STATIONARY DISTRIBUTION OF THE VOLUME AT THE BEST QUOTE IN A POISSON ORDER BOOK MODEL

Ioane Muni Toke

Stationary distribution of the volume at the best quote in a Poisson order book model

Ioane Muni Toke[1]

[1]MICS Laboratory, Chair of Quantitative Finance, CentraleSupelec, France.

January 2, 2017

Abstract

We develop a Markovian model that deals with the volume offered at the best quote of an electronic order book. The volume of the first limit is a stochastic process whose paths are periodically interrupted and reset to a new value, either by a new limit order submitted inside the spread or by a market order that removes the first limit. Using applied probability results on killing and resurrecting Markov processes, we derive the stationary distribution of the volume offered at the best quote. All proposed models are empirically fitted and compared, stressing the importance of the proposed mechanisms.

Keywords : limit order book ; volume of the best quote ; aggressive limit orders ; aggressive market orders ; killing and resurrecting Markov processes.

1 Introduction

The limit order book has been for the past few years the subject of a growing interest among academics and practitioners studying financial markets. This electronic structure centralizes all the orders submitted to a given market by all participants, finds matchable buy and sell orders, and therefore defines the price of the financial product exchanged. The fundamental question is thus to understand how the sequences of submitted orders – the orders flows – are translated into price dynamics. Biais et al. (1995) and Bouchaud et al. (2002) are pioneer investigations on the empirical properties of the limit order book. Smith et al. (2003) is a pioneer theoretical framework for the study of the continuous double auction that is used in limit order books.

The order book is a complex system. Basic mathematical models, such as Cont et al. (2010), rely on simplifying assumptions: only three main types of orders are submitted (limit, market and cancellations, ignoring exchange-specific rules and specificities) ; all orders flows are Poisson processes ; all orders have the same unit size. Abergel & Jedidi (2013) shows that under appropriate assumptions, some limit theorems apply and such Markovian models lead to a diffusion equation for the price. Muni Toke (2015) shows that with similar Poisson models, the average shape of the order book is analytically computable, even when relaxing the unit-volume hypothesis.

Among the important quantities that describe a limit order book, the volume offered at the best quote (bid or ask), i.e. the total number of shares available at the first limit of the order book, is fundamental. One reason is that this quantity is often the only information easily available to market participants (known as Level-1 data). Another (linked) reason is that this quantity is heavily used in trading strategies by market participants : Farmer et al. (2004) shows that the majority of market orders that move the price have a size exactly equal to the volume of the best quote at the time of submission (see below for more details). However, this volume remains difficult to model. When the volume of the best quote drops to zero (because of large market orders or cancellations that remove all the liquidity of the best quote), or when a new best quote is submitted inside the spread, the price changes, and the volume of the best quote is reset to a new quantity that may have no link to the quantities describing the last order submitted. Therefore, computing the volume of the best quote after a new event requires to keep track of the whole order book.

Recently, Cont & De Larrard (2012) proposes a model in which the order book is restricted to its first limits. When the price does not move, the volume at the best quote obviously varies according to the arriving orders flows, and when this volume drops to zero, the price moves and the volume at the best quote is immediately reset to some random value. Therefore, the volume of the best quotes (bid and ask) is a two-dimensional process with values on the positive orthant that jumps randomly inside the orthant each time it reaches an axis. Cont & De Larrard (2012) shows that under appropriate assumptions and using limit theorems in the spirit of queueing theory, this volume may approximately exhibit a jump-diffusive behaviour. There is however a very restrictive assumption for this model to be valid: the spread has to always be equal to one tick. Indeed, one cannot allow for limit orders submitted inside the spread in this framework, since it would make the process jumps even when it does not reach an axis. Such an assumption may be an appropriate model for busy periods of trading of so-called large tick stocks, but probably not in the general case.

Even more recently, Huang et al. (2015) studied a limit order book modelled (with some constraints) by a collection of queueing systems, one per limit price, in which the intensities of arrival of orders (market, limit and cancellations) are dependent on the size of the queue at this limit price. It is shown that this model with size-dependent intensities is able to give a more realistic distribution of the volume available at the first limit than the classic Poisson model. Figure 1 (left panel) reproduces the distribution of the best quote observed in this model. It appears that such models of limit order books based on point processes with state-dependent intensities may provide fruitful results. Muni Toke & Yoshida (2017) obtain similar figures, also reproduced in Figure 1 (right panel), with a model in which the intensities of the orders flows depend on the spread and

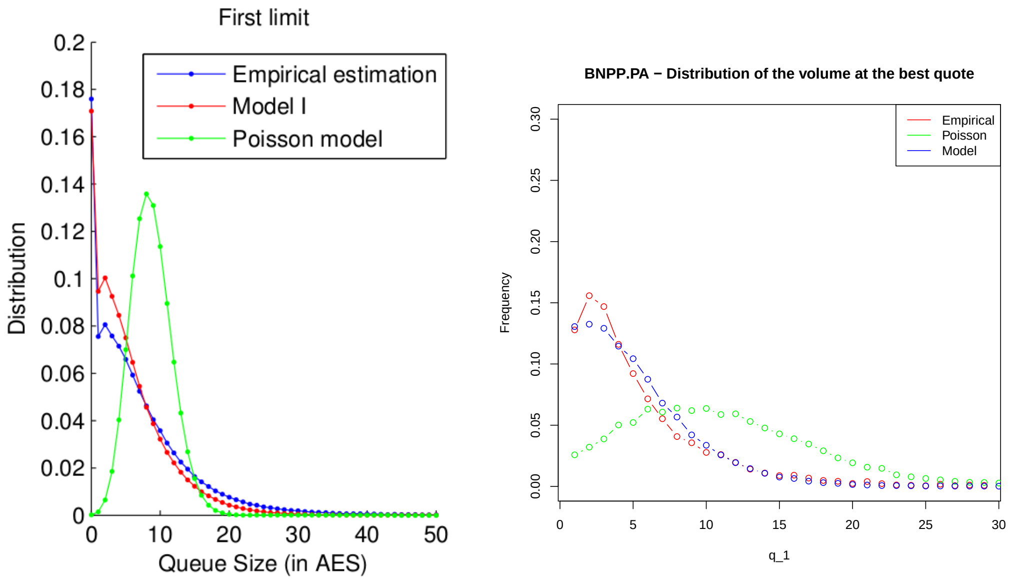

Figure 1: Distribution of the best quotes obtained in state-dependent limit order book models. Left panel reproduced from Huang et al. (2015) and right panel reproduced from Muni Toke & Yoshida (2017).

several volumes characterizing the order book.

In this paper, we show that taking a different route, namely keeping the classical zero-intelligence order book model with Poisson processes, can still lead to interestingly similar results. In the full Poisson model of the order book in this work, the stationary distribution of the volume offered at the best quote can be analytically and then numerically computed. The proposed model is basic but flexible. It does not assume that the spread is always equal to one tick, i.e. all types of events that make the volume of the best quote jump are taken into account: aggressive market orders that matches the full first limit, as well as aggressive limit orders submitted inside the spread as well. The main idea is that such jumps of the volume at the best quote can be identified as killing and resurrecting a Markov process (Pakes 1997). The assumptions that all orders have to be unit-sized can even be lifted, with additional restrictions on the volume of market orders.

The remainder of the paper is organized as follows. Section 2 describes precisely the general zero-intelligence model of the order book with Poisson processes that is used here. The main result on killing and resurrecting Markov processes is recalled and adapted to the order book context. Sections 3 and 4 then explore two types of restrictive assumptions that allow the analytical computation of the stationary distribution of the volume at the best quote: Section 3 excludes market orders that partially match the best quote, while Section 4 allows for all types of market orders but with some size restrictions. Finally, Section 5 provides empirical fittings of all the analytical models proposed, with a comparison to the simulated "best effort" of the unrestricted

Poisson model.

2 A Markovian model of the one-side order book

2.1 Model definition

Let us consider the best ask quote of an electronic order book (the model for the best bid is strictly identical). Three types of events can alter the quantity offered at the best quote : market orders, limit orders and cancellation of existing limit orders. For each type of order, we will distinguish between orders that move the price, called "aggressive" from now on, and orders that do not move the price, labelled "passive" hereafter[1].

Let us start with passive orders. We assume that passive limit orders (i.e. submitted at the best quote) are submitted according to a Poisson process with rate λ_1 and that the size of these passive limit orders form a set of independent and identically distributed random variables with probability distribution $(g_{1,n})_{n\in\mathbb{N}}$. We then assume that passive market orders are submitted according to a Poisson process with rate μ and are all unit-sized (one share). Since a passive market order should not move the price, it cannot be submitted when there is only one share left. The third type of passive orders, passive cancellations, are modelled as follows: each unit-size component of a limit order (i.e. each share) standing at the best quote is cancelled some random time after its submission. All these random times are assumed to form a set of independent and identically distributed random variables with exponential distribution with parameter $\theta_1 > 0$. Since passive cancellations should not move the price, the last share standing should not be removed by this process, so that if there is n_1 shares at the best quote, the total cancellation intensity is $(n_1 - 1)\theta_1$.

Let us now turn to aggressive orders. Aggressive limit orders are limit orders submitted inside the spread, i.e. at a price lower than the current best ask. We will assume that aggressive limit orders are submitted according to a Poisson process with rate λ_0, and that the size of all aggressive limit orders form a set of independent and identically distributed random variables with probability distribution $(g_{0,n})_{n\in\mathbb{N}}$. The effect of the submission of an aggressive limit order is simple: at the moment of the submission, this order instantly becomes the best quote, i.e. the ask price is reset to the price of the aggressive limit order, and the quantity available at the best quote is reset to the volume of the submitted limit order.

Aggressive market orders are submitted according to a Poisson process μ_A, and their size is equal to the available quantity at the best quote at the time of the submission. In other words, aggressive

[1]The concept of aggressiveness of orders is somewhat common in the financial microstructure literature, even though the precise definition of "aggressive" may vary with frameworks and authors. Such a terminology is already used in early works such as Harris & Hasbrouck (1996) and Biais et al. (1995) to classify orders: a limit order submitted inside the spread (i.e. moving the price) is labeled "more aggressive" than a limit order submitted at the best quote, which in turn is labeled "more aggressive" than a limit order submitted inside the book ; similarly, a market order larger than the volume available at the best quote (hence moving the price) is more aggressive than a market order matching only partially the first limit.

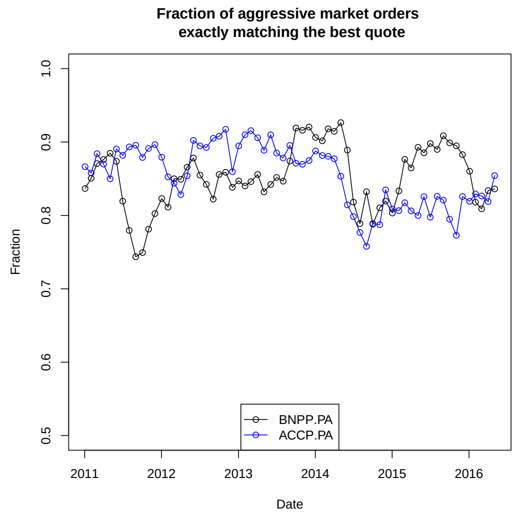

Figure 2: Fraction of aggressive market orders that have a size exactly equal to the volume of the best quote from 2011 to 2016 for BNPP.PA and ACCP.PA.

market orders match all the liquidity available at the best quote. This is not an unreasonable assumption: Farmer et al. (2004) shows that on a 16-stock sample from the London stock exchange, 86% of the buy market orders that change the price have a size that is exactly equal to the volume offered at the best ask. It turns out that this figure is still valid with recent data: on Figure 2, the fraction of aggressive market orders that have a size equal to the volume of the best quote is plotted from 2011 to 2016 for two stocks traded on the Paris stock exchange, and in this period it indeed varies between roughly 80% and 90%. When an aggressive market order is submitted, the volume available at the best quote drops to zero, the price moves up and the quantity available at the best limit is instantaneously reset to the quantity available at the second limit of the order book (we'll say second limit for convenience, it is more precisely the next non-empty limit of the order book).

Note that an aggressive cancellation would be the cancellation of the last share at the best quote with volume one. Its effect is therefore strictly equivalent to the one of an aggressive market order. We can then without loss of generality for our model assume that the intensity μ_A includes aggressive cancellations.

Because of the effect of aggressive market orders and cancellations, we also need to model the second limit of the order book. Using assumptions coherent with the Markovian setting we are establishing, we will assume the following: limit orders are submitted at the second limit (or more generally at any limit inside the book) according to a Poisson process with rate λ_2 ; the sizes of these limit orders form a set of independent and identically distributed random variables with probability distribution $(g_{2,n})_{n\in\mathbb{N}}$; each unit-size component of a limit order (i.e. each share) standing at the

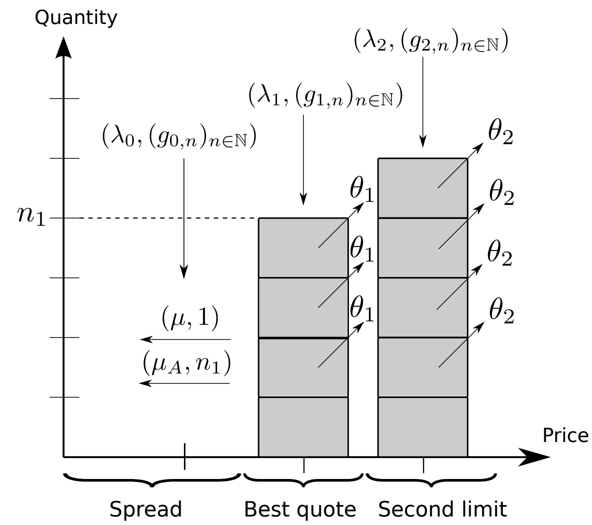

Figure 3: Schematic diagram describing the stylized model of the order book and its three type of orders: limit orders (↓), market orders (←) and cancellations (↗).

second limit is cancelled some random time after its submission, and all these random times are assumed to form a set of independent and identically distributed random variables with exponential distribution with parameter $\theta_2 > 0$.

Figure 3 summarizes the types of orders and notations introduced.

Remark 1. It is of course very well known that the memoryless assumptions made to build this stylized model (Poisson processes, exponentially distributed lifetimes of orders) are not satisfied in practice on financial markets, and that long-memory phenomena, clustering of events, etc. are commonly observed (see e.g. Abergel et al. (2016) and references therein). More realistic modeling of the limit order book can for example rely on Hawkes processes (see e.g. Bacry et al. (2015) for a review), or non-Poisson point processes with state-dependent intensities (see e.g. Huang et al. (2015), Muni Toke & Yoshida (2017)). The interest of this work is to show that even with basic non-realistic assumptions, our purely mechanistic description of the order book can produce empirically realistic distributions otherwise obtained with different principles.

Remark 2. In a full model, submission of aggressive limit orders should be spread-dependent, since these orders cannot be submitted when the spread is equal to one tick. This restriction is not included in the model described here, which is not state-dependent. This assumption makes the model more suitable for small tick stocks, for which the spread is generally not constrained to one tick, rather than for large tick stocks.

2.2 Stationary distribution of the best quote

In this general Poisson model, the continuous-time stochastic process $X = \{X(t), t \in [0, \infty)\}$ describing the volume available at the best quote evolves as follows. Let τ_1 be the random time of the first price move. During the time interval $[0, \tau_1)$, in the absence of events that move the price (aggressive market orders/cancellations, aggressive limit orders), X evolves as the stochastic process $1 + Y = \{1 + Y(t), t \in [0, \infty)\}$, which is one (the last share that cannot be cancelled or executed without moving the price) plus the size of a queue with the infinitesimal generator:

$$\begin{pmatrix} -\lambda_1 & \lambda_1 g_{1,1} & \lambda_1 g_{1,2} & \lambda_1 g_{1,3} & \lambda_1 g_{1,4} & \dots \\ \mu + \theta_1 & -(\mu + \lambda_1 + \theta_1) & \lambda_1 g_{1,1} & \lambda_1 g_{1,2} & \lambda_1 g_{1,3} & \dots \\ 0 & \mu + 2\theta_1 & -(\mu + \lambda_1 + 2\theta_1) & \lambda_1 g_{1,1} & \lambda_1 g_{1,2} & \dots \\ 0 & 0 & \mu + 3\theta_1 & -(\mu + \lambda_1 + 3\theta_1) & \lambda_1 g_{1,1} & \dots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}. \tag{1}$$

Now, at time τ_1, the price moves because of an aggressive limit order or an aggressive market order. The process X is instantaneously reset to a new random variable $X_{\tau_1} = H$ that, depending on the direction of the price move, represents either the size of the incoming aggressive limit order (downward price move), or the volume offered at the second limit inside the order book (upward price move). Then, if τ_2 is the random time of the next price movement, X on $[\tau_1, \tau_2)$ behaves according to the infinitesimal generator (1), and so on.

This mechanism is identifiable to what is known in applied probability as killing and resurrecting a Markov process. The process of the volume of the best quote starts at time 0 and evolves according to the infinitesimal generator (1). Then, upon the submission of an aggressive limit or market order, it is killed, and (instantaneously) resurrected to a random variable H with distribution $(h_i)_{i \in \mathbb{N}^*}$, from where it restarts its course according the previous dynamics. Such a mechanism is studied in Pakes (1997), where the following result is proved.

Theorem 1 (rephrased from Pakes (1997))**.** *Let $Z = \{Z_t, t \in [0, \infty)\}$ be a Markov process on $\mathbb{N}$ with initial distribution $(h_i)_{i \in \mathbb{N}}$. Zero state is assumed to be absorbing for Z. Let $R = \{R_t, t \in [0, \infty)\}$ be the process constructed as follows : R starts following some path of the process Z ; at some random time (exponentially distributed with parameter $\beta > 0$), R is killed, i.e. reset to 0 ; after some random time (exponentially distributed with parameter $\alpha > 0$), R is resurrected, i.e. restarts following a new path of the process Z ; and so on.*

Then the process R admits a stationary distribution $(\pi_j)_{j \in \mathbb{N})}$ given by:

$$\pi_0 = \frac{\beta}{\alpha + \beta - \alpha\beta \hat{f}_0(\beta)}, \quad \pi_j = \alpha \pi_0 \hat{f}_j(\beta), \tag{2}$$

where $\hat{f}_j$ is the Laplace transform of the series $\sum_{i\in\mathbb{N}} h_i p_{i,j}(t)$, in which $p_{i,j}(t) = \mathbf{P}(Z(t) = j|Z(0) = i)$ is the transient probability of the (non-killed) process Z from state i to state j.

In our order book model, the resurrection is instantaneous, i.e. $\alpha \to +\infty$; the killing events are aggressive limit and market orders, i.e. $\beta = \lambda_0 + \mu_A$ by standard properties of the Poisson processes ; the state 0 is not accessible without killing (classic cancellations and partial market orders cannot deplete the best limit), i.e. $\hat{f}_0 = 0$. Another consequence of the last fact is that the distribution $(h_i)_{i\in\mathbb{N}^*}$ of H represents exactly the new volume available at the best quote after an aggressive event. This distribution is therefore a mix of the distribution $(g_{0,i})_{i\in\mathbb{N}^*}$ with probability $\frac{\lambda_0}{\mu_A + \lambda_0}$ and the stationary distribution of the volume at the second limit, denoted $(\pi_{2,i})_{i\in\mathbb{N}}$, with probability $\frac{\mu_A}{\mu_A + \lambda_0}$. We thus obtain the following result.

Proposition 1. *In the general Poisson order book model described in this section, the stationary distribution $(\pi_j)_{j\in\mathbb{N}^*}$ of the total volume offered at the best quote is written for any $j \geq 1$:*

$$\pi_j = (\lambda_0 + \mu_A)\hat{f}_j(\lambda_0 + \mu_A) \tag{3}$$

where $\hat{f}_j(\cdot)$ is the Laplace transform of the series $\sum_{i\in\mathbb{N}} h_i p_{i,j}(t)$ in which $\forall i \in \mathbb{N}^*$,

$$h_i = \frac{\lambda_0 g_{0,i}}{\lambda_0 + \mu_A} + \frac{\mu_A \pi_{2,i}}{\lambda_0 + \mu_A}, \tag{4}$$

and $p_{i,j}(t) = \mathbf{P}(1 + Y(t) = j|1 + Y(0) = i)$ is the transient probability of the (non-killed) process $1 + Y$ from state i to state j.

This result is central in this work. In the following, we will study several specifications of the above general model, all of them allowing analytical tractability at some cost. The first criterion dividing the different types of models is the presence or absence of "partial" market models. In the first type of model (Type-1 models), we assume that all market orders are aggressive market orders. In other words $\mu = 0$, and there are no "partial" market orders that match only partially the best limit. In the second type of models (Type-2 Models), this restriction is lifted, i.e. market orders may or may not be aggressive ($\mu > 0$), but some restrictions on the distributions of the volumes will be added. These different types of models are studied in the next two sections.

3 Models with aggressive market orders exclusively

Type-1 models assume that all market orders are aggressive, i.e. that $\mu = 0$. In this setting, we are able to analytically compute the stationary distribution π of the process X through a direct approach: we compute the transient probabilities $(p_{i,j}(t))_{i,j\in\mathbb{N}}$ by a standard generating function method, as well as the stationary distribution $(\pi_{2,i})_{i\in\mathbb{N}^*}$ by a direct method.

Let us start with the transient probabilities $(p_{i,j}(t))_{i,j\in\mathbb{N}}$. Let $p_{i,j} = r_{i-1,j-1}$ for any $(i,j) \in (\mathbb{N}^*)^2$. $(r_{i,j}(t))_{i,j\in\mathbb{N}}$ are the transition probabilities of the process Y. The Kolmogorov forward equations are written for any $(i,j) \in \mathbb{N}^2$:

$$r'_{ij}(t) = -(\lambda_1 + j\theta_1)r_{ij}(t) + (j+1)\theta_1 r_{i,j+1}(t) + \sum_{k=0}^{j-1} \lambda_1 g_{1,j-k} r_{i,k}(t). \tag{5}$$

Let $G_1(z) = \sum_{j=1}^{+\infty} g_{1,j} z^j$ be the generating function of the distribution of the sizes of incoming limit orders at the best limit. By multiplicating this equation by z^j and summing over j, we obtain that the generating function $\varphi_i(z,t) = \sum_{j=0}^{\infty} r_{ij}(t) z^j$ is solution of the partial differential equation:

$$0 = \frac{\partial \varphi_i}{\partial t}(z,t) - \theta_1(1-z)\frac{\partial \varphi_i}{\partial z}(z,t) + \lambda_1(1 - G_1(z))\varphi_i(z,t), \tag{6}$$

subject to the initial condition $\varphi_i(z,0) = z^i$.

As for the stationary distribution $(\pi_{2,i})_{i\in\mathbb{N}^*}$, we use the Markovian setting described in Section 2 : $\lambda_2 > 0$ is the rate of arrival of limit orders, $(g_{2,n})_{n\in\mathbb{N}}$ is the distribution of their sizes, and $1/\theta_2$ is the average lifetime of a share standing inside the book. Similarly to what has been assumed for the best quote, we assume that the last share cannot be cancelled, so that the size of the queue does not drop to zero (it is by definition the next-non empty limit of the order book). With these assumptions, the size of the book at the second limit is the process $\{1 + Y_2(t), t \in [0,\infty)\}$ with infinitesimal generator

$$\begin{pmatrix}
-\lambda_2 & \lambda_2 g_{2,1} & \lambda_2 g_{2,2} & \lambda_2 g_{2,3} & \lambda_2 g_{2,4} & \dots \\
\theta_2 & -(\lambda_2+\theta_2) & \lambda_2 g_{2,1} & \lambda_2 g_{2,2} & \lambda_2 g_{2,3} & \dots \\
0 & 2\theta_2 & -(\lambda_2+2\theta_2) & \lambda_2 g_{2,1} & \lambda_2 g_{2,2} & \dots \\
0 & 0 & 3\theta_2 & -(\lambda_2+3\theta_2) & \lambda_2 g_{2,1} & \dots \\
\vdots & \vdots & \ddots & \ddots & \ddots & \ddots
\end{pmatrix}. \tag{7}$$

The process Y_2 admits a stationary distribution $(\rho_{2,i})_{i\in\mathbb{N}}$, and obviously $\pi_{2,i} = \rho_{2,i-1}$. Writing the classical balance equations and solving the derived ODE for the generating function $\psi(z) = \sum_{n=0}^{+\infty} \rho_{2,n} z^n$, we obtain :

$$\psi(z) = \rho_{2,0} e^{\frac{\lambda}{\theta}\int_0^z \frac{1-G_2(u)}{1-u}\,du} \tag{8}$$

where $G_2(u) = \sum_{n=0}^{\infty} g_{2,i} u^i$ is the generating function of the distribution of the sizes of limit orders submitted inside the book.

Therefore, if we specify the distributions g_1 and g_2 of the sizes of incoming limit orders respectively at the best quote and inside the book, and if subsequent computations are analytically tractable, then we can derive the distribution π. We study two variants 1a and 1b of the model 1 setting. On the one hand, model 1a assumes that all limit orders submitted at the best quote or inside the book are unit-sized, i.e. $g_{1,1} = g_{2,1} = 1$ and $g_{1,n} = g_{2,n} = 0$ for any $n \geq 2$. This assumptions is the one usually made in zero-intelligence models that look for some analytical tractability (see e.g. Cont et al. 2010). On the other hand, model 1b assumes that all limit orders submitted at the best quote or inside the book are geometrically distributed with parameters $0 < q_1 < 1$ and $0 < q_2 < 1$ respectively. This assumption has been used in Muni Toke (2015) in which it has been used to compute a general average shape of an order book. The results in these two cases are now stated.

Proposition 2 (Model 1a)**.** *If all limit orders at the best quote or inside the book are unit-sized, then the stationary distribution $(\pi_j)_{j\in\mathbb{N}^*}$ of the volume offered at the best quote is:*

$$\pi_j = (\lambda_0 + \mu_A) \sum_{i=1}^{\infty} h_i \sum_{k=0}^{\min(i-1,j-1)} \binom{i-1}{k} \frac{1}{(j-1-k)!} \left(\frac{\lambda_1}{\theta_1}\right)^{j-1-k} \times \int_0^{\infty} e^{-(\lambda_0+\mu)t} e^{-k\theta_1 t} (1 - e^{-\theta_1 t})^{i+j-2-2k} e^{-\frac{\lambda_1}{\theta_1}(1-e^{-\theta_1 t})} \, dt, \quad (9)$$

and

$$h_i = \frac{\lambda_0 g_{0,i}}{\mu_A + \lambda_0} + \frac{\mu_A}{\mu_A + \lambda_0} e^{-\frac{\lambda_2}{\theta_2}} \frac{(\lambda_2)^{i-1}}{(\theta_2)^{i-1}(i-1)!}. \quad (10)$$

Proof. The unit-size assumption gives $G_1(z) = z$, and inserting this in equation (6) allows a direct solving of the latter as:

$$\varphi_i(z,t) = \left[1 - (1-z)e^{-\theta_1 t}\right]^i \exp\left[\frac{\lambda_1}{\theta_1}\left(z - (1-z)e^{-\theta_1 t}\right)\right], \quad (11)$$

which then gives after some computations, using Leibniz differentiation formula, the transition probabilities:

$$r_{ij}(t) = \sum_{k=0}^{\min(i,j)} \frac{i!}{k!(i-k)!(j-k)!} \left(\frac{\lambda_1}{\theta_1}\right)^{j-k} e^{-k\theta_1 t} (1 - e^{-\theta_1 t})^{i+j-2k} e^{-\frac{\lambda_1}{\theta_1}(1-e^{-\theta_1 t})}. \quad (12)$$

Shifting both indices by one, multiplying by h_i, summing over i and taking the Laplace transform yields equation (9).

Furthermore, still using the unit-size assumption, the stationary distribution of a process with generator defined in equation (7) is a Poisson distribution with parameter $\frac{\lambda_2}{\theta_2}$. This readily gives equation (10). $\square$

Proposition 3 (Model 1b)**.** *If all limit orders submitted at the best quote and inside the book are i.i.d. and geometrically-distributed with parameter q_1 and q_2 respectively, then the stationary distribution $(\pi_j)_{j\in\mathbb{N}^*}$ of the volume offered at the best quote is:*

$$\begin{aligned}\pi_j =&(\lambda_0+\mu_A)\left[\sum_{i=1}^{\infty} h_i \left[q_1\frac{\lambda_1}{\theta_1}\sum_{k=0}^{\min(i-1,j-2)}\binom{i-1}{k}\frac{1}{(j-1-k)!}\sum_{l=0}^{j-2-k}\binom{j-1-k}{l}(-1)^l\right.\right.\\ &\times\prod_{\alpha=1}^{l}\left(\frac{\lambda_1}{\theta_1}-\alpha(1-q_1)\right)\prod_{\beta=1}^{j-2-k-l}\left(\frac{\lambda_1}{\theta_1}+\beta(1-q_1)\right)\\ &\left.\times\int_0^{+\infty}(1-e^{-\theta_1 t})^{i-k}e^{-(l+k)\theta_1 t}\left[q_1+(1-q_1)e^{-\theta_1 t}\right]^{\frac{\lambda_1}{\theta_1(1-q_1)}-l-1}e^{-(\lambda_0+\mu)t}\,dt\right]\\ &\left.+\sum_{i=j}^{\infty}h_i\binom{i-1}{j-1}\int_0^{+\infty}e^{-(j-1)\theta_1 t}(1-e^{-\theta_1 t})^{i-j}\left[q_1+(1-q_1)e^{-\theta_1 t}\right]^{\frac{\lambda_1}{\theta_1(1-q_1)}}e^{-(\lambda_0+\mu)t}\,dt\right], \quad (13)\end{aligned}$$

and

$$h_i = \frac{\lambda_0 g_{0,i}}{\mu_A+\lambda_0}+\frac{\mu_A}{\mu_A+\lambda_0}q_2^{\frac{\lambda_2}{(1-q_2)\theta_2}}\frac{(1-q_2)^{i-1}}{(i-1)!}\frac{\Gamma(i-1+\frac{\lambda_2}{(1-q_2)\theta_2})}{\Gamma(\frac{\lambda_2}{(1-q_2)\theta_2})}. \quad (14)$$

Proof. The assumption of a geometric distribution of the sizes of limit orders inside the book gives $G_1(z)=\frac{q_1 z}{1-(1-q_1)z}$. With this definition of G_1, we can solve equation (6) to obtain:

$$\varphi_i(z,t)=[1-(1-z)e^{-\theta_1 t}]^i\left[\frac{q_1+(1-q_1)(1-z)e^{-\theta_1 t}}{1-(1-q_1)z}\right]^{\frac{\lambda_1}{\theta(1-q_1)}}. \quad (15)$$

The Leibniz differentiation formula and some computations lead to the transient probabilities:

$$\begin{aligned}r_{ij}(t) =&q_1\frac{\lambda_1}{\theta_1}\sum_{k=0}^{\min(i,j-1)}\frac{i!}{k!(i-k)!(j-k)!}(1-e^{-\theta_1 t})^{i-k+1}\\ &\times\sum_{l=0}^{j-1-k}\binom{j-1-k}{l}(-1)^l e^{-(l+k)\theta_1 t}\left[q_1+(1-q_1)e^{-\theta_1 t}\right]^{\frac{\lambda_1}{\theta_1(1-q_1)}-l-1}\\ &\times\prod_{\alpha=1}^{l}\left(\frac{\lambda_1}{\theta_1}-\alpha(1-q_1)\right)\prod_{\beta=1}^{l}\left(\frac{\lambda_1}{\theta_1}+\beta(1-q_1)\right)\\ &+\mathbf{1}_{j\le i}\binom{i}{j}e^{-j\theta_1 t}(1-e^{-\theta_1 t})^{i-j}\left[q_1+(1-q_1)e^{-\theta_1 t}\right]^{\frac{\lambda_1}{\theta_1(1-q_1)}}. \quad (16)\end{aligned}$$

As in the previous case, shifting the indices by one, multiplying by h_i, summing over i and taking the Laplace transform yields equation (13).

Now, as for the stationary distribution of the volume offered at the second limit, the assumption of geometrically distributed sizes of limit orders gives $G_2(z) = \dfrac{q_2 z}{1-(1-q_2)z}$, and equation (8) gives by derivation and after some computations:

$$\forall i \in \mathbb{N}, \rho_{2,i} = q_2^{\frac{\lambda_2}{(1-q_2)\theta_2}} \frac{(1-q_2)^i}{i!} \frac{\Gamma(i + \frac{\lambda_2}{(1-q_2)\theta_2})}{\Gamma(\frac{\lambda_2}{(1-q_2)\theta_2})}, \tag{17}$$

hence the result. Note that the distribution $(\rho_{2,i})_{i\in\mathbb{N}}$ is a negative binomial distribution (or Polya distribution) with non-integer size parameter $\dfrac{\lambda_2}{(1-q_2)\theta_2}$ and probability parameter q_2 (see e.g. Feller 1968, Chap. VI.). $\square$

4 Models with both partial and aggressive market orders

Type-2 models allow for both partial and aggressive market orders to be submitted, i.e. $\mu > 0$ and $\mu_A > 0$. With both types of market orders, the direct approach of the previous section does not provide analytically tractable results. Therefore we propose a different strategy. In this new setting, we can keep the tractability of the model at the cost of assuming that classic orders directly affecting the best quote are unit-sized. This is the only restriction: orders submitted inside the spread or inside the book can be kept with a general distribution, and the aggressive market orders are still defined the same way, obviously, with a size equal to the volume at the best quote. With this assumption, the best quote is a birth-and-death process, for which we can compute the Laplace transform of its transition probabilities, which can be expressed using continuous fractions. The original result dates back to Murphy & O'Donohoe (1975) but a modern derivation of the result is found in Crawford & Suchard (2012). Therefore, we are able to study the stationary distribution of the volume available at the best quote without computing the transient probabilities of the process Y with infinitesimal generator (1) and $\mu > 0$ (to our knowledge, such a computation is still an unresolved challenge).

Assume that both partial market orders and limit orders submitted at the best quote are unit-sized. Then the process Y, which translates the evolution of the volume at the best quote (minus one) without any price movement, is a birth-and-death process with constant birth-immigration rate λ_1 and linear death-emigration rate $\mu+n\theta_1$ for any $n \geq 0$. Let $(q_{m,n}(t)), (m,n) \in \mathbb{N}^2, t \in [0,\infty)$ be the transition probabilities of the process Y, and $(\hat{q}_{m,n}(s)), (m,n) \in \mathbb{N}^2, s \in \mathbb{C}$ their Laplace transform, if it exists. Let $(B_n(s))_{n\in\mathbb{N}}$ the real sequence defined by the two-step recurrence:

$$\begin{cases} B_0(s) = 1, \;\; B_1(s) = s + \lambda_1, \\ B_n(s) = (s + \lambda_1 + \mu + (n-1)\theta_1)\, B_{n-1} - \lambda_1\, (\mu + (n-1)\theta_1)\, B_{n-2}, \; n \geq 2. \end{cases} \tag{18}$$

Then, adapting Crawford & Suchard (2012, Theorem 1) to our special case, we have for any $(m,n) \in \mathbb{N}^2$ such that $m \leq n$:

$$\hat{q}_{m,n}(s) = \lambda_1^{n-m} \frac{\hat{a}_1}{\hat{b}_1 + \frac{\hat{a}_2}{\hat{b}_2 + \frac{\hat{a}_3}{\hat{b}_3 + \dots}}} \triangleq \lambda_1^{n-m} \frac{\hat{a}_1(s)}{\hat{b}_1(s)+} \frac{\hat{a}_2(s)}{\hat{b}_2(s)+} \frac{\hat{a}_3(s)}{\hat{b}_3(s)+} \dots \tag{19}$$

(The symbol $\triangleq$ is the definition equality that introduce a simplified notation for the continuous fractions.) In the above equation the sequences $(\hat{a}_i)_{i\in\mathbb{N}^*}$ and $(\hat{b}_i)_{i\in\mathbb{N}^*}$ are defined as follows:

$$\hat{a}_i = \begin{cases} B_m(s) & \text{if } i = 1, \\ -\lambda_1 \left(\mu + (n+1)\theta_1\right) B_n(s) & \text{if } i = 2, \\ -\lambda_1 \left(\mu + (n+i-1)\theta_1\right) & \text{if } i \geq 3, \end{cases} \tag{20}$$

and

$$\hat{b}_i = \begin{cases} B_{n+1}(s) & \text{if } i = 1, \\ s + \lambda_1 + \mu + (n+i-1)\theta_1 & \text{if } i \geq 2. \end{cases} \tag{21}$$

The result for any $(m,n) \in \mathbb{N}^2$ such that $m \geq n$ is similarly written:

$$\hat{q}_{m,n}(s) = \left(\prod_{j=n+1}^{m} (\mu + j\theta_1) \right) \frac{\hat{\alpha}_1(s)}{\hat{\beta}_1(s)+} \frac{\hat{\alpha}_2(s)}{\hat{\beta}_2(s)+} \frac{\hat{\alpha}_3(s)}{\hat{\beta}_3(s)+} \dots \tag{22}$$

where the sequences $(\hat{\alpha}_i)_{i\in\mathbb{N}^*}$ and $(\hat{\beta}_i)_{i\in\mathbb{N}^*}$ are defined as follows:

$$\hat{\alpha}_i = \begin{cases} B_n(s) & \text{if } i = 1, \\ -\lambda_1 \left(\mu + (m+1)\theta_1\right) B_m(s) & \text{if } i = 2, \\ -\lambda_1 \left(\mu + (m+i-1)\theta_1\right) & \text{if } i \geq 3, \end{cases} \tag{23}$$

and

$$\hat{\beta}_i = \begin{cases} B_{m+1}(s) & \text{if } i = 1, \\ s + \lambda_1 + \mu + (m+i-1)\theta_1 & \text{if } i \geq 2. \end{cases} \tag{24}$$

Hence, the Laplace transforms $(\hat{q}_{m,n}(s)), (m,n) \in \mathbb{N}^2, s \in \mathbb{C}$ are numerically computable using appropriate numerical methods for the computation of continuous functions. If we now go back to the killing and resurrection of Markov processes, we have the following result.

Proposition 4 (Type-2 models). *If all orders submitted at the best quote are unit-sized, then the stationary distribution of the volume offered at the best quote is given by equation* (3), *i.e.* $\pi_j = (\lambda_0 + \mu_A) \sum_{m=1}^{\infty} h_m \hat{q}_{m-1,j-1}(\lambda_0 + \mu_A)$, *where the* $\hat{q}_{m,n}$*'s are given by* (19) *and* (22).

If limit orders submitted inside the book are assumed to be unit-sized as well (model 2a), then the probabilities h_m are given by equation (10). *If the sizes of these limit orders are assumed to be geometrically distributed (model 2b), then the probabilities h_m are obtained with equation* (14).

5 Empirical results

5.1 Data and estimation

We use Thomson-Reuters tick-by-tick data for the stock BNPP.PA traded on the Paris stock exchange, from January 3rd, 2011 to May 20th, 2016, i.e. a five-year-and-five-month long sample. This stock among the largest market capitalizations and most liquid stocks on the Paris stock exchange.

For each available trading day, we consider the data from 10:00 a.m. to 16:00 p.m., i.e. a six-hour period at the heart of the trading day. The idea is to get rid of the very busy opening and closing period where the assumption of a stationary model may be difficult to fulfil. It is well-known that even on the six-hour period considered, one does observe a seasonal activity (the U-shaped pattern of financial activity), and one might get better "stationary" results by shortening the daily period under investigation. However, it will appear that our models actually provides satisfying results even using the full day sample. For each stock, for each trading day, we compute the total numbers and the distributions of the sizes of: limit orders inside the spread; limit orders at the best quote; limit orders at the second best limit; partial market orders; and aggressive market orders. We also compute the time-weighted empirical distribution of the volume offered at the best quote and at the second best limit. Details on data preparation can be found in Muni Toke (2016).

Straightforwardly, the estimators of the Poisson parameters $\lambda_0, \lambda_1, \lambda_2, \mu$ and μ_A (if needed by the model) are defined as the number of the associated events (aggressive limit order, limit order at the best quote, limit order inside the book, partial market orders, aggressive market orders) divided by the length of the time interval. For each of these types of orders, we also compute their respective mean order size $\sigma_0, \sigma_1, \sigma_2, \sigma_\mu$ and σ_{μ_A}. As for the cancellation parameters, we do not have any data allowing us to track the submitted orders individually, and therefore we cannot easily estimate an average lifetime θ_1^{-1} and θ_2^{-1} for cancelled orders at the best quote and inside the book. However, we can get an order of magnitude by using equilibrium relations of incoming and outgoing flows of the order book. Our data let us compute the time-weighted average volume offered at the best quote L_1 and at the second best quote L_2. Then equating the average number of incoming and outgoing share in the order book, we set θ_1 and θ_2 so that $\lambda_1\sigma_1 = \mu\sigma_\mu + \mu_A\sigma_{\mu_A} + \theta_1 L_1$ and $\lambda_2\sigma_2 = \theta_2 L_2$.

Finally, following our theoretical framework, we will assume that all partial market orders are unit-sized, and therefore rescale all size and volume quantities by the median trade size. The rescaling gives us the empirical versions of the distributions $(g_{0,i}), (g_{1,i}), (g_{2,i})$ and $(\pi_{2,i})$ (if needed by the model) with a support roughly included in $1, \ldots, 40$. It may happen that the average rescaled limit order size at the best quote is actually smaller than the median trade size, forbidding a geometric modelling of the distribution (expectation lower than 1). In such cases (frequent from 2014 in our data, very rare before), the scaling size is set to a fraction of the median trade size, so that the average rescaled limit order size at the best quote is larger slightly larger that 1 (we've arbitrarily chosen 1.1 in the following empirical work).

5.2 Fitted models and benchmarks

In the previous section we have presented two types of models (namely type-1, without partial market order, and type-2, with unit-sized partial market orders), each type having two variants (namely, a and b). In variants a, the distribution $(g_{2,i})_{i\in\mathbb{N}^*}$, which represents the volume of the limit orders submitted inside the book, is a Dirac distribution on the atom 1. In variants b, it is geometric with parameter $0 < q_2 \leq 1$. The distribution $(g_{0,i})_{i\in\mathbb{N}^*}$, which represents the volume of the limit orders submitted inside the spread, has not been specified up to now. In line with the other assumptions, we will assume in this empirical section that this distribution is a Dirac on 1 in variants a, and a geometric distribution with parameter $0 < q_0 \leq 1$ in variants b.

We now add further elements of comparison for our models. First, for each type of model, we add a variant c in which both distributions $(g_{0,i})_{i\in\mathbb{N}^*}$ and $(g_{2,i})_{i\in\mathbb{N}^*}$ are taken equal to their empirical counterpart. Furthermore, in order to underline the importance of the mechanism that takes into account aggressive orders and consequent upward and downward movements of the best price, we recall as benchmark the following simplistic model of the best quote, in which aggressive market and limit orders are ignored, which is equivalent to assume a constant price in our setting. This benchmark will be referred to as the Type-0 model. In the Type-0 model, limit orders arrive at rate λ_1 with volume distribution $(g_{1,i})_{i\in\mathbb{N}}$; market orders are unit-sized and arrive at rate μ ; all standing shares have a (i.i.d.) exponential lifetime with parameter $\theta_1 > 0$. It is thus easily shown that the volume at the best quote in the Type-0 model is the Markov process with infinitesimal generator given at equation (1). This process admits a stationary distribution $(\pi_i)_{i\in\mathbb{N}}$ satisfying the following recurrence :

$$\left\{ \begin{array}{lcl} 0 & = & -\lambda_1\pi_0 + (\mu+\theta_1)\pi_1, \\ 0 & = & -(\lambda_1+\mu+n\theta_1)\pi_n + (\mu+(n+1)\theta_1)\pi_{n+1} + \lambda_1 \displaystyle\sum_{i=1}^{n} g_{1,i}\pi_{n-i} \quad (n \geq 1), \end{array} \right. \tag{25}$$

with

$$\pi_0 = \left(\frac{\mu}{\theta_1} \int_0^1 u^{\frac{\mu}{\theta_1}-1} e^{\frac{\lambda_1}{\theta_1}\int_u^1 H(v)\,dv}\, du, \right)^{-1} \tag{26}$$

where $H(u) = \dfrac{1-G_1(u)}{1-u}$, G_1 being the generating function of the distribution $(g_{1,i})_{i\in\mathbb{N}}$ (see e.g.

Model type	$(g_{0,i})_{i\in\mathbb{N}}$	$(g_{1,i})_{i\in\mathbb{N}}$	$(g_{2,i})_{i\in\mathbb{N}}$	$(\pi_{2,i})_{i\in\mathbb{N}}$	$\mu > 0$?	$\lambda_0 > 0, \mu_A > 0$?
Model 0a	None	Unit-size	None	None	Yes	No
Model 0b	None	Geometric	None	None	Yes	No
Model 1a	Unit-size	Unit-size	Unit-Size	*Poisson*	No	Yes
Model 1b	Geometric	Geometric	Geometric	*Neg. binomial*	No	Yes
Model 1c	Empirical	Geometric	*None*	Empirical	No	Yes
Model 2a	Unit-size	Unit-size	Unit-Size	*Poisson*	Yes	Yes
Model 2b	Geometric	Unit-size	Geometric	*Neg. binomial*	Yes	Yes
Model 2c	Empirical	Unit-size	*None*	Empirical	Yes	Yes
Model 3	Empirical	Empirical	*None*	Empirical	Yes	Yes

Table 1: Summary of the different models and their variants. Italic means that the distribution $(\pi_{2,i})_{i\in\mathbb{N}}$ (resp. $(g_{2,i})_{i\in\mathbb{N}}$) in these variants is not a free parameter, but a consequence of the choice of $(g_{2,i})_{i\in\mathbb{N}}$ (resp. $(\pi_{2,i})_{i\in\mathbb{N}}$).

Muni Toke 2015, section 4). We will consider two cases: unit-sized limit orders (Model 0a) and limit orders with geometrically-distributed size with parameter q_1 (Model 0b). In line with the general models, one may assume that the last share cannot be cancelled or executed by shifting the indices of distribution $(\pi_i)_{i\in\mathbb{N}}$ by 1 (i.e. on $\mathbb{N}^*$).

Finally, we add a second benchmark by simulating our general zero-intelligence model of the best quote with all distributions of the model equal to their empirical counterpart. This could be considered as the "best effort" of a zero-intelligence mechanism with both classic and aggressive market and limit orders to compute the distribution of the volume offered at the best quote. Note that we cannot solve analytically the distribution in this general setting: the distribution in this case is numerically estimated by simulation. This simulated benchmark will be referred to as Type-3 model.

Table 1 summarizes the models that are under study here.

5.3 Empirical results

We split the sample into 65 monthly calendar periods. For each period, we compute the analytical distribution of the volume offered at the best quote for all our models and variants, as well as the distribution obtained by simulation of the model 3.

On figure 4 we compare some examples of these analytical distributions with their empirical counterpart. Six examples are provided (every month of January on the sample, from 2011 to 2016), that illustrate different qualities of fits that are found on the data: at least visually, January 2011, 2012 and 2014 are good fits, 2016 is average while the 2013 and 2015 are of lesser quality. The interesting result is that the killing and resurrecting mechanism used to compute the analytical distributions of models of types 1 and 2 is able to produce an empirically-sound shape of the distribution of the volume offered at the best quote in an order book. Indeed, all models of

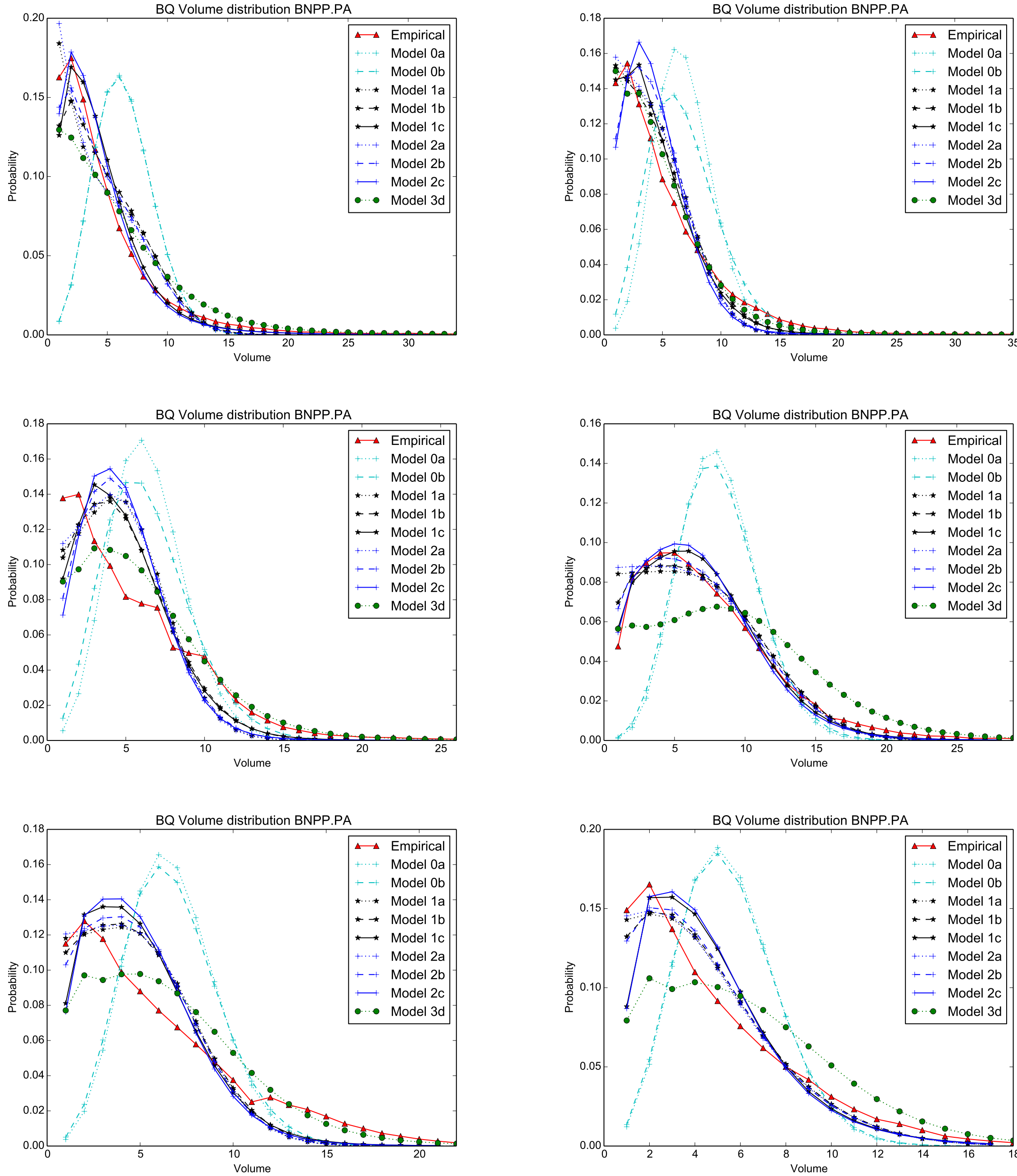

Figure 4: Distribution $(\pi_i)_{i\in\mathbb{N}}$ of the volume offered at the best quote for all the models described, compared to the empirical one. Data fitted on the January period, from 2011 to 2016 (top left to bottom right, by row).

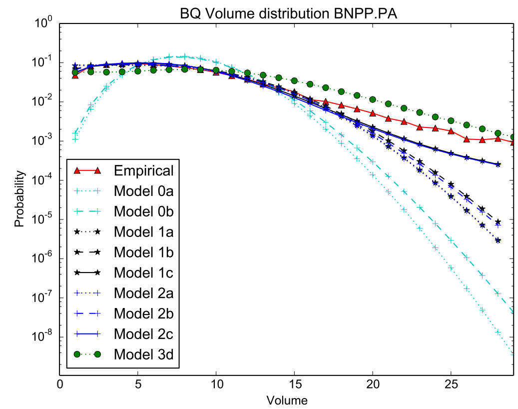

Figure 5: Distribution $(\pi_i)_{i\in\mathbb{N}}$ of the volume offered at the best quote for all the models described, compared to the empirical one, in semi-log-scale, for January 2014.

type 1 and 2 outperform the basic models of type 0, the latter failing to reproduce the empirical distribution. This outlines the importance of the flows of aggressive limit orders and aggressive market orders when describing the volume at the best quote. This quantity cannot be modeled by a simple queue with limit order, partial cancellation and partial market order.

It is enlightening to compare these shapes with the ones produced by more complete state-dependent order book models such as the ones mentioned in Section 1. Figure 1 reproduces two distributions of the best quotes obtained by state-dependent models. More precisely, to obtain these shapes, Huang et al. (2015) use a model in which the intensities of submission of orders depend and the volume of the queue in which it is submitted, while Muni Toke & Yoshida (2017) use a model in which intensities of submission of orders depend on the spread, the volume at the best quote and the total volume of the book. It is very interesting to observe that a basic non-state-dependent time-homogeneous model such as the one proposed here is still able to reproduce sound shapes of the distribution of the volume at the best quote, very similar to the ones obtained by fuller models.

We complement these observations on the main body of the distribution by looking at the tail on Figure 5 (for brevity only one period is shown, but the analysis is valid for all periods). As expected, large differences between the models are observed when looking at the tail of the distribution. Variants a, with unit-sized orders, exhibit the thinner tails. The geometric distributions for the size of orders at the best quote and in the book (variants b) allows for slightly heavier tails, but still much thinner than the empirically observed one. Interestingly, variants c exhibit a tail of the analytical distribution of the volume at the best quote that is much heavier. This hints that using the empirical distributions for the distributions $(g_{0,i})_{i\in\mathbb{N}}$ (aggressive limit orders) and $(\pi_{2,i})_{i\in\mathbb{N}}$ (after an aggressive market order) is sufficient to obtain a heavy tail at the best quote, even if the distribution of the size of the limit orders arriving at the best quote is thin-tailed. This

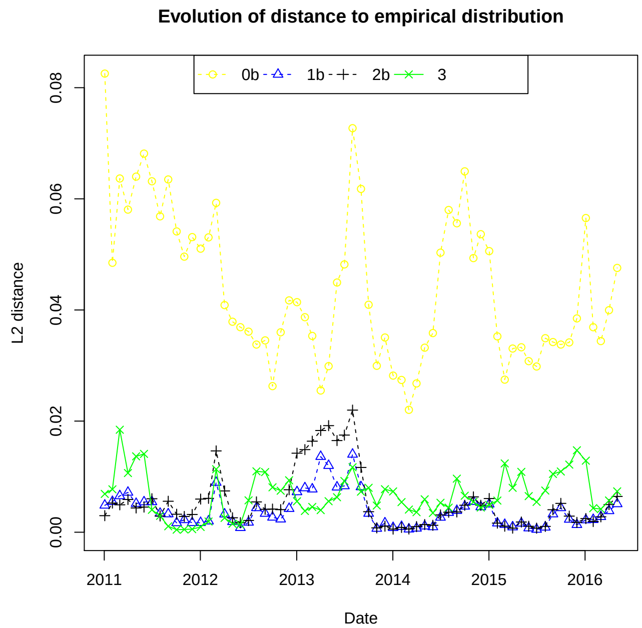

Figure 6: Distance in L^2-norm between the distribution of the volume at the best quote computed by each type of models and the empirical observation.

again underlines the prime importance of the aggressive order flows modeled with the killing and resurrecting mechanism in our models.

We finally investigate the performances of the model over time. For each period, we plot on 6 the L^2-distance between the computed (or simulated) distributions and the empirical one. For readability we have only included the versions of the models with geometric sizes (variants b), but other versions are very close. As observed above, the Poisson benchmark is outperformed by the proposed models. The performances of the proposed models are roughly constant through time on the tested sample, with a temporary decrease in the first part of 2013. The quality of the fit of models of type 1 and 2 is very similar, yet on this sample models of type 1 are in average slightly better that models of type 2. Recall that models of type 1 have only aggressive market orders (no partial market orders) but allow for a flexible (geometric) size of limit orders at the best quote, while models of type 2 require that all limit orders at the best quote are unit-sized. This observation underlines the potential importance of allowing general sizes of orders in a limit order book model. We also observe that the performances of the proposed models (of both types 1 and 2) is comparable to the one of model 3. This is a good achievement, since model 3 is the result of a simulation of a complete limit order book with all zero-intelligence parameters available while others models are limited. We may also observe a bit surprisingly that models of types 1 and 2 even provide in average a better fit of the volume at the best quote than the model 3. It may not

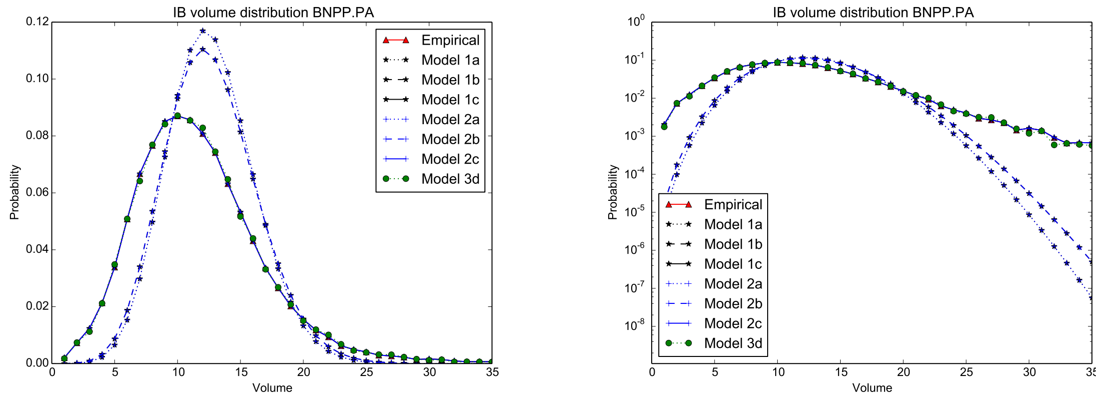

Figure 7: Distribution $(\pi_2, i)_{i\in\mathbb{N}}$ (empirical, Poisson obtained with models 1a and 2a, and negative binomial obtaind with models 1b and 2b) of the volume at the second best quote, main body (left panel) and tail in semi-log-scale (right panel) for January 2014.

be straightforward to explain this fact, so we will not risk a conjecture at the moment.

Remark 3. Models 1b is on average the best model on the whole sample. However, formula (14) involves several series and integrals that have to be numerically evaluated, which makes this model much more time-costly than the others. Using basic (non-optimized) straightforward Python implementations of the above formulas, using Scipy for the numerical integration with default numerical parameters, model 1b/1c require about 4 minutes to compute the whole distribution while models 1a needs less than 1 second and models of type 2 less than 2 seconds. These absolute values are obviously only indicative, since they are hardware-dependent and numerical optimization and enhancements might be performed, but the relative values may be important in selecting the appropriate model for a potential use.

6 Conclusion

This paper has shown that a basic zero-intelligence model of the limit order book is able to accurately describe the stationary distribution of volume offered at the best quote providing it includes a proper mechanism to take into account aggressive orders that move the price: aggressive limit orders submitted inside the spread, as well as aggressive market orders that remove the whole liquidity available at the first level of the book. We have modelled these aggressive orders using results on killing and resurrecting a Markov process (this process being here the quantity at the best quote when the price is constant), which allows us to provide analytical formulas for the distribution of interest.

We end this section by one last observation that may trigger future enhancements. Figure 7 shows that the theoretical models for $(\pi_{2,i})_{i\in\mathbb{N}}$ have the same defects in modeling the empirical distribution of the volume inside the book as Type-0 models with respect to the volume at the best quote: namely, a body slightly shifted to the right and a tail too thin (only one period is plotted, but the observation is valid on the whole sample). This is not surprising as the volume inside the book as been treated here as one entity, without taking into account killing and resurrecting (as in models of Type-0). Future work may include the killing and resurrecting as a cascading effect: when the price move, *all* the limits in the book are shifted, i.e. are killed and resurrected.

Acknowledgements

The author wishes to thank the participants of the 2015 IPAM Workshop on The Mathematics of High Frequency Financial Markets, as well as an anonymous referee for comments that helped improved the manuscript.

References

Abergel, F., Anane, M., Chakraborti, A., Jedidi, A. & Muni Toke, I. (2016), *Limit order books*, Cambridge University Press.

Abergel, F. & Jedidi, A. (2013), 'A mathematical approach to order book modeling', *International Journal of Theoretical and Applied Finance* **16**(05).

Bacry, E., Mastromatteo, I. & Muzy, J.-F. (2015), 'Hawkes processes in finance', *Market Microstructure and Liquidity* **1**(01), 1550005.

Biais, B., Hillion, P. & Spatt, C. (1995), 'An empirical analysis of the limit order book and the order flow in the paris bourse', *Journal of Finance* **50**(5), 1655–1689.

Bouchaud, J.-P., Mézard, M. & Potters, M. (2002), 'Statistical properties of stock order books: empirical results and models', *Quantitative Finance* **2**(4), 251–256.

Cont, R. & De Larrard, A. (2012), 'Order book dynamics in liquid markets: limit theorems and diffusion approximations', *arXiv preprint arXiv:1202.6412* .

Cont, R., Stoikov, S. & Talreja, R. (2010), 'A stochastic model for order book dynamics', *Operations research* **58**(3), 549–563.

Crawford, F. W. & Suchard, M. A. (2012), 'Transition probabilities for general birth–death processes with applications in ecology, genetics, and evolution', *Journal of mathematical biology* **65**(3), 553–580.

Farmer, J., Gillemot, L., Lillo, F., Mike, S. & Sen, A. (2004), 'What really causes large price changes?', *Quantitative Finance* **4**(4), 383–397.

Feller, W. (1968), *An Introduction to Probability Theory and Its Applications: Volume One*, 3rd edn, John Wiley & Sons.

Harris, L. & Hasbrouck, J. (1996), 'Market vs. limit orders: the superdot evidence on order submission strategy', *Journal of Financial and Quantitative analysis* **31**(02), 213–231.

Huang, W., Lehalle, C.-A. & Rosenbaum, M. (2015), 'Simulating and analyzing order book data: The queue-reactive model', *Journal of the American Statistical Association* **110**(509), 107–122.

Muni Toke, I. (2015), 'The order book as a queueing system: average depth and influence of the size of limit orders', *Quantitative Finance* **15**(5), 795–808.

Muni Toke, I. (2016), 'Reconstruction of order flows using aggregated data', *Market Microstructure and Liquidity* **02**, 1650007.

Muni Toke, I. & Yoshida, N. (2017), 'Modelling intensities of order flows in a limit order book', *Quantitative Finance* . (to appear).

Murphy, J. & O'Donohoe, M. (1975), 'Some properties of continued fractions with applications in markov processes', *IMA Journal of Applied Mathematics* **16**(1), 57–71.

Pakes, A. G. (1997), 'Killing and resurrection of markov processes', *Communications in Statistics, Stochastic Models* **13**(2), 255–269.

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