. Fractional brownian motion with $H < 1/2$ as a limit of scheduled traffic. *Working Paper* **to appear**.

Avellaneda, Marco, Sasha Stoikov, Josh Reed. 2011. Forecasting prices from Level-I quotes in the presence of hidden liquidity. *Algorithmic Finance* **1** 35–43.

Baccelli, François, Guy Fayolle. 1987. Analysis of models reducible to a class of diffusion processes in the positive quarter plane. *SIAM Journal on Applied Mathematics* **47**(6) 1367–1385.

Bacry, Emmanuel, Sylvain Delattre, Marc Hoffman, Jean-Francois Muzy. 2010. Scaling limits for Hawkes processes and application to financial statistics. Working paper.

Bayraktar, Erhan, Ulrich Horst, Ronnie Sircar. 2006. A limit theorem for financial markets with inert investors. *Mathematics of Operations Research* **2006** 33–54.

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.

Bouchaud, J. P., M. Mezard, M. Potters. 2002. Statistical properties of stock order books: empirical results and models. *Quantitative Finance* **2** 251.

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, 57–160.

Bougerol, Philippe, Nico Picard. 1992. Strict stationarity of generalized autoregressive processes. *Ann. Probab.* **4** 1714–1730.

Cont, Rama. 2011. Statistical modeling of high-frequency financial data. *IEEE Signal processing* **28**(5) 16–25. URL http://ssrn.com/abstract=1748022.

Cont, Rama, Adrien de Larrard. 2010. Price dynamics in a markovian limit order market. Working paper. URL http://ssrn.com/abstract=1735338.

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

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

Dai, J. G., Vin Nguyen. 1994. On the convergence of multiclass queueing networks in heavy traffic. *Annals of Applied Probability* **4** 26–42.

Engle, R., J. Russell. 1998. Autoregressive conditional duration: a new model for irregularly-spaced transaction data. *Econometrica* **66** 1127–1162.

Engle, Robert F. 2000. The econometrics of ultra-high-frequency data. *Econometrica* **68**(1) 1–22.

Galakhov, E.I., A.L. Skubachevskii. 2001. On feller semigroups generated by elliptic operators with integro-differential boundary conditions. *Journal of Differential Equations* **176**(2) 315–355.

Ghysels, Eric, Joanna Jasiak. 1998. GARCH for irregularly spaced financial data: the ACD-GARCH model. *Studies in Nonlinear Dynamics and Econometrics* **2**(4) 133–149.

Gopikrishnan, Parameswaran, Vasiliki Plerou, Xavier Gabaix, H. Eugene Stanley. 2000. Statistical properties of share volume traded in financial markets. *Phys. Rev. E* **62** R4493–R4496.

Harrison, J. Michael. 1978. The diffusion approximation for tandem queues in heavy traffic. *Advances in Applied Probability* **10**(4) 886–905.

Harrison, J. Michael, Vien Nguyen. 1993. Brownian models of multiclass queueing networks: Current status and open problems. *Queueing Systems* **13** 5–40.

Hasbrouck, Joel. 2007. *Empirical Market Microstructure*. Oxford University Press.

Hautsch, Nikolaus. 2004. *Modelling Irregularly Spaced Financial Data*. Springer.

Iglehart, Donald L., Ward Whitt. 1970. Multiple channel queues in heavy traffic. I. *Advances in Appl. Probability* **2** 150–177.

Iglehart, Donald L., Ward Whitt. 1971. The equivalence of functional central limit theorems for counting processes and associated partial sums. *The Annals of Mathematical Statistics* **42**(4) pp. 1372–1378.

Iyengar, Satish. 1985. Hitting lines with two-dimensional brownian motion. *SIAM Journal on Applied Mathematics* **45** 983–989.

Jacod, Jean, Albert N. Shiryaev. 2003. *Limit theorems for stochastic processes*. Springer, Berlin.

Kruk, Łukasz. 2003. Functional limit theorems for a simple auction. *Math. Oper. Res.* **28**(4) 716–751.

Lindvall, Torgny. 1973. Weak convergence of probability measures and random functions in the function space $D(0, \infty)$. *Journal of Applied Probability* **10**(1) 109–121.

Lipton, Alex. 2001. *Mathematical methods in foreign exchange*. World Scientific.

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

Mendelson, Haim. 1982. Market behavior in a clearing house. *Econometrica* 1505–1524.

Metzler, Adam. 2010. On the first passage problem for correlated brownian motion. *Statistics & Probability Letters* **80**(5-6) 277–284.

Ramanan, Kavita, Martin I. Reiman. 2003. Fluid and heavy traffic diffusion limits for a generalized processor sharing model. *Ann. Appl. Probab.* **13**(1) 100–139.

Resnick, Sidney. 2006. *Heavy-tail phenomena: Probabilistic and Statistical Modeling*. Springer.

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

Spitzer, Frank. 1958. Some theorems concerning 2-dimensional brownian motion. *Transactions of the American Mathematical Society* **87** 187–197.

Taira, Kazuaki. 1991. *On the existence of Feller semigroups with Boundary conditions*. American Mathematical Society.

Whitt, Ward. 1980. Some useful functions for functional limit theorems. *Math. Oper. Res.* **5**(1) 67–85.

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

Yoshida, Hidenobu, Ikuko Miyamoto. 1999. Harmonic functions in a cone which vanish on the boundary. *Math. Nachr.* **202** 177–187.

Zhou, C. 2001. An analysis of default correlations and multiple defaults. *Review of Financial Studies* **14** 555–576.