The London School of Economics and Political Science

Stochastic Models for the Limit Order Book

Filippo Riccardi

A thesis submitted to the Department of Statistics of the London School of Economics for the degree of Master of Philosophy, London, January 2014.

Declaration

I certify that the thesis I have presented for examination for the MPhil degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of 12,924 words.

Abstract

The aim of this thesis is to outline a new approach to the Limit Order Book's (LOB) modelling. This is accomplished starting from the existing literature and then proposing several models with different features and complexity levels, the aim being to compute some quantities of interest.

Chapter 1 is an introduction to the LOB. It explains what this object is, why we want to model it and what has been done so far in literature. We also outline the general ideas behind this work.

Chapter 2 is focused on the avalanche approach: orders in the LOB accumulate on some levels and get executed when the price process crosses such values. This idea, as it will be explained, belongs to my supervisor Dr Rheinlander. This model uses the theory related to the local time of a Brownian motion.

Chapter 3 defines a model introducing a Poisson process for incoming orders and cancellations. The framework outlined in this way is used to calculate quantities of interest: a simulation technique is described and a refinement of the model is also presented, in order to be consistent with empirical behaviours observed in the markets.

Finally, in Chapter 4, the local time approach is once again taken into considerations and two models are proposed using downcrossings and excursions. It is also shown how to link the two frameworks.

Contents

Chapter 1

Introduction

1.1 The Limit Order Book

Most financial markets are organised as pure electronic Limit Order Books (LOB). In this chapter we describe this entity and outline why it is studied and what this thesis intends to add.

The LOB is the collection of all orders placed but not yet executed: e.g. the current price of a stock is £100 and a trader wants to sell her shares at £105. She will place an order in the LOB and, as soon as the price reaches level £105, the shares will be sold.

More precisely, a LOB can be described by what follows:

- a limit order is an ex-ante commitment made at time t to trade a certain security up to a given amount at a specified price;
- orders accumulate in the LOB and are removed if cancelled or if traded against an incoming market order;
- traders' orders are handled sequentially: price-priority first and then time-priority;
- in order to make any inference from an order submission, it is essential to condition on the state the trader faces: the massive state space of the LOB dynamics makes modelling very complex.

Traders have two main reasons to trade, often at the same time. On one hand, they can have a view on the fundamental value of the asset. On the other hand, they may have to trade by a deadline, i.e. there could be portfolio reasons.

Agents optimally choose between two type of orders.

1. *Limit orders*: these feature a better price but have execution uncertainty. They could be also picked off if, after a news announcement for instance, someone executes against them before they can be cancelled.

2. *Market orders*: the execution is guaranteed but at a worse price.

Studying the LOB turns out to be very interesting since it can give an idea of what the investors' expectations are for an asset. In fact, for instance, if the limit price is chosen to be very close to the current stock price, then the investor is impatient (since there is a high probability of having the order executed soon). Furthermore, it is an effective tool to study liquidity risk.

We conclude the description of the LOB presenting an example. The following figure shows what happens when a market buy order for 1000 shares is placed:

Price	Book at time t		Δt	Book at time $t + \Delta t$	
59	1000		...	1000	
58	1500		...	1500	
57	2000	*Ask*	-1000	1000	*Ask*
56	-1200	*Bid*	...	-1200	*Bid*
55	-1000		...	-1000	
54	-800		...	-800	

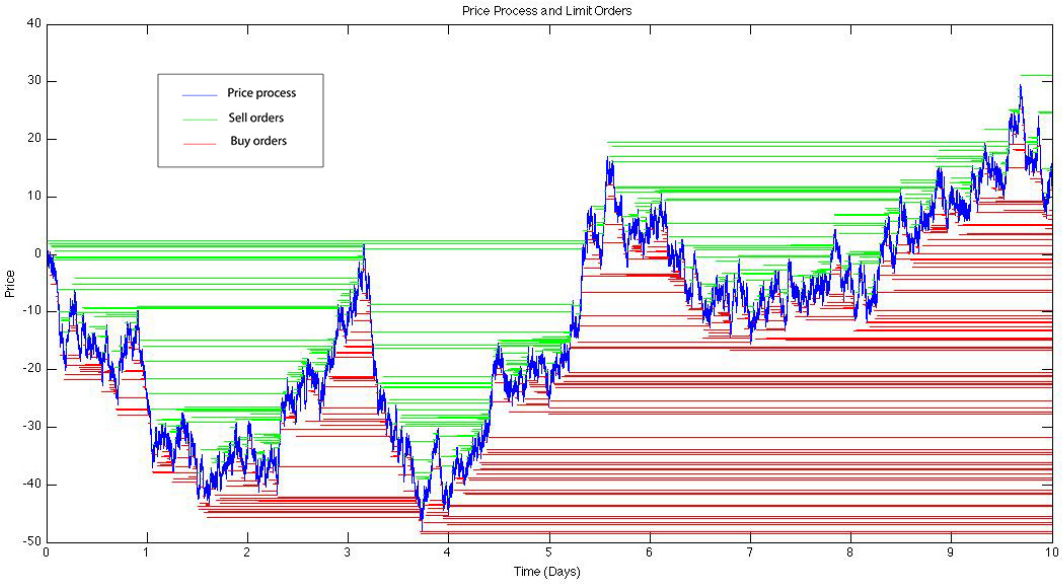
Since there are 2000 shares at the best ask quote, after the execution the best ask level hasn't changed. Next, we show what happens if a market buy order for 2500 shares is placed:

Price	Book at time t		Δt	Book at time $t + \Delta t$	
59	1000		...	1000	
58	1500		-500	1000	*Ask*
57	2000	*Ask*	-2000		
56	-1200	*Bid*	...	-1200	*Bid*
55	-1000		...	-1000	
54	-800		...	-800	

In this case, 2000 shares will be bought at level 57 and the remaining ones at 58, moving the best ask one tick higher. In the next table, we show what happens when a limit buy order for 1500 shares at price 57:

Price	Book at time t		Δt	Book at time $t + \Delta t$	
59	1000		...	1000	
58	1000	*Ask*	...	1000	*Ask*
57			-1500	-1500	*Bid*
56	-1200	*Bid*	...	-1200	
55	-1000		...	-1000	
54	-800		...	-800	

Thanks to this additional limit buy order, the best bid quote moves from 56 to 57. Finally, we can see in a graph how the LOB evolves in time:



1.2 Existing literature and motivation

There exist two main types of research related to the order books: a first line in financial Economics and a second one in Physics. The former uses econometric models for a static order process, while the latter consists of artificial models, fully dynamic since the order process is affected by the changes in prices.

One of the first economic models was developed by Cohen et al. in [7] by allowing limit orders at two given prices (the best quotes). Using this framework and applying queueing theory, they managed to compute some relevant quantities like the average volume of stored limit orders, the expected time to execution and the relation between probability of execution and cancellation. Domowitz and Wang in [13] introduced arbitrary order placement and cancellation processes and multiple price levels. The distributions of the bid-ask spread, of the transaction prices and of the waiting times can be derived but since the processes they used are time-stationary, they don't respond to changes in the best bid or ask. Bollerslev et al. in [2] provide an empirical test of this model, showing a good prediction for the distribution of the spread but it cannot predict the price diffusion from which errors in the estimation of spread and of stored supply and demand arise. Finally, Rosu in [23] presents a model of the order book where strategic and informed traders can choose between limit and market orders, i.e. between execution price and waiting costs. He shows that in equilibrium the best quotes depend only on the numbers of buy and sell orders in the book.

The Physics literature addresses also the dynamic issues, since the order placement process reacts in response to changes in price. Initially Bak et al. in [1] and then Eliezer and Kogan in [15] and Tang in [24], presented limit prices of orders placed at fixed distance from the mid-point and then they are randomly shuffled until transactions occur. Maslov in [20] builds a model where traders don't use any particular strategy, but they exhibit a purely random order placement. This leads to anomalous behaviours both for the price diffusion and for the stored limit orders, which either go to zero or grow without bound, unless we assume equal probabilities for limit and market orders placement. This issue can be overcome if a Poisson order cancellation is introduced, as it is shown by Challet and Stinchcombe in [6].

In more recent years, many studies were produced on the empirical observation of the order books, in order to understand their statistical properties. Most importantly, Bouchaud et al. in [4] and [5] explain several important features of the order books. Firstly, the incoming limit order prices follow a power-law distribution around the current price with a diverging mean. Then they investigate the shape of the average order book and the typical life time of an order (until cancellation or execution) as a function of the distance from the best price. Moreover, they study the price impact function, finding a logarithmic dependence of the price response on the volume. Another interesting empirical study is provided by Eisler et al. in [14]: they focus on the first passage time of the order book prices needed to observe a given price change, the time to fill for executed limit orders and the time to cancel. They show that all these quantities decay asymptotically as power laws.

Some stochastic models for the LOB are given by Cont et al. in [10]. They initially present a model able to capture the short-term dynamics of the limit order book, while keeping it analytically tractable. They also study the probability of some relevant quantities (conditional on the state of the order book) such as an increase in the mid-price and execution of an order at the bid before the ask quote moves. Then in [9] Cont et al. show that the price impact of order book events is mainly driven by the order flow imbalance, between supply and demand at the best quotes. Finally in [8], Cont and Lallard use a stochastic model where orders arrivals are described as a Markovian queueing system, providing again effective results on several quantities of interest.

At the market microstructure level, the notion of *asset price* is ambiguous but very important in order for participants to choose their strategies. Delattre et al. in [12] study this problem, proposing a notion of efficient price. They consider as driving factor the imbalance of the LOB, i.e. the difference between the available volumes at the best quotes. This is assumed to be a deterministic function (order flow response function) of the distance between the best quotes and the efficient price. Such function is then estimated in a non parametric way and the efficient price is derived from it.

Lorenz and Osterrieder in [19] and Osterrieder in [21] present a stochastic model for the order book where a reference price process sweeps the LOB as it diffuses. Since there is no need to model market order flow, this model is quite tractable and provides the very first step for this thesis, as it will be explained in more detail in the following chapters.

The aim of this work is to present a new approach to the order book modelling, starting from the idea of having a reference price process, which we call *virtual* in what follows, that diffuses independently of the limit order flow and determines orders executions. We therefore propose several models with different features and complexity levels, the aim being to compute some quantities of interest and to provide an additional point of view in the existing literature of the LOB modelling, believing that this framework can be used as a tool to expand the research in this field.

1.3 Notation

In the next chapters the following notation will be used, unless otherwise specified.

- $(\Omega, \mathcal{F}, \mathbb{F}, P)$ is a filtered probability space, where $\mathbb{F}$ is a filtration $(\mathcal{F}_t)_{0\leq t\leq T}$.
- $E[X]$ is the expectation of a random variable X (measurable with respect to $(\Omega, \mathcal{F})$) computed with the probability measure P.
- $S = (S_t)_{0\leq t\leq T}$ refers to a stochastic process with values in $\mathbb{R}$, adapted with respect to $(\Omega, \mathcal{F}, \mathbb{F}, P)$.
- $B = (B_t)_{0\leq t\leq T}$ is a Brownian motion with respect to $(\Omega, \mathcal{F}, \mathbb{F}, P)$.
- $L^\alpha = (L^\alpha_t)_{0\leq t\leq T}$ refers to the local time of the Brownian motion at level $\alpha \in \mathbb{R}$. A complete description of local times can be found, for instance, in [22], Chapter VI.
- $M = (M_t)_{0\leq t\leq T}$ is the running maximum of the Brownian motion: $M_t = \sup_{0\leq s\leq t} B_s$.
- $T = (T_x)_{x\geq 0}$ is the process of first passage times at levels $x \geq 0$ for the Brownian motion. Note that T is a subordinator.
- $\mathbb{I}_A(x)$ is the indicator function of a set A. $\mathbb{I}_A(x) = 1$ if $x \in A$ and $\mathbb{I}_A(x) = 0$ if $x \in A^c$.

Chapter 2

The avalanche approach

In this chapter we present a first approach to the LOB modelling aimed at introducing the main framework of this thesis. We use a virtual price process determining executions in the LOB and focus on the first issues arising from this initial step of our modelling. In particular, we will see how the study of the evolution of the number of orders in the LOB can be addressed using what is called *avalanche approach*: this method allows to compute some quantities analytically, but it requires several simplifications, given the high complexity of the state space.

2.1 Introduction

The first question we want to answer is: *how many orders are in the LOB at a given time?* A simple initial model is the following. Let $S = (S_t)_{0 \le t \le T}$ be the stock (mid-) price process and assume that orders are placed at random times $(T_i)_{i \ge 1}$ at level $S_{T_i} \pm \mu$ (we have to consider both buy and sell orders). For instance, we could assume that times T_i's are relative to a Poisson process and μ is Gamma distributed: however, this would already represent a model very difficult to study.

Let us consider a simpler setting. Let S be a process (which will be studied more in depth in the following chapters) defined by $S_t = \sigma B_t$, where $\sigma > 0$. Hence, S can be seen as a log-price process. Let us also assume that μ is constant and the order volume is 1. We will study only the sell side since the same considerations hold for the buy side. We also introduce a deterministic sequence of order times, i.e. we work in business time.

From now on, we call S the *virtual price process.* This element is what distinguishes the present work from most of the literature concerning market microstructure. The idea is very similar to what is outlined in by Osterrieder in [21] and by Lorenz and Osterrieder in [19].

Another key feature in this first description is the assumption that the number of orders placed up to time t at a limit price x is L_t^α where $\alpha = (x - \mu)/\sigma$.

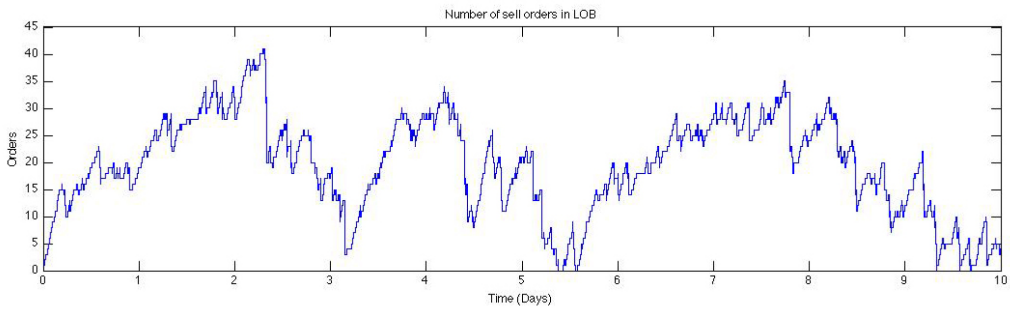
Therefore, all orders at a given level are executed when the virtual price process crosses it. Let X_t denote the number of executed orders in $[0, t]$ (on one side, say the sell side). Then, using the occupation times formula ([22], Chapter VI, Corollary 1.6), we have

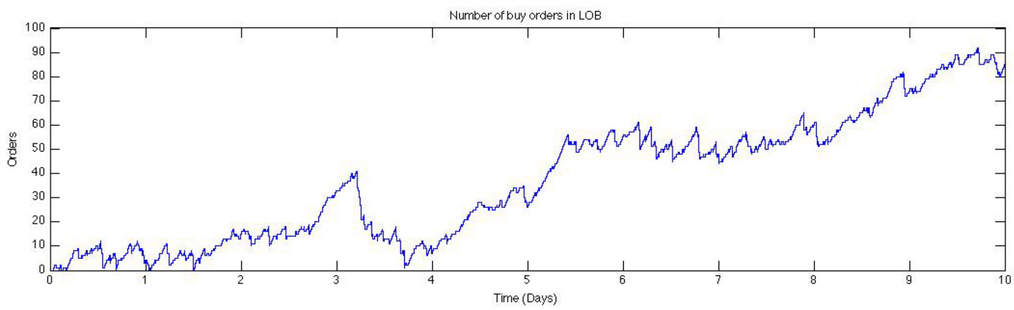
$$X_{t+\epsilon} - X_t \approx \int_{m_{t,t+\epsilon}-\mu}^{M_{t,t+\epsilon}-\mu} L_t^a \, da = \int_0^t \mathbb{I}_{[m_{t,t+\epsilon}-\mu, M_{t,t+\epsilon}-\mu]}(\sigma B_u) \, du \tag{2.1}$$

where $m_{t,t+\epsilon}$ and $M_{t,t+\epsilon}$ are the minimum and maximum of σB between time t and $t+\epsilon$. This approximation doesn't take into consideration the orders executed before time t and the ones placed in $[t, t+\epsilon]$. Therefore, since even a very simplified model faces some complex problems in its early stages, in what follows we study the problem from another point of view: the *avalanches*.

2.2 Study of avalanches

The number of orders in the LOB evolves with avalanches. In other words, limit orders may accumulate on some levels and when the price process crosses those values, we will see a sudden decrease of the number of orders in the LOB. In the next graph we show a possible realisation of the evolution of the number of orders in the LOB, where the avalanche dynamics are more clear.





Fix $\epsilon > 0$. We call *ϵ-avalanche* an avalanche such that the time length between two consecutive executions is always less than ϵ. Let us then define two related quantities that we are going to study for the first avalanche.

- The *avalanche length* (of an ϵ-avalanche) is the difference between the last and the first limit prices of executed orders.
- The *avalanche size* (of an ϵ-avalanche) is the number of executed orders during the entire length.

In this section we start from the idea of Dr Rheinlander, who pioneered the *avalanche* approach in market microstructure. He set out the main roots of this method and developed the study of the distributions of avalanche length and size. On my own, I then deepened the analysis of the sojourn problem.

2.2.1 Probability of avalanche length

We defined the length of the first ϵ-avalanche to be the number of order levels that get executed between the first order execution and the last order execution such that there is no new execution during the next ϵ-interval. Assume that the first order is executed at level $x > 0$ at time $T_{x/\sigma}$. Let $\delta > 0$:

$$\begin{aligned} &P(\text{Avalanche Length } > \delta) = \\ &\quad = P(M \text{ has no level stretch } > \epsilon \text{ in } [T_{x/\sigma}, T_{(x+\delta)/\sigma}]) + R_\epsilon \\ &\quad = P(T \text{ has no jumps with size } > \epsilon \text{ in } [x/\sigma, (x+\delta)/\sigma]) + R_\epsilon \end{aligned}$$

where R_ϵ is an error term. Define the counting process

$$N^\epsilon_y = \#\{\text{ jumps of } T \text{ with size } > \epsilon \text{ up to } y\}$$

which is a Poisson process with intensity $\lambda_\epsilon > 0$. Therefore,

$$P(T \text{ has no jumps with size } > \epsilon \text{ in } [x/\sigma, (x+\delta)/\sigma]) = P(N^\epsilon_{\delta/\sigma} = 0) = e^{-\lambda_\epsilon \delta/\sigma}$$

where, thanks to Lévy-measure ν of T, we have

$$\lambda_\epsilon = E\left[\sum_{0<y\leq 1} \mathbb{I}_{\{\Delta T_y > \epsilon\}}\right] = \int_\epsilon^\infty \nu(dy) = \int_\epsilon^\infty \frac{dy}{\sqrt{2\pi y^3}} = \sqrt{\frac{2}{\pi\epsilon}}.$$

We can actually prove that the error term R_ϵ is negative and converges to 0 as $\epsilon \to 0^+$. In fact, if the avalanche length is greater than δ then M has definitely no level stretch greater than ϵ in $[T_{x/\sigma}, T_{(x+\delta)/\sigma}]$, but the vice-versa is not always true (intuitively, this happens if B increases "slowly" so that M has no level stretch greater than ϵ but at the same time it doesn't touch the previously placed orders). Hence,

$$P(\text{Avalanche Length } > \delta) \leq P(M \text{ has no level stretch } > \epsilon \text{ in } [T_{x/\sigma}, T_{(x+\delta)/\sigma}])$$

which shows that $R_\epsilon \leq 0$. Moreover, we have already calculated the right-hand side of this inequality:

$$\exp\left\{-\sqrt{\frac{2}{\pi\epsilon}}\frac{\delta}{\sigma}\right\}$$

which is infinitesimal as $\epsilon \to 0^+$. Therefore,

$$\lim_{\epsilon\to 0^+} P(\text{Avalanche Length } > \delta) - R_\epsilon = 0$$

which proves, together with $R_\epsilon \leq 0$, that R_ϵ is infinitesimal as $\epsilon \to 0^+$.

2.2.2 Probability of avalanche size

Let us assume once again that the first ϵ-avalanche begins with an execution at limit price x and ends at price y. Then the number Y of executed orders during its length can be modelled by the following integral:

$$Y = \int_0^{T_{y/\sigma}} \int_x^y L_t^{\frac{a-\mu}{\sigma}} \, da \, dt. \tag{2.2}$$

For the time being, let us replace the stopping time $T_{y/\sigma}$ with a constant time horizon T. In this case, using the occupation times formula just like we did in (2.1), formula (2.2) becomes

$$Y = \int_0^T \int_x^y L_t^{\frac{a-\mu}{\sigma}} \, da \, dt = \sigma \int_0^T \int_0^t \mathbb{I}_{[\frac{x-\mu}{\sigma}, \frac{y-\mu}{\sigma}]}(B_u) \, du \, dt. \tag{2.3}$$

We can therefore study the distribution of Y. Fix $z \in \mathbb{R}^+$:

$$\begin{aligned} P(Y \leq z) &= \\ &= P\left(\text{Time spent by } B \text{ in } \left[\frac{x-\mu}{\sigma}, \frac{y-\mu}{\sigma}\right] \text{ before } T \text{ is } \leq \sqrt{\frac{2z}{\sigma}}\right) \\ &= P\left(\int_0^T \mathbb{I}_{[\frac{x-\mu}{\sigma}, \frac{y-\mu}{\sigma}]}(B_u) \, du \leq \sqrt{\frac{2z}{\sigma}}\right). \end{aligned} \tag{2.4}$$

Hence, in the end, studying the law of Y actually means solving an occupation time problem, also known as sojourn problem, to which we dedicate Appendix A.

More specifically, in Appendix A we study the distribution of the following random variable:

$$\int_0^t \mathbb{I}_{[b,c]}(B_u)\, du$$

where t is a positive constant and $0 \leqslant b < c \leqslant +\infty$. In other words, we want to study the law of the time spent by a Brownian motion starting at 0 in a specified interval $[b, c]$ within the time horizon t. The case $b = 0$ and $c = +\infty$ is a well known result obtained by P. Lévy, the so-called arc-sine law, and well explained, for instance, in [18]. Another remarkable result in this direction was achieved by G. Fusai in [16], where the characteristic function of such an occupation time is calculated. Finally, it is important to highlight that this distribution has been found with more than one method and the result can be found in [3], p. 166, formula 1.7.4.

In Appendix A, M. Kac's method (outlined in [17]) is used, since it is one of the first classical approaches in literature to this kind of problems. Our goal is to find the distribution for $0 < b < c < +\infty$. We show how to find a double Laplace transform used in Kac's method and a further step for future research in this direction would be to invert it analytically in order to find the sojourn distribution.

2.3 Conclusions

In this chapter we presented the avalanche approach as a first step towards a stochastic modelling of the LOB. Given the huge state space, it is clear how complex such modelling can be even with such a simple approach. In any case, we managed to compute some quantities relevant for this framework: the distributions of avalanche length and size

Finally, we believe this setting can lead to some interesting developments, however in the next chapters we will focus our attention on articulated models still involving the idea of a virtual price process.

Chapter 3

Some developed models

In this chapter we outline a complete stochastic model for the LOB: similarly to the previous chapter, the main idea is a virtual price process sweeping the orders accumulated in the book at the best quotes. The dynamics of the limit orders and cancellations are explained in detail. Market orders are slightly more complex and, therefore, are treated according to two approaches: initially a numerical one and then an analytical one, matching the first passage time distribution to empirical data. Finally, we show different results related to several quantities of interest. In fact, this is the main model developed in this thesis, as it is at the same time quite adaptable to empirical data but also mathematically treatable.

3.1 Introduction

Let us consider one liquid stock and the relative LOB. This model is made up of a fixed price grid $\{k\,\tau\}_{k\in\mathbb{Z}}$, where $\tau > 0$ is the tick size. Since prices can be negative, they may be thought of as log-prices without loss of generality. At time 0, the best ask price is τ while the best bid is $-\tau$. Moreover, as already proposed in the existing literature (see Cont and De Lallard in [8]), we consider an initial full book: i.e. each quote (except from price 0) has one order.

The spread is defined as the difference between the best quotes: "best ask − best bid". From the previous description it is clear that at time 0 the spread is equal to $2\,\tau$. Just like in [8], in this model the spread is assumed to be (tight and) constant. This is a meaningful approximation because such a behaviour can be observed in empirical studies ([8], [9]). In addition, when the liquidity at one quote is consumed, an order between the best quotes is placed by a trader to get time priority or by a market maker (if there is one) to provide liquidity. In the following table, we give an intuitive idea of what happens in case of an upward movement of the best quotes.

Price	Book at time t		Δt	Book at time $t + \Delta t$	
61	2		...	2	
60	7		...	7	
59	5		...	5	*Ask*
58	1	*Ask*	-1	0	
57	0		...	-1	*Bid*
56	-8	*Bid*	...	-8	
55	-9		...	-9	
54	-3		...	-3	

A market order is placed at quote 58: since all liquidity at such quote is consumed, the best ask moves upwards and an additional limit buy order is placed at quote 57, making the best bid move one tick upwards as well..

This model features only three types of *order book events*: limit orders, cancellations and market orders and their dynamics are explained in the rest of the chapter.

3.2 Limit orders and cancellations

Limit orders are placed in correspondence of the jumps of a Poisson process with intensity $\lambda > 0$: the drawback is that the choice of a memory-less process is not realistic, but this is just a starting point that can be improved (e.g. via a time-varying intensity). Every time the process has a jump, an order is placed in the book. More precisely, let us call $\{T_i\}_{i\geq 1}$ the sequence of stopping times associated with the Poisson process. The inter-arrival time is exponentially distributed with parameter λ:

$$T_{i+1} - T_i \sim Exp(\lambda).$$

At each time T_i, firstly it is chosen on which side of the book the order is placed. Ask side and bid side are equally likely: the order can be placed on one of the two sides with probability $1/2$. Once the side is picked, the order is placed μ ticks away from the current mid-price, where μ is a random variable defined by

$$P[\mu = j] = \frac{k}{j^{1+\alpha}}, \qquad j = 1, \dots, 100.$$

Here, $\alpha > 0$ and k is a normalizing constant given by

$$k = \left[\sum_{i=1}^{100} i^{-(1+\alpha)} \right]^{-1}.$$

This power-law is justified by several empirical studies: for instance, Bouchaud et al. in [4] and [5], and Cont et al. in [9] and [10]. Limiting the possible values of μ to $\{1,\ldots,100\}$ ticks is considered in the literature as well (see Bouchaud et al. in [4]). The value 100 is arbitrarily chosen and observed for a stock in [4]. In our model all orders are of unit volumes: this feature can be improved (with iid variables modelling orders' volumes, for instance) or, as it is done by Cont et al. in [8] and [10], we can think of the unit volume as the average of actual orders' volumes.

Cancellations are modelled in the same way, considering an independent Poisson process with intensity $\gamma > 0$. The corresponding sequence of stopping times is $\{T_i^c\}_{i\geq 1}$. At every time T_i^c an order is placed (with equal probability for each side of the book) at μ^c ticks away from the mid-price, where

$$P[\mu^c = j] = \frac{k^c}{j^{1+\beta}}, \qquad j = 1,\ldots,100.$$

Again, we have $\beta > 0$ and

$$k^c = \left[\sum_{i=1}^{100} i^{-(1+\beta)}\right]^{-1}.$$

If the chosen quote has only one order, then no cancellation happens: this is a technical condition imposed to keep the full-book structure.

Market orders are more complex and the next section is devoted to their dynamics: this aspect is what makes the model deeply different from most of the previous literature, while its main ideas are in line with the models presented by Osterrieder in [21] and by Osterrieder and Lorenz in [19].

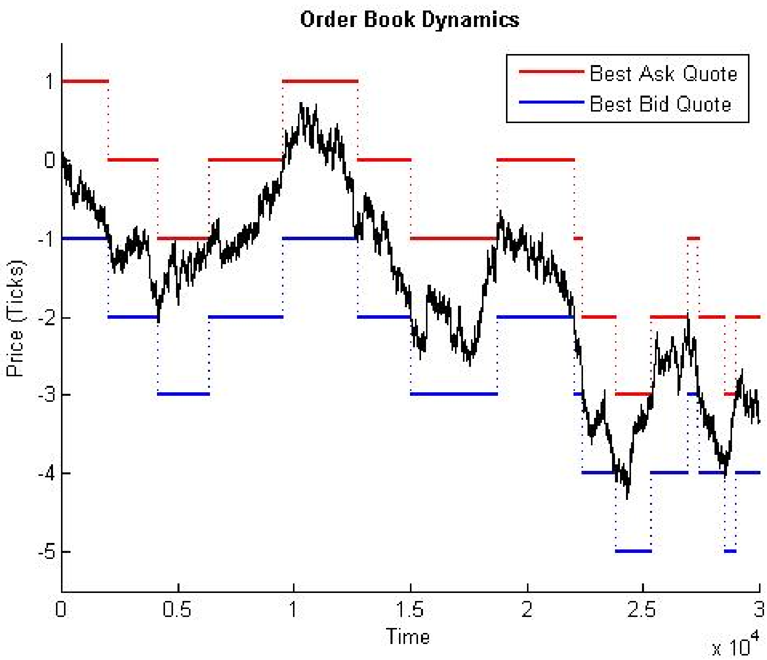
3.3 Market orders

3.3.1 First approach

Let us introduce the virtual price process S, like in the previous chapter. This stochastic process is given by:

$$S_t = \sigma B_t, \qquad t \geq 0, \tag{3.1}$$

where $\sigma > 0$. Now the idea is that the market order that consumes all liquidity at one quote (hence provoking a price change) is placed when S hits one of the two barriers represented by the best quotes. In the following graph, we can observe a possible path:



This model looks quite simple but, nonetheless, it is a good starting point, both because it shows a realistic behaviour anyway and because it is mathematically tractable. It will also be improved, introducing a time-varying volatility parameter σ_t.

3.3.2 Numerical simulations

In order to investigate this setting, numerical simulations are essential. Moreover, execution dynamics can be explained more precisely starting from an approximation of S. This is needed in order to control the number of quotes hit on each single time step by the Brownian motion. A total freedom would make the model very hard to be studied, hence, we switch to a discrete setting.

Let us consider a sequence of iid random variables $\{\xi_j\}_{j\geq 1}$ such that

$$E[\xi_1] = 0 \quad \text{and} \quad Var[\xi_1] = s^2$$

where $0 < s^2 < \infty$. Define with Z the partial sums

$$Z_0 = 0, \quad Z_k = \sum_{j=1}^{k} \xi_j, \quad k \geq 1.$$

From Z we can build a continuous linear interpolation Y in the following way:

$$Y_t = Z_{\lfloor t \rfloor} + (t - \lfloor t \rfloor)\xi_{\lfloor t \rfloor + 1}, \qquad t \in \mathbb{R}^+,$$

where $\lfloor t \rfloor$ indicates the largest integer smaller than t (i.e. the integer part). Properly scaling the process Y (in space and time):

$$X_t^{(n)} = \frac{1}{s\sqrt{n}} Y_{nt}, \qquad t \in \mathbb{R}^+,$$

for $n \in \mathbb{N}$, we obtain the so-called Donsker's theorem.

Theorem 1 *Using the objects just introduced, we have*

$$(X_{t_1}^{(n)}, \ldots, X_{t_k}^{(n)}) \xrightarrow[n\to\infty]{d} (B_{t_1}, \ldots, B_{t_k})$$

for all $k \in \mathbb{N}$ and $(t_1, \ldots, t_k) \in \mathbb{R}^k$, where the convergence is in distribution.

In our model we choose $s = 1$ and the sequence $\{\xi_j\}_{j\geq 1}$ such that

$$P[\xi_1 = 1] = P[\xi = -1] = \frac{1}{2}.$$

Therefore, once we fix an integer value for n high enough, we can approximate the noise process with the random walk provided by Donsker's theorem:

$$S_t = \sigma B_t \approx \sigma X_t^{(n)} = \frac{\sigma}{\sqrt{n}} Y_{nt}.$$

Let us now see step by step what happens when the virtual price process hits a barrier. Firstly, note that for times t of the form $t = k/n$, $k \in \mathbb{N}$, we have

$$X_t^{(n)} = \frac{1}{\sqrt{n}} Y_k = \frac{1}{\sqrt{n}} Z_k = \frac{1}{\sqrt{n}} \sum_{j=1}^{k} \xi_j.$$

Hence, a possible way to proceed is the following. If we take one day as time unit, then the time increment in our model (i.e. the time step used in simulations) can be taken as $h = 1/n$. Therefore, one day is made up of n time increments each of length h. The status of the virtual price process is checked every second, i.e. after each increment of length h. This means that, in order to see a price change, we have to add one realisation ξ_j to the random walk and check if the approximated virtual price process has crossed one of the barriers. Hence, we have

$$\begin{aligned} S_t &\xrightarrow{\text{after time } h} S_{t+h} \\ \sigma X_{k/n}^{(n)} &\xrightarrow{\text{after time } h} \sigma X_{(k+1)/n}^{(n)} \\ \frac{\sigma}{\sqrt{n}} \sum_{j=1}^{k} \xi_j &\xrightarrow{\text{after time } h} \frac{\sigma}{\sqrt{n}} \sum_{j=1}^{k+1} \xi_j. \end{aligned}$$

Note that we check only *at* every second and not between the seconds: this shouldn't be considered a drawback of the model because the possible increment of the random walk, $\sigma/\sqrt{n}$, is very small (if n is chosen to be high enough, as we do).

Let us assume that the buy market order that consumes all liquidity at the best ask is placed at time $t = k/n = k\,h$ (in what follows a superscript "new" will indicate the value of the quantity after the execution at the same time t). For sell market orders the dynamics are exactly the same. If the current best ask quote is $p_A(t)$, then

$$\frac{\sigma}{\sqrt{n}} \sum_{j=1}^{k} \xi_j \geq p_A(t).$$

This means that all orders at $p_A(t)$ are executed and the mid-price will move up by one tick. This is accomplished through the following updates

$$\begin{aligned} p_A^{new}(t) &= p_A(t) + \tau \\ p_B^{new}(t) &= p_B(t) + \tau. \end{aligned}$$

where $p_B(t)$ and $p_B^{new}(t)$ are the best bid quotes before and after the transaction at time t. The new volume $v_A^{new}(t)$ at the best ask quote will become the volume at $p_A^{new}(t)$. Then, in order to keep the spread constant and equal to $2\,\tau$, a new buy limit order is placed at the quote $p_B^{new}(t)$ (where the last execution just happened) and this makes the new volume at the best bid $v_B^{new}(t) = 1$.

Finally, it is important to understand why the choice of a discrete random walk is necessary, especially for simulations. Dealing with the Brownian motion, the increment after each time step h is normally distributed and can theoretically take values in all $\mathbb{R}$. Hence, it is useful to have control on the Brownian shock's size. With Donsker's theorem approximation, we can limit the increment of the approximated virtual price process and impose that at every step at most one barrier is hit. Therefore, we just need that each increment (in absolute value) is smaller than or equal to the tick size:

$$\frac{\sigma}{\sqrt{n}} \leq \tau,$$

which implies that n has to be chosen such that

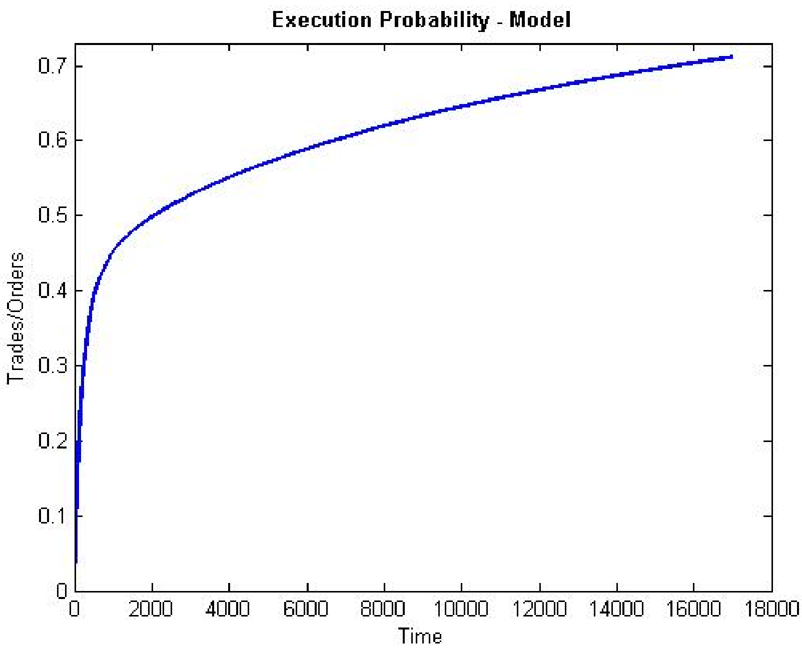
$$n \geq \left(\frac{\sigma}{\tau}\right)^2.$$

3.3.3 Probability of executions

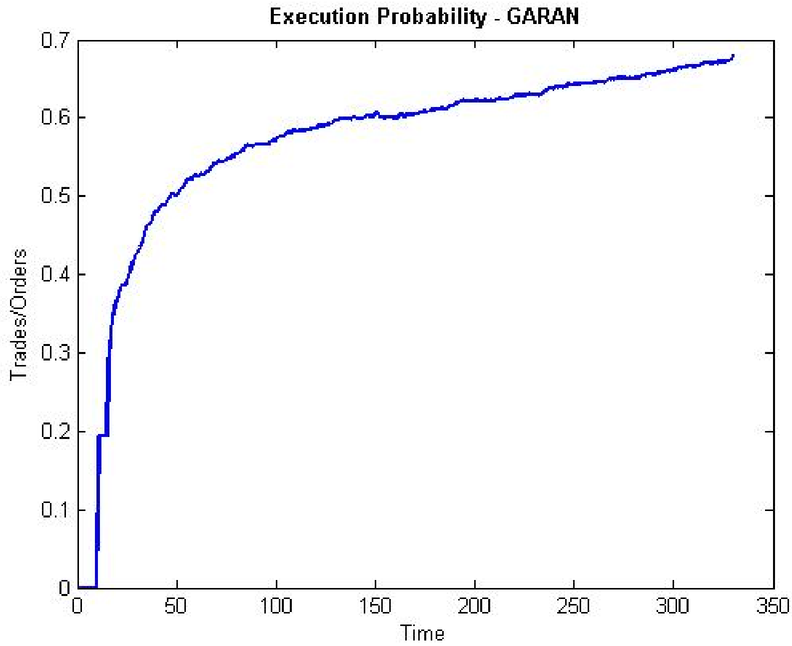
Using this setting, we have studied the probability of executions through simulations. This quantity is defined as the ratio between number of trades and number of orders in the book. The results show a shape very similar to what was found in Istanbul Stock Exchange with the following parameters choice:

$$\sigma = 50$$
$$\lambda = 0.6$$
$$\alpha = 0.6$$
$$\gamma = 0.01$$
$$\beta = 50.$$

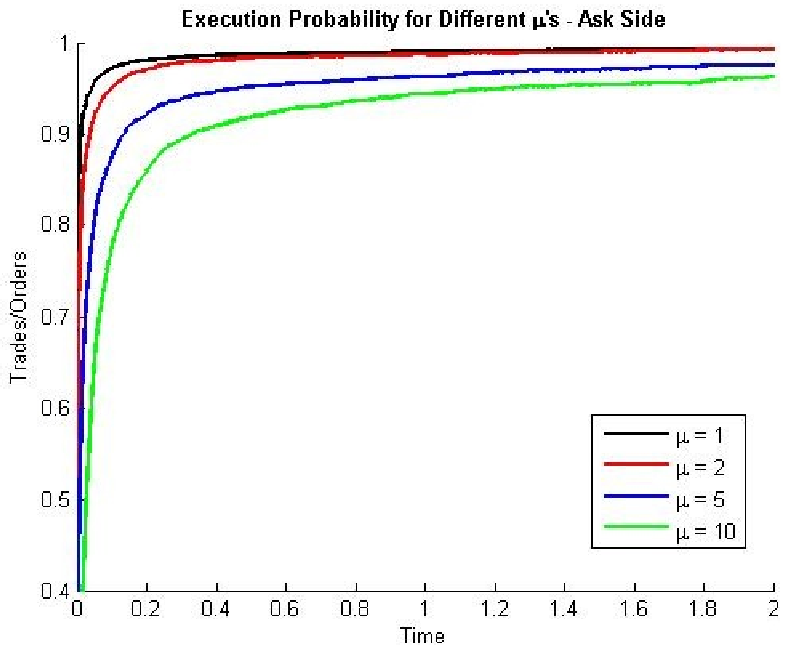
A high value for σ let us observe many trades. The choices for λ and γ make cancellations much less frequent than limit orders. The value for α is the one found by Bouchaud et al. in [4]. When a cancellation happens, big β results in a high probability of choosing a quote close to the mid-price: this is meaningful because usually limit orders far away from the mid-price are placed by patient investor, unlike those close to the mid-price. The following plot was done with 1000 simulations over a period of 100 days.



The following graph is what has been observed by M. Valenzuela (LSE) and I. Zer (LSE) for a very liquid stock, GARAN, on the Istanbul Stock Exchange, averaging over all working days in July 2008.



A further step should be to study the same quantity, but focusing on orders placed at a fixed displacement from the mid-price. This would answer the question: if at time 0 an order is placed at a distance of μ ticks away from the best quote, what is the probability of it being executed? This is a valuable piece of information for a trader: if such value is high, then it is very likely that many informed agents are operating in the market and she could adjust her aggressiveness in placing orders. The following plot shows what happens with numerical simulations using different displacements.



3.3.4 A refinement

It has been observed that the actual first passage time distribution of transaction prices is quite different from the one implied by the model described so far. This is highlighted in particular by Eisler et al. in [14]. The well known distribution of the first passage time for a Brownian motion at a given barrier has density

$$h_{BM}(t) = \frac{\Delta}{\sqrt{2\pi}} t^{-3/2} \exp\left(\frac{\Delta^2}{2t}\right)$$

(where Δ is the barrier's distance at time 0), which is asymptotically proportional to $t^{-3/2}$ for long times. Eisler et al. in [14] show that this behaviour is realistic for long times, while for short times h_{BM} is very different from what is observed. Therefore they fit the actual density in the following way:

$$h(t) = \frac{C\,t^{-\theta}}{1 + [t/F(\Delta)]^{-\theta+\theta'}} \tag{3.2}$$

where C is a normalising constant, F is a function of the barrier Δ only and θ and θ' are two given constants. Normalisation conditions also require that $1 < \theta < 2$ and $\theta' < 1$. The three parameters F, θ and θ' can be estimated from observed data. Moreover, from the shape of h, we can see that for $t \ll F(\Delta)$ we have $h(t) \propto t^{-\theta'}$, while for $t \gg F(\Delta)$ we have $h(t) \propto t^{-\theta}$. As mentioned before, it is actually found in [14] that θ has values close to 3/2 for five different stocks and for different values of Δ.

In what has been presented so far, we firstly defined the virtual price process as a Brownian motion, which has h_{BM} as corresponding density for the first passage time. Now, we proceed the other way around: we take h as density of the first passage time of a new virtual price process, which can be created using a result by Davis and Pistorius in [11]. Before applying their theorem, let us properly define the new problem.

Inverse first-passage time problems

Given a probability distribution H on $\mathbb{R}_+$ and a family $\mathcal{Y}$ of càdlàg stochastic processes $Y = \{Y_t\}_{t\geq 0}$ starting from 0, the inverse first passage time problem to zero (IFTP) is to find a $Y \in \mathcal{Y}$ such that the first-passage time T_0^Y of Y below zero,

$$T_0^Y := \inf\{t \geq 0 : Y_t < 0\},$$

follows the distribution H. Let us restrict ourselves to a family $\mathcal{Y}$ of Gaussian processes. Let W be a standard Wiener process on the previously defined probability space $(\Omega, \mathcal{F}, P)$ and let A be a random variable independent of W. Consider the set of linear Gaussian processes Y of the form

$$Y_t = A + \int_0^t \nu\, \sigma_s^2\, ds + \int_0^t \sigma_s\, dW_s, \qquad t \geq 0,$$

where $\sigma : \mathbb{R}_+ \to \mathbb{R}$ is a function such that

$$I_t := \int_0^t \sigma_s^2\, ds < \infty, \qquad t \geq 0.$$

Note that Y is equal in law to the time-changed Brownian motion $\{X(I_t), t \geq 0\}$ where

$$X_t = A + \nu\, t + W_t, \qquad t \geq 0.$$

From now on, by P_x we denote the measure P conditioned on $Y_0 = x$. Let us also recall the form of the first-passage time distribution $K_a^{(\nu)}(t) = P_a(T_0^X \leq t)$ to a constant level for a Brownian motion with drift ν. Under P_a the stopping time T_0^X has Laplace transform given by

$$E_a[e^{-qT_0^X}] = \exp\left(-|a|\left[\nu + \sqrt{\nu^2 + 2q}\right]\right),$$

while the density $k_a^{(\nu)}(t)$ is given by

$$k_a^{(\nu)}(t) = \frac{|a|}{\sqrt{2\pi t^3}} \exp\left(-\frac{(a+\nu t)^2}{2t}\right).$$

For any distribution H with density h, a solution to the IFTP can be found by suitably choosing the form of the volatility function σ:

Theorem 2 *Let H be a cumulative distribution function on $\mathbb{R}_+ \cup \{\infty\}$ with $H(0) = 0$ that is continuous on $\mathbb{R}_+$ with density h. Fix $a > 0$ and ν satisfying*

$$\nu \leq -\frac{1}{2a} \log(H(\infty)), \tag{3.3}$$

and define σ_s for $s \geq 0$ by

$$\sigma_s^2 = \begin{cases} 0 \ , & 0 \leq s \leq s_0 := \inf\{s : H(s) > 0\} \\ \frac{h(s)}{k_a^{(\nu)}\left((K_a^{(\nu)})^{-1}(H(s))\right)} \ , & s_0 < s < \infty \ , \end{cases} \tag{3.4}$$

where $(K_a^{(\nu)})^{-1}$ denotes the right-continuous inverse of $K_a^{(\nu)}$. Then it holds that $T_0^Y \sim H$, that is,

$$P_a(T_0^Y \leq t) = H(t), \qquad t \in \mathbb{R}_+.$$

See [11] for a proof of this theorem.

A new virtual price process

We can now use the result outlined in Theorem (2) in our setting. Let us start from the distribution h in (3.2) of the first-passage time found by Eisler et al. in [14]. We can then build the volatility process σ via formula (3.4):

$$\sigma_s^2 = \begin{cases} 0 \ , & s = 0 \\ \frac{h(s)}{k_a^{(\nu)}\left((K_a^{(\nu)})^{-1}(H(s))\right)} \ , & s > 0 \ . \end{cases} \tag{3.5}$$

Note that condition (3.3) becomes

$$\nu \leq -\frac{1}{2a} \log(H(\infty)) = 0$$

since $H(\infty) = 1$, letting us choose $\nu = 0$. Therefore, our new virtual price process can be given by:

$$Y_t = \int_0^t \sigma_s \, dW_s, \qquad t \geq 0. \tag{3.6}$$

In fact, we can easily see that, thanks to Theorem (2), the probability density function of the first passage time of Y defined in (3.6) at a given barrier is exactly h from formula (3.2): fix a quote in the price grid Δ and define (with a slight change of notation)

$$T_\Delta^Y = \inf\{t \geq 0 : Y_t > \Delta\}.$$

Then, from Theorem (2), we obtain, for $t \geq 0$, the distribution of the first passage time for this newly defined virtual price process:

$$P\left(T_\Delta^Y \leq t\right) = \int_0^t h(s) \, ds = \int_0^t \frac{C \, s^{-\lambda_F}}{1 + [s/T_F(\Delta)]^{-\lambda_F + \lambda_F'}} \, ds. \tag{3.7}$$

3.4 Quantities of interest

This refined model, whose virtual price process is defined in (3.6), allows us to study some important quantities in addition to the first passage time's distribution calculated in (3.7). These quantities are useful in order to understand the LOB's dynamics and to choose trading strategies.

Let us start from some distributions introduced by Cont et al. in [10] and [8], where they were studied via queueing theory. Define T_A and T_B as the stopping times when all liquidity is consumed for the first time at the best ask and at the best bid, respectively. Hence, in our model, recalling that τ stands for the tick size,

$$T_A = \inf\{t \geq 0 : Y_t > \tau\}$$
$$T_B = \inf\{t \geq 0 : Y_t < -\tau\}$$

Notice that, thanks to the definition of this model and the Markov property of the virtual price process, it doesn't matter whether we start at time 0 or at another time, as long as we know the initial state of the book. Moreover, most of the results shown in this section are a straightforward consequence of the (already explained) identity in law between Y_t and W_{I_t}.

1. The probability of having a price increase before a price decrease is simply

$$P(T_A < T_B) = \frac{1}{2}.$$

This is due to the symmetry of increments for the virtual price process.

2. We can then derive the distribution of the duration before the first mid-price move: for $t > 0$, we have

$$\begin{aligned} P(T_A \wedge T_B \geq t) &= P(T_A \geq t, T_B \geq t) = P(T_\tau^W \geq I_t, T_{-\tau}^W \geq I_t) \\ &= P(T_\tau^W \wedge T_{-\tau}^W \geq I_t) = \int_{I_t}^{\infty} f_1(s)\, ds \end{aligned}$$

where we defined

$$T_\tau^W = \inf\{t \geq 0 : W_t > \tau\}$$
$$T_{-\tau}^W = \inf\{t \geq 0 : W_t < -\tau\}$$

and f_1 is a well known result (see [18]):

$$f_1(t) = \tau \sqrt{\frac{2}{\pi t}} \sum_{n=-\infty}^{\infty} (4n+1) \exp\left\{ -\frac{(4n+1)^2\, \tau^2}{2\, t} \right\}.$$

Therefore, we can write down explicitly the expression of the pdf for the distribution of the duration before the first mid-price move:

$$P(T_A \wedge T_B \in dt) = f_1(I_t)\, \sigma_t^2\, dt.$$

3. Let us now assume that we place a sell limit order at the best ask quote. We can calculate the distribution of getting the order executed before the mid-price moves down. A similar result holds for the bid side.

$$P(T_A < T_B,\, T_A \leq t) = P(T_\tau^W < T_{-\tau}^W,\, T_\tau^W \leq I_t), \quad t \geq 0$$

since $I(T_A) \sim T_\tau^W$ and $I(T_B) \sim T_{-\tau}^W$. Hence, the density function becomes

$$P(T_A < T_B,\, T_A \in dt) = f_2(I_t)\, \sigma_t^2\, dt$$

where, again from [18], we know that $f_2(t) = (2t)^{-1} f_1(t)$. Moreover, in order to further justify this fact, we could refer to the assumption in [14] according to which the time between the price reaching the quote and the agent's execution is negligible. This is meaningful if the agent we are considering simply waits for his order to get executed and doesn't cancel it (while all other agents, seen as a whole, can cancel their orders).

Other important quantities were studied by Cont et al. in [8] and [10]. These are mostly conditional distributions of some random variables given the shape of the book. For instance, the probability of making the spread (execution of two orders placed one at the best ask and one at the best bid) before the mid-price moves.

3.5 Conclusions

In this chapter we presented the main stochastic model for the LOB analysed in this thesis: we assumed a constant spread, limit orders and cancellations driven by a Poisson process and a power law and a full book structure. We considered a virtual price process given initially by a Brownian motion. Then, we managed to find a virtual price process matching the empirical density of the first passage time. This entire framework allowed us to compute several quantities of interest. However, there is room for future research: for instance, the main problem with this framework is to find a balance between the discrete and the continuous settings. In fact, the former allows a very nice description of partial executions but results in very difficult formulae. On the other hand, the latter gives easier ways for calculations but makes very difficult the modelling of partial executions. This issue is partially addressed in the next chapter.

Chapter 4

The local time approach

The modelling approach presented in this chapter involves again a virtual price process determining orders executions. Unlike the previous models, however, the LOB dynamics are now determined by downcrossings and executions. These aspects are more complex and harder to be analytically treated, but they are included in this thesis because they represent a different approach to this type of stochastic modelling of the LOB and could lead to more significant developments for future research.

4.1 Modelling via downcrossings

Let us partially resume the models described in previous chapters, introducing new dynamics for limit orders: from now on, the model outlined in this section will be referred to as "Model-1". As in section 3.1, consider a fixed grid for prices $\{k\tau\}_{k\in\mathbb{Z}}$, where $\tau > 0$ is the tick size. We assume that at time 0 all levels have one order except for level 0, in particular: levels in the grid greater than 0 have one sell limit order, levels smaller than 0 have one buy limit order, while level 0 is empty. Hence, at time 0, the best quotes are τ and $-\tau$.

In this model the bid-ask spread is kept constant and equal to 2τ. Moreover, to determine limit orders, let us introduce a virtual price process $\{S_t\}_{t\geq 0}$ defined as

$$S_t = \sigma\, B_t,$$

where $\sigma > 0$. The idea underlying the generation of limit orders is that they are placed according to *how long* the process S is *close* to level 0. The goal is to have continuos order arrivals (hence, a macroscopic point of view) and to link this heuristic notion to the local time of a Brownian motion. Let us start from the more intuitive scenario in which orders arrive discretely.

Fix $\epsilon > 0$. Define, for $n \in \mathbb{N}$ and $t > 0$,

$$\begin{aligned}\tau_0 &\equiv 0\\ \sigma_n &:= \inf\{s \geq \tau_{n-1} : |B_s| = \epsilon/\sigma\}\\ \tau_n &:= \inf\{s \geq \sigma_n : |B_s| = 0\}\\ D_t(\epsilon/\sigma) &:= \sup\{n \in \mathbb{N} : \tau_n \leq t\}.\end{aligned}$$

$D_t(\epsilon/\sigma)$ represents the number of *downcrossings* of $|S|$ from level ϵ to level 0 before time t. Model-1 assumes that orders can only be placed at the best quotes and the number of orders arrived by time t is $D_t(\epsilon/\sigma)$. Moreover, we assume that the only volume size available is ϵ/σ. Hence, for all times t before the first mid-price change, the volume of orders placed at the best quotes is

$$V_A^\epsilon(t) = V_B^\epsilon(t) = 1 + \frac{\epsilon}{\sigma} D_t(\epsilon/\sigma),$$

where V_A^ϵ and V_B^ϵ describe the ask and bid sides respectively. When $\epsilon \downarrow 0$ the order size becomes smaller and smaller but, at the same time, more and more orders arrive since more and more downcrossings are considered.

Intuitively, the virtual price process before the first mid-price change lies between the best quotes and limit orders are placed according to how much time it spends *close* to the actual mid-price (this is the idea behind the definition of local times). In fact, for a fixed $\epsilon > 0$, when $|S|$ touches 0 an order is placed if that is a downcrossing. In the limit $\epsilon \downarrow 0$, orders are placed at *all* hitting times, with a smaller and smaller size to avoid volume explosion. Therefore, we can use a well-established result for local times (see for instance [22], Chapter XII) and state that for all $t > 0$

$$\lim_{\epsilon \downarrow 0} \frac{\epsilon}{\sigma} D_t(\epsilon/\sigma) = L_t^0 \quad \text{a.s.}$$

In order to analyse executions and cancellations, Model-1 splits into total and partial executions that are discussed in the following subsections.

4.1.1 Total executions

The first and most basic way to model executions is assuming that *all* orders placed at one of the best quotes are executed when the virtual price process S hits one of the two barriers τ or $-\tau$. This happens either at time $T_{\tau/\sigma}$ or at time $T_{-\tau/\sigma}$.

Assume that S hits τ first: then the new best ask becomes 2τ while the new best bid is 0 and the new volumes are both one. In fact, in order to keep the spread equal to 2τ, a new buy limit order is placed right after the execution at level 0 (as we explained in section 3.1). This is meaningful because it implies that a buy market order makes the mid-price move upwards. In addition, the new limit order may be placed to get time priority. Note that at level $-\tau$ we still have $1 + \frac{\epsilon}{\sigma} D_{T_\tau}(\epsilon/\sigma)$ orders in the discrete setting or $1 + L^0_{T_\tau}$ in the continuous one.

4.1.2 Partial executions

In order to make Model-1 more realistic, we have to allow for partial executions. This is very important to compute relevant quantities needed in practice. In this new setting, when S hits level τ or $-\tau$, not all orders are executed but just some of them. The idea here is to have reflexive barriers at the best quotes and market orders (just like cancellations) are placed in the same way we used for limit orders.

To understand these dynamics, let us initially consider the discrete setting. Fix $\epsilon > 0$: the number of buy market orders and cancellations (since there is actually no difference for our present purposes) in the ask side is given by the downcrossings of $|\tau - S_t|$ from level ϵ to level 0. So we can just repeat what we have previously done and define, for the ask side,

$$\begin{aligned}
\tau_0^A &\equiv 0 \\
\sigma_n^A &:= \inf\{s \geq \tau_{n-1}^A : |\tau/\sigma - B_s| = \epsilon/\sigma\} \\
\tau_n^A &:= \inf\{s \geq \sigma_n^A : |\tau/\sigma - B_s| = 0\} \\
D_t^A(\epsilon/\sigma) &:= \sup\{n \in \mathbb{N} : \tau_n^A \leq t\}.
\end{aligned}$$

Therefore, for all times t before the first change in the mid-price, the number of orders at the best ask is

$$V_A^\epsilon(t) = \left[1 + \frac{\epsilon}{\sigma} D_t(\epsilon/\sigma) - \frac{\epsilon}{\sigma} D_t^A(\epsilon/\sigma)\right]^+$$

where we assume that for fixed $\epsilon > 0$ the volume of market orders or cancellations can only be ϵ/σ. The extension to the continuous case relies on the result (see [22], Chapter XII)

$$\lim_{\epsilon \downarrow 0} \frac{\epsilon}{\sigma} D_t^A(\epsilon/\sigma) = L_t^{\tau/\sigma} \quad \text{a.s.}$$

for all $t > 0$. Hence, when $\epsilon \downarrow 0$, for all times t before the first change in the mid-price, the number of orders at the best ask is

$$V_A(t) = \left[1 + L_t^0 - L_t^{\tau/\sigma}\right]^+.$$

Similarly, for the bid side, we have

$$\begin{aligned}
\tau_0^B &\equiv 0 \\
\sigma_n^B &:= \inf\{s \geq \tau_{n-1}^B : |\tau/\sigma + B_s| = \epsilon/\sigma\} \\
\tau_n^B &:= \inf\{s \geq \sigma_n^B : |\tau/\sigma + B_s| = 0\} \\
D_t^B(\epsilon/\sigma) &:= \sup\{n \in \mathbb{N} : \tau_n^B \leq t\}
\end{aligned}$$

and, for all times t before the first change in the mid-price, the number of orders at the best bid is

$$V_B^\epsilon(t) = \left[1 + \frac{\epsilon}{\sigma} D_t(\epsilon/\sigma) - \frac{\epsilon}{\sigma} D_t^B(\epsilon/\sigma)\right]^+$$

where we assume again that for fixed $\epsilon > 0$ the volume of market orders or cancellations can only be ϵ/σ. We can then use the following result for the extension to the continuous case:

$$\lim_{\epsilon \downarrow 0} \frac{\epsilon}{\sigma} D_t^B(\epsilon/\sigma) = L_t^{-\tau/\sigma} \quad \text{a.s.}$$

for all $t > 0$. So, taking the limit $\epsilon \downarrow 0$, for all times t before the first change in the mid-price, the number of orders at the best bid is

$$V_B(t) = \left[1 + L_t^0 - L_t^{-\tau/\sigma}\right]^+ .$$

It is important to notice that the first change in the mid-price happens when either V_A^ϵ or V_B^ϵ hit 0 in the discrete scenario or when either V_A or V_B hit 0 in the continuous scenario.

4.2 Modelling via excursions

Another possible way to model the order book is described in this section. This framework is equal to Model-1 for the price grid, the spread and the starting shape and we will refer to it as "Model-2".

Consider two stochastic processes: the usual virtual price S^1, that lives between the quotes, to determine market orders and cancellations. Then we will use an additional process S^2, called *auxiliary*, that regulates limit orders. More precisely, for $t > 0$,

$$S_t^1 = \sigma_1 B_t^1$$
$$S_t^2 = \sigma_2 B_t^2$$

where σ_1 and σ_2 are positive constants, while B^1 and B^2 are independent Brownian motions.

Let us now consider limit orders in the discrete case and fix $\epsilon > 0$. Unlike Model-1, in Model-2 the only order size allowed is

$$\sqrt{\frac{\pi\epsilon}{2\sigma_2}}.$$

The number of orders placed is instead related to the excursion processes and this is the main reason why we have to use two Brownian motions. In Appendix B a brief overview of the topic is given, but refer to [22], Chapter XII for a more detailed explanation. In what follows we use the notation introduced in Appendix B.

In our framework, i.e. Model-2, the number of orders arrived up to time t before the first price move, is given by $\eta_t(\epsilon/\sigma_2)$ (i.e. the number of excursions with duration bigger than ϵ). Hence, for the discrete case, the volume of limit orders placed at the best quotes is

$$V_A^\epsilon(t) = V_B^\epsilon(t) = 1 + \sqrt{\frac{\pi\epsilon}{2\sigma_2}}\,\eta_t(\epsilon/\sigma_2)$$

where η is here related to B^2. With these assumptions, the convergence to the continuous setting is easily given by the following result:

$$\lim_{\epsilon\downarrow 0} \sqrt{\frac{\pi\epsilon}{2}}\,\eta_t(\epsilon) = L_t^0 \quad \text{a.s.}$$

where L_t^0 is the local time of B^2 at level 0. So, as $\epsilon \downarrow 0$ the number of limit orders placed at the best quotes is

$$V_A(t) = V_B(t) = 1 + L_t^0.$$

The interesting part of this approximation in the discrete case is that, as it is observed in Appendix B,

$$\eta_t(\epsilon/\sigma) = N_{L_t^0}^{\epsilon/\sigma_2} = \sum_{s\le t} \mathbb{I}_{[\epsilon/\sigma_2,\infty[}(R(e_s))$$

which is a Poisson process (evaluated at a local time) and this gives an immediate intuitive interpretation of this quantity. Moreover, we can explicitly calculate the intensity of this Poisson process using the characteristic measure of the Poisson point process $R(e_s)$. Namely, we have that the intensity is

$$\bar{n}(]\epsilon/\sigma_2,\infty[) = \left(\frac{2\sigma_2}{\pi\epsilon}\right)^{1/2}.$$

Note that as ϵ decreases, the intensity increases and this is the behaviour we would expect from Model-2.

Similarly to what was done for Model-1, we can define market orders and cancellations in two ways that are discussed in what follows.

4.2.1 Total executions

The framework to model executions and cancellations is built exactly in same way developed in Model-1. Here we use S^1 as virtual price process and repeat what we observed in Subsection 4.1.1.

4.2.2 Partial executions

Again, for partial executions in Model-2 we have analogous considerations to those made in Subsection 4.1.2 using S^1 as virtual price process, but this time with excursions instead of downcrossings.

4.3 Quantities of interest

Given the previous definitions of Model-1 and Model-2, we can compute some of the quantities already studied in Section 3.4. Here we simply state the problems without solving all of them in depth, as they would require a much wider analysis, beyond the effort of this work.

Actually, if we consider Model-2 in these problems, nothing really changes. We just have to rewrite the same considerations done for Model-1 with the proper processes S^1 and S^2 and with η instead of D. Hence, we will only study Model-1 in what follows.

4.3.1 Duration of price moves

Let us start with the conditional distribution of the duration between price moves, given the shape of the book. For simplicity, consider the known shape we have at time 0.

Assume Model-1 under total executions: then the stopping time of the first price move is given by $T = T_{\tau/\sigma} \vee T_{-\tau/\sigma}$ and, therefore, its distribution is, for $u > 0$

$$P[T > u] = P[T_{\tau/\sigma} > u, T_{-\tau/\sigma} > u] = 2\, P[T_{\tau/\sigma} > u]$$

which is a well-known quantity. It is the same both for the discrete and the continuous cases.

If we allow for partial executions in Model-1, we need to define some more objects. In the discrete setting, for a fixed $\epsilon > 0$, set

$$\begin{aligned}
T^{A,\epsilon} &:= \inf\{t \geq 0 : V_A^\epsilon(t) = 0\} \\
&= \inf\left\{t \geq 0 : \left(1 + \frac{\epsilon}{\sigma}\left[D_t(\epsilon/\sigma) - D_t^A(\epsilon/\sigma)\right]\right)^+ = 0\right\} \\
T^{B,\epsilon} &:= \inf\{t \geq 0 : V_B^\epsilon(t) = 0\} \\
&= \inf\left\{t \geq 0 : \left(1 + \frac{\epsilon}{\sigma}\left[D_t(\epsilon/\sigma) - D_t^B(\epsilon/\sigma)\right]\right)^+ = 0\right\} \\
T^\epsilon &:= T^{A,\epsilon} \vee T^{B,\epsilon}.
\end{aligned}$$

We are therefore interested in the distribution of T^ϵ which seems to be quite difficult to study. The generalisation to the continuous scenario is straightforward:

$$
\begin{aligned}
T^A &:= \inf\{t \geq 0 : V_A(t) = 0\} \\
&= \inf\left\{t \geq 0 : \left(1 + L_t^0 - L_t^{\tau/\sigma}\right)^+ = 0\right\} \\
T^B &:= \inf\{t \geq 0 : V_B(t) = 0\} \\
&= \inf\left\{t \geq 0 : \left(1 + L_t^0 - L_t^{-\tau/\sigma}\right)^+ = 0\right\} \\
T &:= T^A \vee T^B.
\end{aligned}
$$

4.3.2 Mid-price increase

The conditional probability of a mid-price increase given the shape of the book is investigated in this section. In Model-1 with total executions, this quantity doesn't depend on the shape of the book and can be written, both for the discrete and continuous cases, as

$$
\begin{aligned}
P[\text{mid-price increase}] &= P[\, T_{\tau/\sigma} < T_{-\tau/\sigma} \,] \\
&= \frac{\tau}{\sigma\sqrt{2\pi}} \int_0^\infty t^{-3/2} \sum_{n=-\infty}^{\infty} (4n+1) \exp\left\{ -\frac{(4n+1)^2\tau^2}{2t\sigma^2} \right\} dt,
\end{aligned}
\tag{4.1}
$$

where the last equality follows from a well-established result (see [18], Chapter II).

If we allow for partial executions, then, recalling the notation previously introduced, we have

$$P[\text{mid-price increase}] = P[\, T^{A,\epsilon} < T^{B,\epsilon} \,]$$

in the discrete scenario. In the limit $\epsilon \downarrow 0$:

$$P[\text{mid-price increase}] = P[\, T^A < T^B \,].$$

These last two quantities can be computed like in formula (4.1).

4.3.3 Order execution before mid-price moves

The conditional probability of executing an order placed at the best quote before the mid-price moves, given the shape of the book, is very easy to write for Model-1 with total executions. In fact, thanks to the framework of that model, we can write such probability as

$$P[\, T_{\tau/\sigma} < T_{-\tau/\sigma} \,],$$

in the case of a sell limit order for instance (see formula (4.1) for the computation).

If we allow for partial executions, things are a bit more complex. The idea is that we assume that at time 0 we have volume a at the best ask and volume b at the best bid. Then we place a sell limit order of volume x at the best ask (analogous considerations can be done for the bid side): we want to compute the probability that our order gets executed before the mid-price moves. Define

$$\begin{aligned}
T_{a,x}^{A,\epsilon} &:= \inf\left\{t \geq 0 : a + x \leq \frac{\epsilon}{\sigma} D_t^A(\epsilon/\sigma)\right\} \\
T_{a+x}^{A,\epsilon} &:= \inf\left\{t \geq 0 : \left(a + x + \frac{\epsilon}{\sigma}\left[D_t(\epsilon/\sigma) - D_t^A(\epsilon/\sigma)\right]\right)^+ = 0\right\} \\
T_b^{B,\epsilon} &:= \inf\left\{t \geq 0 : \left(b + \frac{\epsilon}{\sigma}\left[D_t(\epsilon/\sigma) - D_t^B(\epsilon/\sigma)\right]\right)^+ = 0\right\} \\
T_{a,b,x}^{\epsilon} &:= T_{a+x}^{A,\epsilon} \vee T_b^{B,\epsilon}.
\end{aligned}$$

The interpretation of these objects is the following. $T_{a,x}^{A,\epsilon}$ is the time in which our order gets executed. $T_{a+x}^{A,\epsilon}$ is the first time when the orders volume at the best ask is 0 starting with $a + x$ orders. $T_b^{B,\epsilon}$ is the first time when the orders volume at the best bid is 0 starting with b orders. So $T_{a,b,x}^{\epsilon}$ is simply the time in which the mid-price moves. Hence, the required quantity can be written as

$$P\left[T_{a,x}^{A,\epsilon} < T_{a,b,x}^{\epsilon}\right].$$

In the continuous scenario, we just have to define the analogous quantities

$$\begin{aligned}
T_{a,x}^{A} &:= \inf\left\{t \geq 0 : a + x \leq L_t^{\tau/\sigma}\right\} \\
T_{a+x}^{A} &:= \inf\left\{t \geq 0 : \left(a + x + L_t^0 - L_t^{\tau/\sigma}\right)^+ = 0\right\} \\
T_b^{B} &:= \inf\left\{t \geq 0 : \left(b + L_t^0 - L_t^{-\tau/\sigma}\right)^+ = 0\right\} \\
T_{a,b,x} &:= T_{a+x}^{A} \vee T_b^{B}.
\end{aligned}$$

that have analogous interpretations. Finally, the probability to be studied is

$$P\left[T_{a,x}^{A} < T_{a,b,x}\right].$$

4.4 Linking the models

In this section we show how Model-1 and Model-2 can be linked. In fact, we know the limiting behaviour of D and η and used it in the analysis of the continuous case. On the other hand, in the discrete scenario, we can approximate D using η: this would introduce the idea of a Poisson process in Model-1, which can become useful for future research.

For ϵ small enough we can write this approximation:

$$\frac{\epsilon}{\sigma} D_t(\epsilon/\sigma) \approx \sqrt{\frac{\pi\epsilon}{2\sigma}} \eta_t(\epsilon/\sigma) = \sqrt{\frac{\pi\epsilon}{2\sigma}} N^{\epsilon/\sigma}_{L^0_t}$$

and therefore

$$D_t(\epsilon/\sigma) \approx \sqrt{\frac{\pi\sigma}{2\epsilon}} N^{\epsilon/\sigma}_{L^0_t}.$$

This means that, for ϵ small enough, D can be approximated by a scaled Poisson process with intensity

$$\left(\frac{2\sigma}{\pi\epsilon}\right)^{1/2}$$

evaluated at the local time.

Similarly, the other quantities D^A and D^B can be approximated by the same scaled Poisson process (with the same intensity) evaluated at the local time at different levels

$$D^A_t(\epsilon/\sigma) \approx \sqrt{\frac{\pi\sigma}{2\epsilon}} N^{\epsilon/\sigma}_{L^{\tau/\sigma}_t}$$
$$D^B_t(\epsilon/\sigma) \approx \sqrt{\frac{\pi\sigma}{2\epsilon}} N^{\epsilon/\sigma}_{L^{-\tau/\sigma}_t}.$$

4.5 Conclusions

In this chapter we presented a more complex model compared to the previous ones, as it involves deep mathematical objects: downcrossings and excursions of a Brownian motion. Even if, given the scope of this work, we didn't have the resources to accurately study the quantities of interest, it is important to include the models of this chapter. In fact, they provide the future research with interesting tools to study partial and total executions using the theory of local times.

Chapter 5

Conclusions

This thesis aims at outlining a new approach to the study of market microstructure, based on a virtual price process that determines the orders' dynamics. Several models with different features and complexity levels are presented in order to compute some quantities of interest and to provide an additional point of view in the existing literature of the LOB modelling.

In Chapter 1 we presented the avalanche approach, where orders in the LOB accumulate on some levels and get executed when the virtual price process crosses such values. This framework involved the theory of the local time of a Brownian motion and allowed us to compute some relevant distributions.

In Chapter 2 we outline the main stochastic model for the LOB analysed in this thesis, using also a Poisson process for incoming orders and cancellations. This framework features a constant bid-ask spread, limit orders and cancellations driven by a Poisson process and a power law and a full book structure. We also managed to match the empirical density of the first passage time of the price process at a given barrier.

In Chapter 3, two more complex models are proposed using downcrossings and excursions of a Brownian motion. They provide useful tools to study partial and total executions in these frameworks.

The models described in this work can be studied more widely in many directions. Sell and buy orders should be allowed to arrive at different times. What happens after the first execution could be investigated, together with volatility and autocorrelations of prices. Moreover, the possibility of placing limit orders not only at the best quotes but also at different price levels should be considered.

In addition to the quantities already presented, the probability of making the spread could be studied, i.e. the probability of executing two orders, one at the ask and one at the bid, before the mid-price moves. Finally, also the long-term behaviour and the optimal strategies represent interesting areas for development.

Finally, it is important to stress once more that this work intends to be an overview of possible new ways to model the LOB with the tools provided by the stochastic analysis. Given the scope of this thesis, we could not study in depth all possible issues of this complex and articulated subject. However, this work could represent an interesting first step for a prolific future research in the modelling of market microstructure.

Appendices

Appendix A

The sojourn problem

A.1 Kac's method

In this section we present an overview of one of M. Kac's results in [17]. The calculation of distribution functions σ, defined as

$$\sigma(a,t) = P\left(\int_0^t V(B_\tau)\,d\tau < a\right) , \tag{A.1}$$

for $a \in [0,t]$ and a function $V : \mathbb{R} \to \mathbb{R}$, can be reduced to solving an appropriate differential equation. Let us restrict to the case

$$0 \leqslant V(x) \leqslant M \quad \forall x \in \mathbb{R}$$

for a positive constant M. For $n \in \mathbb{N}$, define the functions $Q_n : \mathbb{R} \times \mathbb{R}^+ \to \mathbb{R}^+$ as follows

$$Q_0(x,t) = \frac{1}{\sqrt{2\pi t}}\, e^{-\frac{x^2}{2t}} \tag{A.2}$$

$$Q_{n+1}(x,t) = \int_0^t \int_{-\infty}^{+\infty} \frac{e^{\frac{-(x-\xi)^2}{2(t-\tau)}}}{\sqrt{2\pi(t-\tau)}}\, V(\xi)\, Q_n(\xi,\tau)\, d\xi\, d\tau \, . \tag{A.3}$$

It can be checked that

$$\mu_n = E\left[\left(\int_0^t V(B_\tau)\,d\tau\right)^n\right] = n! \int_{-\infty}^{+\infty} Q_n(x,t)\,dx$$

and

$$0 \leqslant Q_n(x,t) \leqslant \frac{M^n}{n!}\, t^n Q_0(x,t) \, .$$

Let now

$$Q(x,t,u) = \sum_{n=0}^{+\infty} (-1)^n u^n Q_n(x,t) \tag{A.4}$$

for $u \in \mathbb{R}$. The series converges for all x and u in $\mathbb{R}$ and $t \neq 0$ and moreover

$$|Q(x,t,u)| < e^{uMt} Q_0(x,t) \ .$$

Due to (A.2) and (A.3), we have

$$Q(x,t,u) + \frac{u}{\sqrt{2\pi}} \int_0^t \int_{-\infty}^{+\infty} \frac{e^{\frac{-(x-\xi)^2}{2(t-\tau)}}}{\sqrt{t-\tau}} V(\xi)\, Q(\xi,\tau,u)\, d\xi\, d\tau = Q_0(x,t) \ . \tag{A.5}$$

It is clear that

$$E\left[e^{-u\int_0^t V(B_\tau)\,d\tau}\right] = \int_{-\infty}^{+\infty} Q(x,t,u)\, dx \ ,$$

and since V is positive it follows that $Q(x,t,u)$ is a decreasing function of u. In particular,

$$Q(x,t,u) \leqslant Q_0(x,t)$$

and thus the Laplace transform

$$\Psi(x,s,u) = \int_0^{+\infty} e^{-st} Q(x,t,u)\, dt \ , \tag{A.6}$$

where $s > 0$, exists.

If we take the Laplace transform on both sides of equation (A.5), we obtain

$$\Psi(x,s,u) + \frac{u}{\sqrt{2s}} \int_{-\infty}^{+\infty} e^{-\sqrt{2s}|x-\xi|}\, V(\xi)\, \Psi(\xi,s,u)\, d\xi = \frac{e^{-\sqrt{2s}|x|}}{\sqrt{2s}} \ . \tag{A.7}$$

The integral equation (A.7) is equivalent to the differential equation (where derivatives are taken with respect to x)

$$\frac{1}{2}\Psi'' - [s + uV(x)]\Psi = 0 \tag{A.8}$$

and the conditions

$$\Psi \to 0 \text{ as } x \to \pm\infty \tag{A.9}$$

$$\Psi' \text{ continuous except at } x = 0$$

$$\Psi'(0^-) - \Psi'(0^+) = 2 \ .$$

This is the procedure outlined by M. Kac for calculating distributions of type (A.1). Let us apply it to the case

$$V(x) = \mathbb{I}_{[0,+\infty]}(x) \ .$$

We obtain

$$\Psi(x) = \frac{\sqrt{2}}{\sqrt{s+u}+\sqrt{s}}\, e^{\sqrt{2s}\,x} \qquad x < 0 \tag{A.10}$$

$$\Psi(x) = \frac{\sqrt{2}}{\sqrt{s+u}+\sqrt{s}}\, e^{-\sqrt{2(s+u)}\,x} \qquad x \geqslant 0 \tag{A.11}$$

Thus,

$$\begin{aligned} \frac{1}{\sqrt{s(s+u)}} &= \int_{-\infty}^{+\infty} \Psi(x,s,u)\,dx = \int_{-\infty}^{+\infty}\int_0^{+\infty} e^{-st} Q(x,t,u)\,dt\,dx \\ &= \int_0^{+\infty} e^{-st} \int_0^{+\infty} e^{-ua}\, d_a\sigma(a,t)\,dt\ . \end{aligned}$$

Inverting with respect to s and u we obtain

$$\sigma(a,t) = \frac{2}{\pi}\ \arcsin\sqrt{\frac{a}{t}}\ .$$

The good part of this peculiar case is that we can invert analytically the double Laplace transform.

A.2 Application of Kac's method

Let us consider the function $V = \mathbb{I}_{[b,c]}$, where $0 < b < c < +\infty$, and use Kac's method in Appendix A to find the distribution σ. Solving the ODE (A.8) and applying condition (A.9), we get, after a quick study of the characteristic polynomial, (only the dependence on x is explicit)

$$\Psi(x) = \begin{cases} C_1\, e^{\alpha x} & \text{if } x < 0 \\ C_2\, e^{\alpha x} + C_3 e^{-\alpha x} & \text{if } x \in [0,b) \\ C_4\, e^{\beta x} + C_5 e^{-\beta x} & \text{if } x \in [b,c] \\ C_6\, e^{-\alpha x} & \text{if } x > c \end{cases} \tag{A.12}$$

where $\alpha = \sqrt{2s}$, $\beta = \sqrt{2(s+u)}$ and for some unknown constants C_i, $i = 1,\ldots,6$. In order to calculate these constants we should apply the conditions already outlined:

$$\Psi(0^-) = \Psi(0^+) \tag{A.13}$$

$$\Psi'(0^-) - \Psi'(0^+) = 2 \tag{A.14}$$

$$\lim_{x\to b^+} \Psi(x) = \lim_{x\to b^-} \Psi(x) \tag{A.15}$$

$$\lim_{x\to c^+} \Psi(x) = \lim_{x\to c^-} \Psi(x) \tag{A.16}$$

$$\lim_{x\to b^+} \Psi'(x) = \lim_{x\to b^-} \Psi'(x) \tag{A.17}$$

$$\lim_{x\to c^+} \Psi'(x) = \lim_{x\to c^-} \Psi'(x)\ . \tag{A.18}$$

In particular, note that condition (A.13) is imposed thanks to the continuity of Ψ in 0: this is ensured by its definition, i.e. by (A.2), (A.3), (A.4) and (A.6). Hence, we have six equations for six unknowns. Conditions (A.13) - (A.18) lead to the following system of equations for C_i's:

$$C_1 - C_2 - C_3 = 0 \tag{A.19}$$

$$C_1 - C_2 + C_3 = \frac{2}{\alpha} \tag{A.20}$$

$$C_2\, e^{\alpha b} + C_3\, e^{-\alpha b} - C_4\, e^{\beta b} - C_5\, e^{-\beta b} = 0 \tag{A.21}$$

$$C_2\, \alpha e^{\alpha b} - C_3\, \alpha e^{-\alpha b} - C_4\, \beta e^{\beta b} + C_5\, \beta e^{-\beta b} = 0 \tag{A.22}$$

$$C_4\, e^{\beta c} + C_5\, e^{-\beta c} - C_6\, e^{-\alpha c} = 0 \tag{A.23}$$

$$C_4\, \beta e^{\beta c} - C_5\, \beta e^{-\beta c} + C_6\, \alpha e^{-\alpha c} = 0\ . \tag{A.24}$$

Subtract (A.19) from (A.20) to obtain

$$C_3 = \frac{1}{\alpha} \tag{A.25}$$

$$C_1 - C_2 = \frac{1}{\alpha}. \tag{A.26}$$

Substitute (A.25) and (A.26) in (A.21):

$$C_1 = \frac{1}{\alpha}\,(1 - e^{-2\alpha b}) + e^{(\beta-\alpha)b} C_4 + e^{-(\alpha+\beta)b} C_5. \tag{A.27}$$

Hence:

$$C_2 = -\frac{1}{\alpha}\, e^{-2\alpha b} + e^{(\beta-\alpha)b} C_4 + e^{-(\alpha+\beta)b} C_5. \tag{A.28}$$

Now substitute (A.25) and (A.28) in (A.22). This gives

$$C_4 = \frac{2}{\alpha-\beta}\, e^{-(\alpha+\beta)b} - \frac{\alpha+\beta}{\alpha-\beta}\, e^{-2\beta b}\, C_5. \tag{A.29}$$

Replacing (A.29) in (A.23), we get

$$C_5 = \frac{e^{-\alpha c} C_6 - \frac{2}{\alpha-\beta} e^{-(\alpha+\beta)b+\beta c}}{e^{-\beta c} - \frac{\alpha+\beta}{\alpha-\beta}\, e^{-2\beta b+\beta c}} \tag{A.30}$$

Now insert (A.30) in (A.29) to find

$$C_4 = \frac{2}{\alpha-\beta}\, e^{-(\alpha+\beta)b} - \frac{e^{-\alpha c} C_6 - \frac{2}{\alpha-\beta} e^{-(\alpha+\beta)b+\beta c}}{\frac{\alpha-\beta}{\alpha+\beta} e^{2\beta b-\beta c} - e^{\beta c}}. \tag{A.31}$$

Finally, in order to find an explicit expression for C_6, substitute (A.30) and (A.31) in (A.24). After some easy calculations, we obtain

$$C_6 = \frac{4\beta}{\beta-\alpha}\, e^{-(\alpha+\beta)b+(\alpha+\beta)c} \frac{q(q^2-1)}{(q^2+2q+2)[(\alpha-\beta)q-(\alpha+\beta)]} \tag{A.32}$$

where

$$q = \frac{\alpha-\beta}{\alpha+\beta}\, e^{2\beta b-2\beta c}. \tag{A.33}$$

Formulas (A.25), (A.27), (A.28), (A.30), (A.31) and (A.32) together with (A.12) give us the solution Ψ of the integral equation (A.7) in the case $V = \mathbb{I}_{[b,c]}$, where $0 < b < c < +\infty$.

Let us check if these results are consistent with the arcsine law. In other words, we need to prove that if the interval $[b,c]$ covers $\mathbb{R}^+$, i.e. if $b \to 0^+$ and $c \to +\infty$, then the solution Ψ just found converges to (A.10) and (A.11). In terms of equations, we have to show that

$$\begin{aligned}
&\lim_{b\to 0^+, c\to\infty} C_1 = \frac{2}{\alpha+\beta} = \frac{\sqrt{2}}{\sqrt{s+u}+\sqrt{s}} \\
&\lim_{b\to 0^+, c\to\infty} C_4 = 0 \\
&\lim_{b\to 0^+, c\to\infty} C_5 = \frac{2}{\alpha+\beta} = \frac{\sqrt{2}}{\sqrt{s+u}+\sqrt{s}}.
\end{aligned}$$

Now, it is easy to check that

$$\lim_{b\to 0^+, c\to\infty} q = 0.$$

Therefore

$$\lim_{b\to 0^+, c\to\infty} C_6 = 0.$$

Using (A.30) we can also see that

$$\lim_{b\to 0^+, c\to\infty} C_5 = \frac{2}{\alpha+\beta}$$

It is then easy to show that

$$\lim_{b\to 0^+, c\to\infty} C_4 = 0 \qquad \lim_{b\to 0^+, c\to\infty} C_2 = -\frac{1}{\alpha} + \frac{2}{\alpha+\beta}$$

and finally

$$\lim_{b\to 0^+, c\to\infty} C_1 = \frac{2}{\alpha+\beta}.$$

Appendix B

Excursions of a Brownian motion

Consider the canonical version of a Brownian motion: denote by W the Wiener space, by P the Wiener measure and by $\mathcal{F}$ the Borel σ-field of W completed with respect to P. For $w \in W$, set

$$R(w) = \inf\{t > 0 : w(t) = 0\}$$

and call U the set of the functions $w \in W$ such that $0 < R(w) < \infty$ and $w(t) = 0$ for every $t \geq R(w)$. The point δ is the function identically equal to zero. Set $U_\delta = U \cup \{\delta\}$, $\mathcal{U}$ as the σ-algebra generated by the coordinate mappings and $\mathcal{U}_\delta = \mathcal{U} \vee \{\delta\}$. Define also

$$\tau_t(w) = \inf\{s > 0 : L_s > t\}$$

where L is the local time of w at level 0.

The excursion process is the process $e = (e_s, s > 0)$ defined on $(W, \mathcal{F}, P)$ with values in $(U_\delta, \mathcal{U}_\delta)$ by

1. if $\tau_s(w) - \tau_{s-}(w) > 0$, then $e_s(w)$ is the map

$$r \longmapsto \mathbb{I}_{[r \leq \tau_s(w) - \tau_{s-}(w)]}\ B_{\tau_{s-}(w)+r}(w);$$

2. if $\tau_s(w) - \tau_{s-}(w) = 0$, then $e_s(w) = \delta$.

It is well known (see [22]) that e is a σ-discrete point process. It is actually an $(\mathcal{F}_{\tau_t})$-Poisson point process (PPP). Moreover, $R(e_s(w))$ is a PPP on $\mathbb{R}^+$ with characteristic measure $\bar{n}$ given by $\bar{n}(]x, \infty[) = (2/\pi x)^{1/2}$.

Now, for a fixed $\epsilon > 0$, let us call $\eta_t(\epsilon)$ the number of excursions with duration bigger than ϵ (i.e. $R(e_s) \geq \epsilon$) which end at a time $s \leq t$. If N is the counting measure associated with the PPP $R(e_s)$, then it is easy to see that $\eta_t(\epsilon) = N^\epsilon_{L_t}$, where $N^\epsilon_u = N^{[\epsilon,\infty[}_u$.

Bibliography

[1] Bak, P.; Baczuski, M. and Shubik, M. (1997). *Price Variations in a Stock Market with Many Agents.* Physica A **246**, 430.

[2] Bollerslev, T.; Domowitz, I. and Wang, J. (1997). *Order Flow and the Bid-Ask Spread: An Empirical Probability Model of Screen-Based Trading.* J. Econ. Dynamics Control **21**, 1471.

[3] Borodin, A.N. and Salminen, P. (2002). *Handbook of Brownian Motion - Facts and Formulae. Second Edition.* Birkhäuser, Basel.

[4] Bouchaud, J.P.; Mezard, M. and Potters, M. (June 2002) *Statistical Properties of Stock Order Books: Empirical Results and Models.* Science and Finance. Capital Fund Management.

[5] Bouchaud, J.P. and Potters, M. (2003) *More Statistical Properties of Order Books and Price Impact.* Physica A, **324**, 133.

[6] Challet, D. and Stinchcombe, R. (2001). *Analyzing and Modelling 1+1d Markets.* Physica A **300**, 285.

[7] Cohen, K.J.; Conroy, R.M. and Maier, S.F. (1985). *Order Flow and the Quality of the Market: Market Making and the Changing Structure of the Securities Industry.* Lexington Books, Lexington, MA.

[8] Cont, R. and De Lallard, A. (2012) *Price Dynamics in a Markovian Limit Order Market* . Working paper series. Available at SSRN: http://ssrn.com/abstract=1735338.

[9] Cont, R.; Kukanov, A. and Stoikov, S. (2012) *The Price Impact of Order Book Events.* Journal of Financial Econometrics (2013). Available at SSRN: http://ssrn.com/abstract=1712822.

[10] Cont, R.; Stoikov, S. and Talreja, R. (2008) *A Stochastic Model for Order Book Dynamics.* Operations Research, Volume 58, 549-563.

[11] Davis, M. and Pistorius, M. (2010) *Quantification of Counterparty Risk Via Bessel Bridges*. Working paper series. Available at SSRN: http://ssrn.com/abstract=1722604.

[12] Delattre, S.; Robert, C.Y. and Rosenbaum, M. (2013) *Estimating the Efficient Price from the Order Flow: a Brownian Cox Process Approach*. Stochastic Processes and Their Applications, **123** (7), 2603-2619.

[13] Domowitz, I. and Wang, J. (1994). *Auctions as Algorithms*. J. Econ. Dynamics Control **18** , 29.

[14] Eisler, Z.; Kertész, J.; Lillo, F. and Mantegna, R. (2008) *Diffusive Behaviour and the Modeling of Characteristic Times in Limit Order Executions*. Quantitative Finance, Volume 9, Issue 5, 547-563.

[15] Eliezer, D. and Kogan, I. (1998). *Scaling Laws for the Market Microstructure of the Interdealer Broker Markets*. Preprint cond.mat/9808240.

[16] Fusai, G. (2000). *Corridor Options and Arc-Sine Law*. Annals of Applied Probability, **10**(2), May 2000, 634-63.

[17] Kac, M. (1951). *On Some Connections Between Probability Theory and Differential and Integral Equations*. Proc. Second Berkeley Symp. on Math. Statist. and Prob. (Univ. of Calif. Press), 189-215.

[18] Karatzas, I. and Shreve, S.E. (1998). *Brownian Motion and Stochastic Calculus. Second Edition.* Springer, New York.

[19] Lorenz, J. and Osterrieder, J. (2009). *Simulation of a Limit Order Driven Market*. Journal of Trading, **4**(1), 23-30.

[20] Maslov, S. (2000). *Simple Model of a Limit Order Driven Market*. Physica A **278**, 571.

[21] Osterrieder, J. (2007). *Arbitrage, the Limit Order Book and Market Microstructure Aspects in Financial Market Models*. Ph.D. Thesis, ETH Zurich. DISS. ETH Nr.17121.

[22] Revuz, D. and Yor, M. (1999). *Continuous Martingales and Brownian Motion. Third Edition.* Springer, Berlin.

[23] Rosu, I. (2008). *A Dynamic Model of the Limit Order Book*. Review of Financial Studies, Vol. 22, pp. 4601-4641, 2009.

[24] Tang, L.H. and Tian, G.S. (1999). *Reaction-Diffusion-Branching Models of Stock Price Fluctuations*. Physica A **264**, 543.