HAL
open science

Stochastic order book modelling

Aymen Jedidi

To cite this version:

Aymen Jedidi. Stochastic order book modelling. Other. Ecole Centrale Paris, 2014. English. NNT : 2014ECAP0001 . tel-00997433

HAL Id: tel-00997433

https://theses.hal.science/tel-00997433

Submitted on 28 May 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Modélisation Stochastique des Carnets d'Ordres

Stochastic Order Book Modelling

Thèse présentée par

Aymen Jedidi

pour l'obtention du

Grade de Docteur

Spécialité : Mathématiques Appliquées

Laboratoire d'accueil : MAS

Soutenue le, 09 janvier 2014

devant un jury composé de

Frédéric Abergel (École Centrale Paris), *Directeur de thèse*

Bernard Lapeyre (École des Ponts ParisTech), *Rapporteur*

Mathieu Rosenbaum (Université Pierre-et-Marie-Curie), *Rapporteur*

Emmanuel Bacry (École Polytechnique), *Examinateur*

Jean-Philippe Bouchaud (Capital Fund Management), *Examinateur*

Jim Gatheral (Baruch College), *Examinateur*

Fabrizio Lillo (Scuola Normale Superiore di Pisa), *Professeur invité*

2014ECAP0001

À ma famille

Remerciements

Mes plus vifs remerciements vont à Frédéric Abergel, pour avoir accepté de diriger cette thèse et pour sa générosité constante, sa patience et sa confiance. Qu'il voit dans ce travail l'expression de ma gratitude.

Je tiens aussi à remercier Bernard Lapeyre et Mathieu Rosenbaum pour avoir accepté de relire le manuscrit, et Emmanuel Bacry, Jean-Philippe Bouchaud, Jim Gatheral et Fabrizio Lillo pour avoir fait partie du jury de cette thèse. Leur présence m'honore et m'oblige.

De la convivialité dans FiQuant, à la gentillesse et disponibilité au secrétariat, et la qualité des échanges scientifiques, la réalisation de cette thèse tient aussi aux conditions de travail excellentes offertes par le laboratoire MAS. J'en remercie chaleureusement toute l'équipe, notamment Annie, Marouane, Mehdi, Nicolas, Riadh et Sylvie.

Mon éternelle reconnaissance va à mes parents et mes deux sœurs pour leur soutien sans faille.

Abstract

Résumé

Cette thèse étudie quelques aspects de la modélisation stochastique des carnets d'ordres. Nous analysons dans la première partie un modèle dans lequel les temps d'arrivées des ordres sont Poissoniens indépendants. Nous démontrons que le carnet d'ordres est stable (au sens des chaines de Markov) et qu'il converge vers sa distribution stationnaire exponentiellement vite. Nous en déduisons que le prix engendré dans ce cadre converge vers un mouvement Brownien aux grandes échelles de temps. Nous illustrons les résultats numériquement et les comparons aux données de marché en soulignant les succès du modèle et ses limites. Dans une deuxième partie, nous généralisons les résultats à un cadre où les temps d'arrivés sont régis par des processus auto et mutuellement existants, moyennant des hypothèses sur la mémoire de ces processus. La dernière partie est plus appliquée et traite de l'identification d'un modèle réaliste multivarié à partir des flux des ordres. Nous détaillons deux approches : la première par maximisation de la vraisemblance et la seconde à partir de la densité de covariance, et réussissons à avoir une concordance remarquable avec les données. Nous appliquons le modèle ainsi estimé à deux problèmes concrets de *trading* algorithmique, à savoir la mesure de la probabilité d'exécution et le coût d'un ordre limite.

Abstract

This thesis presents some aspects of stochastic order book modelling. In the first part, we analyze a model in which order arrivals are independent Poisson. We show that the order book is stable (in the sense of Markov chains) and that it converges to its stationary state exponentially fast. We deduce that the price generated in this setting converges to a Brownian motion at large time scales. We illustrate the results numerically and compare

them to market data. In the second part, we generalize the results to a setting in which arrival times are governed by self and mutually existing processes. The last part is more applied and deals with the identification of a realistic multivariate model from the order flow. We describe two approaches: the first based on maximum likelihood estimation and the second on the covariance density function, and obtain a remarkable agreement with the data. We apply the estimated model to two specific algorithmic trading problems, namely the measurement of the execution probability of a limit order and its cost.

Contents

Remerciements 5

Abstract 7
Résumé 7
Abstract 7

Chapter 1. Introduction et Principaux Résultats 13
1. Contexte et Motivations 13
2. Contributions 15

Chapter 2. Markovian Order Book Models I: Stability and Scaling Limits 17
1. Introduction and Background 17
1.1. Related Literature 18
1.2. Outline 20
2. An Elementary Approximation: Perfect Market Making 21
3. Order Book Dynamics 23
3.1. Model Setup: Poissonian Arrivals, Reference Frame and Boundary Conditions 23
3.2. Comparison to Previous Results and Models 24
3.3. Evolution of the Order Book 26
4. Infinitesimal Generator 28
5. Price Dynamics 29
6. Ergodicty and Diffusive Limit 30
6.1. Ergodicity of the Order Book and Rate of Convergence to the Stationary State 32
6.2. The Embedded Markov Chain 33
6.3. The Case of Constant Cancellation Rates 35
6.4. Large-scale Limit of the Price Process 37

Chapter 3. Markovian Order Book Models II: Numerical Example 41
1. Numerical Example 41

1.1. Anomalous Diffusion at Short Time Scales. 44
2. Model Parameters Estimation 50
2.1. Description of the Data 50
2.2. Trades and Tick by Tick Data Processing 50
2.3. Parameters Estimation 53
2.4. A Typical Parmaters Set 54
3. Results for CAC 40 Stocks 54
4. Conclusions 58

Chapter 4. On the Stability and Price Scaling Limit of a Hawkes Process-Based Order Book Model 63
1. Introduction 63
1.1. Outline 63
1.2. Notations 64
2. Preliminary Remarks 64
2.1. Point Processes 64
2.2. Hawkes Processes 65
2.3. The Embedded Discrete-time Hawkes Process 67
2.4. Drift of a Discrte-time Markov Process 68
2.5. A Digression on Stochastic Stability 69
3. Auxiliary Results 70
3.1. $V-$uniform Ergodicity of the Intensity of a Hawkes Process 70
3.2. $V-$uniform Ergodicity of the Intensity of a Multivariate Hawkes Process 72
3.3. V-uniform Ergodicity of a “Birth-death” Hawkes Process 75
4. Application in Order Book Modelling 77
4.1. Model Setup 77
4.2. Stability of the Order Book 78
4.3. Large-scale Limit of the Price Process 80
5. Technical Lemmas 82

Chapter 5. Numerical Results and Applications 85
1. Introduction 85
1.1. Notations 85
1.2. Classic Results 85
2. Maximum Likelihood Estimation of Multivariate Hawkes Processes 86
2.1. Likelihood of Univariate Hawkes Processes 86
2.2. Likelihood of Multivariate Hawkes Processes 88

2.3. Illustration 91
3. Goodness of Fit Assessment 91
3.1. A Pragmatic Assessment 91
3.2. A Theoretically Sounder Assessment: the Time-rescaling Theorem 93
3.3. An Even Sharper Assessment 93
4. Application to Tick By Tick Order Book Data 94
5. Non-parametric Estimation of Multivariate Hawkes Processes 95
5.1. The Covariance Density Matrix 95
5.2. The Hawkes-Wiener-Hopf Integral Equation 99
5.3. Nyström's Method 103
5.4. Empirical Covariance Density 105
5.5. Estimating the Base Rate $\boldsymbol{\mu}$ 105
5.6. Principle of the Non-parametric Kernel Estimation Method 106
5.7. Illustration 106
6. Simulation *via* Markovian Projection 106
6.1. Ogata's Thinning Simulation Method 106
6.2. General Kernel 110
7. Applications 112
7.1. Probability of Fill of a Limit Order 112
7.2. Execution Shortfall of a Limit Order 115

Bibliography 119

CHAPTER 1

Introduction et Principaux Résultats

1. Contexte et Motivations

Les économies modernes doivent une partie de leur richesse au bon fonctionnement de leurs marchés financiers. L'industrie financière a vécu trois changements majeurs depuis les années 1930 et il est pertinent de les rappeler ici [**KL13**].

Le premier a trait à la complexité accrue du système financier lui-même. Cela est un corollaire de l'évolution économique et démographique ainsi que la mondialisation des échanges.

Le deuxième développement est l'émergence de la finance quantitative. Les progrès en économie financière ont été nombreux, pour n'en citer que quelques-uns : la théorie d'optimisation du portefeuille de Markowitz, le modèle d'évaluation des actifs financiers (CAPM) de Sharpe, le modèle multi-facteurs du risque (BARRA) de Rosenberg, la formule de Black, Scholes et Merton pour la valorisation des options et son principe de couverture dynamique, ainsi que la théorie de valorisation par martingales de Harrison et Pliska. Ces travaux et d'autres ont fourni le fondement sur lequel repose un pan de l'industrie financière.

Le troisième bouleversement vient des progrès en technologie informatique, fût-ce au niveau matériel (figure 1), logiciel, collection et stockage des données, ou encore la connectivité et les réseaux. La puissance de calcul a rendu possible la résolution de problèmes auparavant insolubles—le problème d'optimisation du portefeuille en est un des premiers exemples. Les avancées logicielles ont diminué les couts des opérations bancaires, et la connectivité rendu possible l' "électronification" des marchés.

Une autre conséquence de la révolution informatique est qu'elle permet l'enregistrement de toutes les transactions financières, voire tous les ordres passés, offrant au scientifique une quantité astronomique de données et rendant possible une approche quasi-expérimentale des marchés.

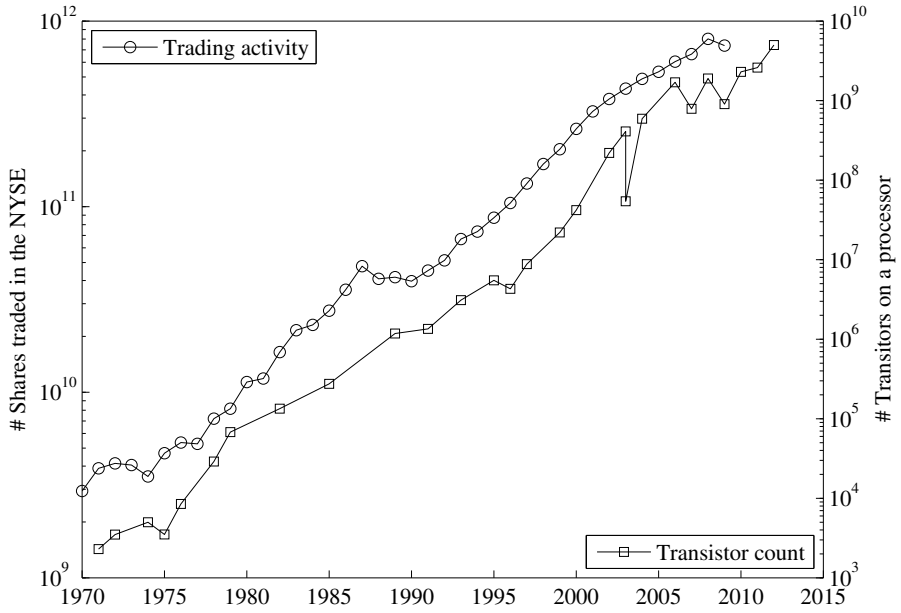


FIGURE 1. Comparative evolution of trading activity and computing power [**Eco11**]. y-axis is in logarithmic scale. Data sources: NYSE-Euronext / Wikipedia.

Au centre des transactions électroniques se trouve le carnet d'ordres. C'est là où l'offre et la demande se rencontrent. Le carnet d'ordres répertorie à chaque instant tous les ordres d'achat ou de vente non-exécutés sur un titre, avec priorité selon le prix et le temps de soumission. Cet objet, dont les règles de fonctionnement sont assez simples, est, regardé de près, très complexe. Le but de cette thèse est d'explorer cet objet sous un angle mathématique.

Plus précisément, notre but est triple:

(1) Proposer quelques fondements théoriques sous des hypothèses simples.
(2) Étudier quelques propriétés empiriques des carnets d'ordres, sur des données dites *tick by tick*.
(3) Montrer le potentiel applicatif de certain modèles de carnet basés sur une classe de processus ponctuel, les processus de Hawkes, en en soulignant les succès et les limites.

2. Contributions

Nous résumons dans cette section les points-clés de cette thèse, chapitre par chapitre.

Chapitre 2. Le but de ce chapitre est d'explorer les liens entre la description microscopique de la formation des prix (modélisation multi-agents) et l'approche par équations différentielles stochastiques utilisée classiquement pour décrire l'évolution des prix à des échelles de temps macroscopiques.

Nous présentons une étude mathématique du carnet d'ordres comme une chaîne de Markov multidimensionnelle en temps continu et prouvons plusieurs résultats dans le cas de temps d'arrivée poissonniens indépendants.

Nous montrons que la structure des annulations est un facteur important pour l'existence d'une distribution stationnaire pour le carnet d'ordres et la convergence exponentielle envers elle. Nous démontrons aussi, par l'intermédiaire du théorème central limite fonctionnel, que la limite à grande échelle du processus de prix est un mouvement Brownien.

Chapitre 3. Nous illustrons l'analyse théorique du deuxième chapitre par simulation numérique et comparons les résultats aux données de marché. C'est aussi l'occasion pour nous de décrire les données *tick by tick* et leur traitement.

Chapitre 4. Depuis leur introduction dans [**Haw71b**], les processus de Hawkes ont été appliqués dans un large éventail de domaines de recherche allant de de la sismologie (à l'origine), au risque de crédit, la contagion financière, et plus récemment la modélisation de la microstructure des marchés. En microstructure des marchés, et en particulier la modélisation de carnet d'ordres, la pertinence de ces processus provient au moins de deux propriétés empiriques du flux des ordres à l'échelle microscopique:

(1) *Time-clustering*: les ordres ont tendance à arriver par grappes.
(2) Dépendance mutuelle: les flux d'ordres présentent des dépendances croisées non négligeables. Par exemple, les ordres au marché "excitent" les ordres limites et *vice versa*.

Dans ce chapitre, nous posons un modèle de carnet d'ordres basé sur le processus de Hawkes dans un cadre markovien, et en utilisant des techniques de la théorie des chaînes de Markov et la stabilité stochastique, montrons que le carnet d'ordres est stable et conduit à un prix diffusif limite à

de grandes échelles de temps, généralisant ainsi les résultats du deuxième chapitre.

Chapitre 5. Ce chapitre est axé sur les aspects pratiques de la modélisation stochastique des carnets d'ordres, à savoir l'estimation d'un modèle réaliste de carnet basé sur les processus de Hawkes et son utilisation pour l'évaluation de stratégies de *trading* algorithmique.

Nous détaillons deux approches d'estimations : la première par maximisation de la vraisemblance et la seconde à partir de la densité de covariance, et réussissons à obtenir une concordance remarquable avec les données. Nous appliquons le modèle ainsi estimé à deux problèmes concrets de *trading* algorithmique, à savoir la mesure de la probabilité d'exécution et le coût d'un ordre limite.

CHAPTER 2

Markovian Order Book Models I: Stability and Scaling Limits

The aim of this chapter is to explore the links between the microscopic description of price formation (agent-based modeling) and the stochastic differential equations approach used classically to describe price evolution at macroscopic time scales. We present a mathematical study of the order book as a multidimensional continuous-time Markov chain and derive several mathematical results in the case of independent Poissonian arrival times. In particular, we show that the cancellation structure is an important factor ensuring the existence of a stationary distribution for the order book and the exponential convergence towards it. We also prove, by means of the functional central limit theorem (FCLT), that the large-scale limit of the price process is a Brownian motion. We illustrate the analysis with numerical simulation and comparison against market data.

1. Introduction and Background

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

- *Limit order*: Specify a price (also called "quote") at which one is willing to buy or sell a certain number of shares;
- *Market order*: Immediately buy or sell a certain number of shares at the best available opposite quote;
- *Cancellation order*: Cancel an existing limit order.

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

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

$$S := P^A - P^B, \tag{1.1}$$

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

$$P := \frac{P^A + P^B}{2}. \tag{1.2}$$

The price dynamics is the result of the interplay between the incoming order flow and the order book [**BMP02**]. Figure 1 is a schematic illustration of this process [**Fer08**]. Note that we chose to represent quantities on the bid side of the book by non-positive numbers.

Although in reality orders can have any size, we shall assume in most of the chapter that all orders have a fixed unit size q. This assumption is convenient to carry out our analysis and is, for now, of secondary importance to the problem we are interested in[1]. Throughout this chapter, we may refer to three different "times":

- *Physical time* (or *clock time*) in seconds,
- *Event time* (or *tick time*): the time counter is incremented by 1 every time an event (i.e. market, limit or cancellation order) occurs,
- *Trade time* (or *transaction time*): the time counter is incremented every trade (i.e. every market order).

1.1. Related Literature. Order book modelling has been an area of intense research activity in the last decade. The remarkable interest in this area is due to two factors:

[1]It will be relaxed in chapter 3 where we resort to numerical simulation.

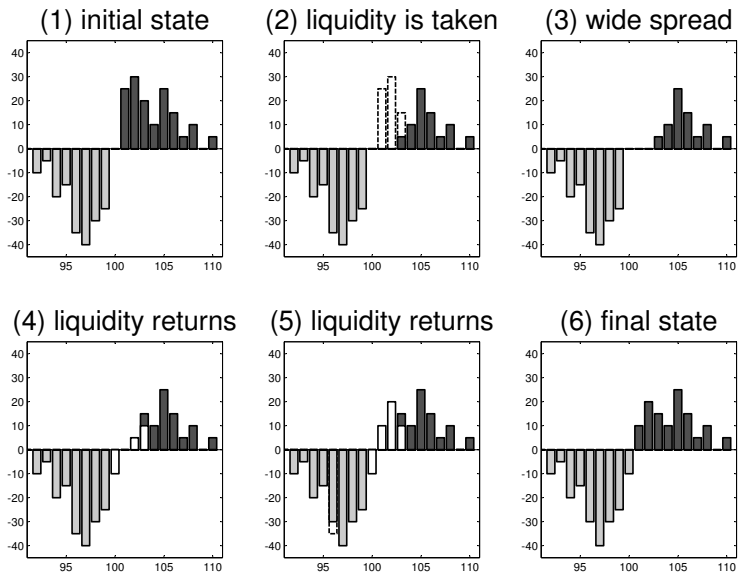


FIGURE 1. Order book schematic illustration: a buy market order arrives and removes liquidity from the ask side, then sell limit orders are submitted and liquidity is restored.

- Widespread use of algorithmic trading in which the order book is the place where offer and demand meet,
- Availability of tick by tick data that record every change in the order book and allow precise analysis of the price formation process at the microscopic level.

Schematically, two modelling approaches have been successful in capturing key properties of the order book—at least partially. The first approach, led by economists, models the interactions between rational agents who act strategically: they choose their trading decisions as solutions to individual utility maximization problems (See e.g. [**PS08**] and references therein).

In the second approach, proposed by econophysicists, agents are assumed to act randomly. This is sometime referred to as zero-intelligence order book modelling, in the sense that order arrivals and placements are entirely stochastic. The focus here is more on the "mechanistic" aspects

of the continuous double auction rather than the strategic interactions between agents. Despite this apparent limitation, zero-intelligence (or statistical) order book models do capture many salient features of real markets (See [**DFG+03, FPZ05**] and references therein). Two notable developments in this strand of research are [**Mas00**] who proposed one of the earliest stochastic order book models, and [**CS01**] who added the possibility to cancel existing limit orders.

In their seminal paper [**SFGK03**], Smith et al. develop a dynamical statistical order book model under the assumption of independent Poissonian order flows. They provide a thorough analysis of the model using simulation, dimensional analysis and mean field approximation. They study key characteristics of the model, namely:

(1) Price diffusion.
(2) Liquidity characteristics: average depth profile, bid-ask spread, price impact and time and probability to fill a limit order.

One of the most important messages of their analysis is that zero-intelligence order book models are able to produce reasonable market dynamics and liquidity characteristics. Our focus here is on the first point, that is, *the convergence of the price process, which is a jump process at the microscopic level, to a diffusive process*[2] *at macroscopic time scales*. The authors in [**SFGK03**] suggest that a diffusive regime is reached. Their argument relies on a mean field approximation. Essentially, this amounts to neglecting the dependence between order fluctuations at adjacent price levels.

Another important paper of interest to us is [**CST10**]. Cont et al. propose to model the order book dynamics from the vantage point of queuing systems. They remarkably succeed in deriving many conditional probabilities of practical importance such as the probability of an increase in the mid-price, of the execution of an order at the bid before the ask quote moves, and of "making the spread". To our knowledge, they are the first to clearly set the problem of stochastic order book modelling in the context of Markov chains, which is a very powerful and well-studied mathematical concept.

1.2. Outline. In this chapter, we build on the models of [**CST10**] and [**SFGK03**] to present a stylized description of the order book, and derive

[2]We mean (abusively) by "diffusive process" or simply "diffusion" the mathematical concept of Brownian motion.

several mathematical results in the case of independent Poissonian arrival times. In particular, we show that the cancellation structure is an important factor ensuring the existence of a stationary distribution for the order book and the exponential convergence towards it. We also prove, by means of the functional central limit theorem (FCLT), that the rescaled-centered price process converges to a Brownian motion, which is a new result.

The remainder of the chapter is organized as follows. In section 2, we motivate our approach using an elementary example where the spread is kept constant ("perfect market making"). In sections 3 trough 5, we compute the infinitesimal generator associated with the order book in a general setting, and link the price dynamics to the instantaneous state of the order book. In section 6, we prove that the order book is *ergodic*—in particular it has a *stationary distribution*—that it converges to its stationary state *exponentially fast*, and that the large-scale limit of the price process is a *Brownian motion*. Our proofs rely on the theory of infinitesimal generators and Foster-Lyapunov stability criteria for Markov chains. We outline an order book simulation algorithm in section 1 and provide a numerical illustration. Finally, section 4 summarizes our results and contains critiques of Markovian order book models.

2. An Elementary Approximation: Perfect Market Making

We start with the simplest agent-based market model:

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

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

$$dP(t) = \Delta P\ (dM^+(t) - dM^-(t))Z, \tag{2.1}$$

where Z is a Bernoulli random variable

$$Z = 0 \text{ with probability } (1-u), \tag{2.2}$$

and

$$Z = 1 \text{ with probability } u. \tag{2.3}$$

The infinitesimal generator[3] $\mathcal{L}$ associated with this dynamics is

$$\mathcal{L}f(P) = u\ \left[\lambda^+\ (f(P+\Delta P) - f) + \lambda^-\ (f(P-\Delta P) - f)\right], \tag{2.5}$$

where f denotes a test function. It is well known that a continuous limit is obtained under suitable assumptions on the intensity and tick size. Noting that (2.5) can be rewritten as

$$\begin{aligned}\mathcal{L}f(P) &= \frac{1}{2}u\ (\lambda^+ + \lambda^-)(\Delta P)^2 \frac{f(P+\Delta P) - 2f + f(P-\Delta P)}{(\Delta P)^2} \\ &+ u\ (\lambda^+ - \lambda^-)\Delta P \frac{f(P+\Delta P) - f(P-\Delta P)}{2\Delta P},\end{aligned} \tag{2.6}$$

and under the following assumptions

$$u\ (\lambda^+ + \lambda^-)(\Delta P)^2 \longrightarrow \sigma^2 \ \text{ as } \Delta P \to 0, \tag{2.7}$$

and

$$u\ (\lambda^+ - \lambda^-)\Delta P \longrightarrow \mu \ \text{ as } \Delta P \to 0, \tag{2.8}$$

the generator converges to the classical diffusion operator

$$\frac{\sigma^2}{2}\frac{\partial^2 f}{\partial P^2} + \mu \frac{\partial f}{\partial P}, \tag{2.9}$$

corresponding to a Brownian motion with drift. This simple case is worked out as an example of the type of limit theorems that we will be interested in in the sequel. One should also note that a more classical approach using the Functional Central limit Theorem (FCLT) as in [**Bil99**] or [**Whi02**] yields similar results ; For given fixed values of λ^+, λ^- and ΔP, the rescaled-centred price process

$$\frac{P(nt) - n\mu t}{\sqrt{n}\sigma} \tag{2.10}$$

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

$$\mathcal{L}f(\mathbf{x}) := \lim_{t\downarrow 0} \frac{\mathbb{E}[f(\mathbf{X}(t))|\mathbf{X}(0) = \mathbf{x}] - f(\mathbf{x})}{t}. \tag{2.4}$$

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

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

$$\sigma = \Delta P \sqrt{(\lambda^+ + \lambda^-)u}, \tag{2.11}$$

and

$$\mu = \Delta P(\lambda^+ - \lambda^-)u. \tag{2.12}$$

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

$$u\ (\lambda^+ + \lambda^-)(\Delta P)^2 \to \sigma^2(P, t), \tag{2.13}$$

and

$$u\ (\lambda^+ - \lambda^-)\Delta P \to \mu(P, t). \tag{2.14}$$

This is the case if the original intensities are functions of P and t themselves.

3. Order Book Dynamics

3.1. Model Setup: Poissonian Arrivals, Reference Frame and Boundary Conditions. We now consider the dynamics of a general order book under the assumption of Poissonian arrival times for market orders, limit orders and cancellations. We shall assume that each side of the order book is fully described by a *finite* number of limits K, ranging from 1 to K ticks away from the best available opposite quote. We will use the notation[4]

$$\mathbf{X}(t) := (\mathbf{a}(t); \mathbf{b}(t)) := (a_1(t), \ldots, a_K(t); b_1(t), \ldots, b_K(t)), \tag{3.1}$$

where $\mathbf{a} := (a_1, \ldots, a_K)$ designates the ask side of the order book and a_i the number of shares available i ticks away from the best opposite quote, and $\mathbf{b} := (b_1, \ldots, b_K)$ designates the bid side of the book. By doing so, we adopt the representation described in [**CST10**] or [**SFGK03**][5], but depart slightly from it by adopting a *finite moving frame*, as we think it is realistic and more convenient when scaling in tick size will be addressed.

Let us now recall the events that may happen:

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

These events are described by *independent* Poisson processes:

[4]In what follows, bold notation indicates vector quantities.

[5]See also [**GO10**] for an interesting discussion.

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

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

We impose constant boundary conditions outside the moving frame of size $2K$: Every time the moving frame leaves a price level, the number of shares at that level is set to a_∞ (or b_∞ depending on the side of the book). Our choice of a finite moving frame and constant[6] boundary conditions has three motivations. Firstly, it assures that the order book does not empty and that P^A, P^B are always well defined. Secondly, it keeps the spread S and the increments of P^A, P^B and $P = (P^A + P^B)/2$ bounded—This will be important when addressing the scaling limit of the price. Thirdly, it makes the model Markovian as we do not keep track of the price levels that have been visited (then left) by the moving frame at some prior time. Figure 1 is a representation of the order book using the above notations.

3.2. Comparison to Previous Results and Models. Before we proceed, we would like to recall some results already present in the literature and highlight their differences with respect to our analysis. Smith et al. have already investigated in [**SFGK03**] the scaling properties of some liquidity and price characteristics in a stochastic order book model. These results are summarized in table 1. In the model of Smith et al. [**SFGK03**], orders arrive on an *infinite* price grid (This is consistent as limit orders arrival rate *per price level* is finite). Moreover, the arrival rates are independent of the price level, which has the advantage of enabling the analytical predictions summarized in table 1.

[6]Actually, taking for a_∞ and $|b_\infty|$ independent positive random variables would not change much our analysis. We take constants for simplicity.

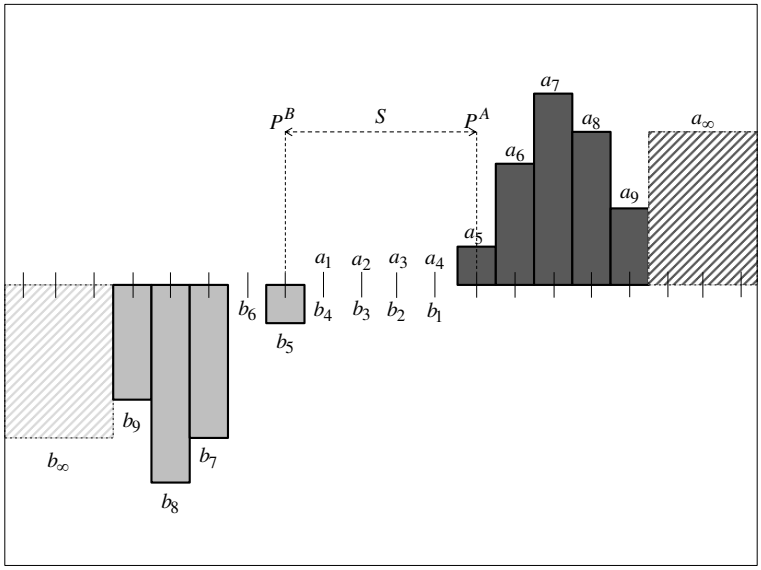


FIGURE 2. Order book dynamics: in this example, $K = 9$, $q = 1$, $a_\infty = 4$, $b_\infty = -4$. The shape of the order book is such that $\mathbf{a}(t) = (0, 0, 0, 0, 1, 3, 5, 4, 2)$ and $\mathbf{b}(t) = (0, 0, 0, 0, -1, 0, -4, -5, -3)$. The spread $S(t) = 5$ ticks. Assume that at time $t' > t$ a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

Quantity	Scaling relation
Average asymptotic depth	λ^L/λ^C
Average spread	$\lambda^M/\lambda^L f(\epsilon, \Delta P/p_c)$
Slope of average depth profile	$(\lambda^L)^2/\lambda^M\lambda^C g(\epsilon, \Delta P/p_c)$
Price "diffusion" parameter at short time scales	$(\lambda^M)^2\lambda^C/\lambda^L\epsilon^{-0.5}$
Price "diffusion" parameter at long time scales	$(\lambda^M)^2\lambda^C/\lambda^L\epsilon^{0.5}$

TABLE 1. Results of Smith et al. $\epsilon := q/(\lambda^M/2\lambda^C)$ is a "granularity" parameter that characterizes the effect of discreteness in order sizes, $p_c := \lambda^M/2\lambda^L$ is a characteristic price interval, and f and g are slowly varying functions.

We stress that, to our understanding, these results are obtained by mean-field approximations, which assume that the fluctuations at adjacent price levels are independent. This allows fruitful simplifications of the complex dynamics of the order book. In addition, the authors do not characterize the convergence of the coarse-grained price process in the sense of Stochastic Process Limits, nor do they show that the limiting process is precisely a Brownian motion (theorem 4.2).

In the model of Cont el al. [**CST10**], arrival rates are indexed by the distance to the best opposite quote, which is more realistic. The order book is constrained to a finite price grid $[1, P_{max}]$ that facilitates the analysis of the Markov chain. Here, we use a combination of the two models in that the arrival rates are not uniformly distributed across prices, and the reference frame is finite but moving. Cont et al. [**CST10**] have considered the question of the ergodicity of their order book model. We also address this question following a different route, and more importantly to our analysis, exhibit the *rate of convergence* to the stationary state, which turns out to be the key of the proof of theorem 4.2.

3.3. Evolution of the Order Book. We can write the following coupled SDEs for the quantities of outstanding limit orders in each side of the order book:[7]

$$\begin{aligned} da_i(t) &= -\left(q - \sum_{k=1}^{i-1} a_k\right)_+ dM^+(t) + q dL_i^+(t) - q dC_i^+(t) \\ &+ (J^{M^-}(\mathbf{a}) - \mathbf{a})_i dM^-(t) + \sum_{i=1}^{K} (J^{L_i^-}(\mathbf{a}) - \mathbf{a})_i dL_i^-(t) \\ &+ \sum_{i=1}^{K} (J^{C_i^-}(\mathbf{a}) - \mathbf{a})_i dC_i^-(t), \end{aligned} \tag{3.2}$$

[7]Remember that, by convention, the b_i's are non-positive.

and

$$\begin{aligned} db_i(t) &= \left(q - \sum_{k=1}^{i-1} |b_k|\right)_+ dM^-(t) - q dL_i^-(t) + q dC_i^-(t) \\ &+ (J^{M^+}(\mathbf{b}) - \mathbf{b})_i dM^+(t) + \sum_{i=1}^{K} (J^{L_i^+}(\mathbf{b}) - \mathbf{b})_i dL_i^+(t) \\ &+ \sum_{i=1}^{K} (J^{C_i^+}(\mathbf{b}) - \mathbf{b})_i dC_i^+(t), \end{aligned} \tag{3.3}$$

where the J's are *shift operators* corresponding to the renumbering of the ask side following an event affecting the bid side of the book and vice versa. For instance the shift operator corresponding to the arrival of a sell market order $dM^-(t)$ of size q is[8]

$$J^{M^-}(\mathbf{a}) = \left(\underbrace{0, 0, \ldots, 0}_{\text{k times}}, a_1, a_2, \ldots, a_{K-k}\right), \tag{3.4}$$

with

$$k := \inf\{p : \sum_{j=1}^{p} |b_j| > q\} - \inf\{p : |b_p| > 0\}. \tag{3.5}$$

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

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

[8]For notational simplicity, we write $J^{M^-}(\mathbf{a})$ instead of $J^{M^-}(\mathbf{a}; \mathbf{b})$ etc. for the shift operators.

4. Infinitesimal Generator

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

$$\begin{aligned}
\mathcal{L}f(\mathbf{a};\mathbf{b}) = {} & \lambda^{M^+}(f\left([a_i-(q-A(i-1))_+]_+; J^{M^+}(\mathbf{b})\right)-f) \\
& + \sum_{i=1}^{K} \lambda_i^{L^+}(f\left(a_i+q; J^{L_i^+}(\mathbf{b})\right)-f) \\
& + \sum_{i=1}^{K} \lambda_i^{C^+} a_i(f\left(a_i-q; J^{C_i^+}(\mathbf{b})\right)-f) \\
& + \lambda^{M^-}\left(f\left(J^{M^-}(\mathbf{a}); [b_i+(q-B(i-1))_+]_-\right)-f\right) \\
& + \sum_{i=1}^{K} \lambda_i^{L^-}(f\left(J^{L_i^-}(\mathbf{a}); b_i-q\right)-f) \\
& + \sum_{i=1}^{K} \lambda_i^{C^-}|b_i|(f\left(J^{C_i^-}(\mathbf{a}); b_i+q\right)-f),
\end{aligned} \tag{4.1}$$

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

$$x_+ := \max(x,0), \qquad x_- := \min(x,0), x\in\mathbb{R}. \tag{4.2}$$

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

$$A(i) := \sum_{k=1}^{i} a_k, \tag{4.3}$$

and

$$B(i) := \sum_{k=1}^{i} |b_k|, \tag{4.4}$$

and the generalized inverse functions thereof

$$A^{-1}(q') := \inf\{p : \sum_{j=1}^{p} a_j > q'\}, \tag{4.5}$$

and

$$B^{-1}(q') := \inf\{p : \sum_{j=1}^{p} |b_j| > q'\}, \tag{4.6}$$

where q' designates a certain quantity of shares[9].

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

$$i_A := A^{-1}(0) = \inf\{p : \sum_{j=1}^{p} a_j > 0\} = \frac{S}{\Delta P} := i_S, \tag{4.7}$$

and

$$i_B := B^{-1}(0) = \inf\{p : \sum_{j=1}^{p} |b_j| > 0\} = \frac{S}{\Delta P} := i_S = i_A. \tag{4.8}$$

5. Price Dynamics

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

$$\begin{aligned} dP^A(t) &= \Delta P[(A^{-1}(q) - A^{-1}(0))dM^+(t) \\ &- \sum_{i=1}^{K} (A^{-1}(0) - i)_+ dL_i^+(t) + (A^{-1}(q) - A^{-1}(0))dC_{i_A}^+(t)], \end{aligned} \tag{5.1}$$

and

$$\begin{aligned} dP^B(t) &= -\Delta P[(B^{-1}(q) - B^{-1}(0))dM^-(t) \\ &- \sum_{i=1}^{K} (B^{-1}(0) - i)_+ dL_i^-(t) + (B^{-1}(q) - B^{-1}(0))dC_{i_B}^-(t)], \end{aligned} \tag{5.2}$$

which describe the various events that affect them: change due to a market order, change due to limit orders inside the spread, and change due to the

[9]Note that a more rigorous notation would be

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

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

cancellation of a limit order at the best price. Equivalently, the respective dynamics of the mid-price and the spread are:

$$\begin{aligned} dP(t) = \frac{\Delta P}{2}\Big[&(A^{-1}(q) - A^{-1}(0))dM^+(t) - (B^{-1}(q) - B^{-1}(0))dM^-(t) \\ &- \sum_{i=1}^{K}(A^{-1}(0) - i)_+ dL_i^+(t) + \sum_{i=1}^{K}(B^{-1}(0) - i)_+ dL_i^-(t) \\ &+ (A^{-1}(q) - A^{-1}(0))dC_{i_A}^+(t) - (B^{-1}(q) - B^{-1}(0))dC_{i_B}^-(t)\Big], \end{aligned} \tag{5.3}$$

$$\begin{aligned} dS(t) = \Delta P\Big[&(A^{-1}(q) - A^{-1}(0))dM^+(t) + (B^{-1}(q) - B^{-1}(0))dM^-(t) \\ &- \sum_{i=1}^{K}(A^{-1}(0) - i)_+ dL_i^+(t) - \sum_{i=1}^{K}(B^{-1}(0) - i)_+ dL_i^-(t) \\ &+ (A^{-1}(q) - A^{-1}(0))dC_{i_A}^+(t) + (B^{-1}(q) - B^{-1}(0))dC_{i_B}^-(t)\Big]. \end{aligned} \tag{5.4}$$

The equations above are interesting in that they relate in an explicit way the profile of the order book to the size of an increment of the mid-price or the spread, therefore linking the price dynamics to the order flow. For instance the infinitesimal drifts of the mid-price and the spread, conditional on the shape of the order book at time t, are given by:

$$\begin{aligned} \mathbb{E}\left[dP(t)|(\mathbf{a};\mathbf{b})\right] = \frac{\Delta P}{2}\Big[&(A^{-1}(q) - A^{-1}(0))\lambda^{M^+} - (B^{-1}(q) - B^{-1}(0))\lambda^{M^-} \\ &- \sum_{i=1}^{K}(A^{-1}(0) - i)_+ \lambda_i^{L^+} + \sum_{i=1}^{K}(B^{-1}(0) - i)_+ \lambda_i^{L^-} \\ &+ (A^{-1}(q) - A^{-1}(0))\lambda_{i_A}^{C^+} a_{i_A} - (B^{-1}(q) - B^{-1}(0))\lambda_{i_B}^{C^-}|b_{i_B}|\Big]\,dt, \end{aligned} \tag{5.5}$$

and

$$\begin{aligned} \mathbb{E}\left[dS(t)|(\mathbf{a};\mathbf{b})\right] = \Delta P\Big[&(A^{-1}(q) - A^{-1}(0))\lambda^{M^+} + (B^{-1}(q) - B^{-1}(0))\lambda^{M^-} \\ &- \sum_{i=1}^{K}(A^{-1}(0) - i)_+ \lambda_i^{L^+} - \sum_{i=1}^{K}(A^{-1}(0) - i)_+ \lambda_i^{L^-} \\ &+ (A^{-1}(q) - A^{-1}(0))\lambda_{i_A}^{C^+} a_{i_A} + (B^{-1}(q) - B^{-1}(0))\lambda_{i_B}^{C^-}|b_{i_B}|\Big]\,dt. \end{aligned} \tag{5.6}$$

6. Ergodicty and Diffusive Limit

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

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

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

$$Q^t(\mathbf{x}, E) := \mathbb{P}\left[\mathbf{X}(t) \in E | \mathbf{X}(0) = \mathbf{x}\right], \; t \in \mathbb{R}_+, \mathbf{x} \in \mathcal{S}, E \subset \mathcal{S}, \tag{6.1}$$

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

$$\lim_{t\to\infty} ||Q^t(\mathbf{x}, .) - \pi(.)|| = 0, \forall \mathbf{x} \in \mathcal{S}, \tag{6.2}$$

where $||.||$ designates for a signed measure ν the *total variation norm*[10] defined as

$$||\nu|| := \sup_{f:|f|<1} |\nu(f)| = \sup_{E\in\mathcal{B}(\mathcal{S})} \nu(E) - \inf_{E\in\mathcal{B}(\mathcal{S})} \nu(E). \tag{6.4}$$

In (2.32), $\mathcal{B}(\mathcal{S})$ is the Borel σ-field generated by $\mathcal{S}$, and for a measurable function f on $\mathcal{S}$, $\nu(f) := \int_{\mathcal{S}} f d\nu$.

V-uniform ergodicity. A Markov process is said *V-uniformly ergodic* if there exists a coercive[11] function $V > 1$, an invariant distribution π, and constants $0 < r < 1$, and $R < \infty$ such that

$$||Q^t(\mathbf{x}, .) - \pi(.)|| \leq R r^t V(\mathbf{x}), \mathbf{x} \in \mathcal{S}, t \in \mathbb{R}_+. \tag{6.5}$$

$V-$uniform ergodicity can be characterized in terms of the infinitesimal generator of the Markov process. Indeed, it is shown in [**MT09, MT93b**] that it is equivalent to the existence of a coercive function V (the "Lyapunov test function") such that

$$\mathcal{L}V(\mathbf{x}) \leq -\beta V(\mathbf{x}) + \gamma, \quad \text{(Geometric drift condition.)} \tag{6.6}$$

[10]The convergence in total variation norm implies the more familiar pointwise convergence

$$\lim_{t\to\infty} |Q^t(\mathbf{x}, \mathbf{y}) - \pi(\mathbf{y})| = 0, \mathbf{x}, \mathbf{y} \in \mathcal{S}. \tag{6.3}$$

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

[11]That is, a function such that $V(\mathbf{x}) \to \infty$ as $|\mathbf{x}| \to \infty$.

for some positive constants β and γ. (Theorems 6.1 and 7.1 in [**MT93b**].) Intuitively, condition (2.34) says that the larger $V(\mathbf{X}(t))$ the stronger $\mathbf{X}$ is pulled back towards the center of the state space $\mathcal{S}$. A similar drift condition is available for discrete-time Markov processes $(\mathbf{X}_n)_{n\in\mathbb{N}}$ and reads

$$\mathcal{D}V(\mathbf{x}) \leq -\beta V(\mathbf{x}) + \gamma \mathbb{I}_C(\mathbf{x}), \tag{6.7}$$

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

$$\mathcal{D}V(\mathbf{x}) := \mathbb{E}[V(\mathbf{X}_{n+1}) - V(\mathbf{X}_n)|\mathbf{X}_n = \mathbf{x}]. \tag{6.8}$$

and $C \subset \mathcal{S}$ a finite set. (Theorem 16.0.1 in [**MT09**].) We refer to [**MT09**] for further details.

6.1. Ergodicity of the Order Book and Rate of Convergence to the Stationary State. Of utmost interest is the behavior of the order book in its stationary state. We have the following result:

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

$$\|Q^t(\mathbf{x}, .) - \pi(.)\| \leq Rr^t V(\mathbf{x}), t \in \mathbb{R}^+, \mathbf{x} \in \mathcal{S}. \tag{6.9}$$

PROOF. Let

$$V(\mathbf{x}) := V(\mathbf{a}; \mathbf{b}) := \sum_{i=1}^K a_i + \sum_{i=1}^K |b_i| + q \tag{6.10}$$

be the total number of shares in the book ($+q$ shares). Using the expression of the infinitesimal generator (4.1) we have

$$\begin{aligned}
\mathcal{L}V(\mathbf{x}) &\leq -(\lambda^{M^+} + \lambda^{M^-})q + \sum_{i=1}^K (\lambda_i^{L^+} + \lambda_i^{L^-})q - \sum_{i=1}^K (\lambda_i^{C^+} a_i + \lambda_i^{C^-}|b_i|)q \\
&\quad + \sum_{i=1}^K \lambda_i^{L^+}(i_S - i)_+ a_\infty + \sum_{i=1}^K \lambda_i^{L^+}(i_S - i)_+ |b_\infty| \qquad (6.11)\\
&\leq -(\lambda^{M^+} + \lambda^{M^-})q + (\Lambda^{L^-} + \Lambda^{L^+})q - \underline{\lambda^C} q V(\mathbf{x}) \\
&\quad + K(\Lambda^{L^-} a_\infty + \Lambda^{L^+}|b_\infty|), \qquad (6.12)
\end{aligned}$$

where

$$\Lambda^{L^\pm} := \sum_{i=1}^{K} \lambda_i^{L^\pm} \text{ and } \underline{\lambda^C} := \min_{1\le i\le K} \{\lambda_i^{C^\pm}\} > 0. \tag{6.13}$$

The first three terms in the right hand side of inequality (6.11) correspond respectively to the arrival of a market, limit or cancellation order—ignoring the effect of the shift operators. The last two terms are due to shifts occurring after the arrival of a limit order inside the spread. The terms due to shifts occurring after market or cancellation orders (which we do not put in the r.h.s. of (6.11)) are negative, hence the inequality. To obtain inequality (6.12), we used the fact that the spread i_S is bounded by $K+1$—a consequence of the boundary conditions we impose— and hence $(i_S - i)_+$ is bounded by K.

The drift condition (6.12) can be rewritten as

$$\mathcal{L}V(\mathbf{x}) \le -\beta V(\mathbf{x}) + \gamma, \tag{6.14}$$

for some positive constants β, γ. Inequality (6.14) together with theorem 7.1 in [**MT93b**] let us assert that $(\mathbf{X}(t))$ is V-uniformly ergodic, hence (6.9). □

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

6.2. The Embedded Markov Chain. Let $(\mathbf{X}_n)$ denote the *embedded* Markov chain associated with $(\mathbf{X}(t))$. In event time, the probabilities of each event are "normalized" by the quantity

$$\Lambda(\mathbf{x}) := \lambda^{M^+} + \lambda^{M^-} + \Lambda^{L^+} + \Lambda^{L^-} + \sum_{i=1}^{K} \lambda_i^{C^+} a_i + \sum_{i=1}^{K} \lambda_i^{C^-} |b_i|. \tag{6.15}$$

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

$$\mathbb{P}[\text{"Buy market order at time } n\text{"}|\mathbf{X}_{n-1} = \mathbf{x}] := p^{M^+}(\mathbf{x}) = \frac{\lambda^{M^+}}{\Lambda(\mathbf{x})}. \tag{6.16}$$

The choice of the test function $V(\mathbf{x}) = \sum_i a_i + \sum_i b_i + q$ does not yield a geometric drift condition, and more care should be taken to obtain a suitable test function. Let $z > 1$ be a fixed real number and consider the function[12]

$$V(\mathbf{x}) := z^{\sum_i a_i + \sum_i |b_i|} := z^{\varphi(\mathbf{x})}. \tag{6.17}$$

[12]To save notations, we always use the letter V for the test function.

We have

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

$$\|U^n(\mathbf{x}, .) - \nu(.)\| \leq R_2 r_2^n V(\mathbf{x}), n \in \mathbb{N}, , \mathbf{x} \in \mathcal{S}. \tag{6.18}$$

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

PROOF.

$$\begin{aligned}
\mathcal{D}V(\mathbf{x}) &\leq \frac{\lambda^{M^+}}{\Lambda(\mathbf{x})}(z^{\sum_i a_i - q + \sum_i |b_i|} - V(\mathbf{x})) \\
&+ \sum_j \frac{\lambda_j^{L^+}}{\Lambda(\mathbf{x})}(z^{\sum_i a_i + q + \sum_i |b_i| + K|b_\infty|} - V(\mathbf{x})) \\
&+ \sum_j \frac{\lambda_j^{C^+} a_j}{\Lambda(\mathbf{x})}(z^{\sum_i a_i - q + \sum_i |b_i|} - V(\mathbf{x})) \\
&+ \frac{\lambda^{M^-}}{\Lambda(\mathbf{x})}(z^{\sum_i a_i + \sum_i |b_i| - q} - V(\mathbf{x})) \\
&+ \sum_j \frac{\lambda_j^{L^-}}{\Lambda(\mathbf{x})}(z^{\sum_i a_i + K a_\infty + \sum_i |b_i| + q} - V(\mathbf{x})) \\
&+ \sum_j \frac{\lambda_j^{C^-} |b_j|}{\Lambda(\mathbf{x})}(z^{\sum_i a_i + \sum_i |b_i| - q} - V(\mathbf{x})).
\end{aligned} \tag{6.19}$$

If we factor out $V(\mathbf{x}) = z^{\sum a_i + \sum b_i}$ in the r.h.s of (6.19), we get

$$\begin{aligned}
\frac{\mathcal{D}V(\mathbf{x})}{V(\mathbf{x})} &\leq \frac{\lambda^{M^+} + \lambda^{M^-}}{\Lambda(\mathbf{x})}(z^{-q} - 1) \\
&+ \frac{\Lambda^{L^-} + \Lambda^{L^-}}{\Lambda(\mathbf{x})}(z^{q + K d_\infty} - 1) \\
&+ \frac{\sum_j \lambda_j^{C^+} a_j + \sum_j \lambda_j^{C^-} |b_j|}{\Lambda(\mathbf{x})}(z^{-q} - 1),
\end{aligned} \tag{6.20}$$

where

$$d_\infty := \max\{a_\infty, |b_\infty|\}. \tag{6.21}$$

Then

$$\begin{aligned}\frac{\mathcal{D}V(\mathbf{x})}{V(\mathbf{x})} &\leq \frac{\lambda^{M^+}+\lambda^{M^-}}{\lambda^{M^+}+\lambda^{M^-}+\Lambda^{L^+}+\Lambda^{L^-}+\overline{\lambda^C}\varphi(\mathbf{x})}(z^{-q}-1)\\ &+ \frac{\Lambda^{L^+}+\Lambda^{L^-}}{\lambda^{M^+}+\lambda^{M^-}+\Lambda^{L^+}+\Lambda^{L^-}+\underline{\lambda^C}\varphi(\mathbf{x})}(z^{q+Kd_\infty}-1)\\ &+ \frac{\underline{\lambda^C}\varphi(\mathbf{x})}{\lambda^{M^+}+\lambda^{M^-}+\Lambda^{L^+}+\Lambda^{L^-}+\overline{\lambda^C}\varphi(\mathbf{x})}(z^{-q}-1), \qquad (6.22)\end{aligned}$$

with the usual notations

$$\underline{\lambda^C} := \min \lambda_i^{C^\pm} \text{ and } \overline{\lambda^C} := \max \lambda_i^{C^\pm}. \qquad (6.23)$$

Denote the r.h.s of (6.22) $B(\mathbf{x})$. Clearly

$$\lim_{\varphi(\mathbf{x})\to\infty} B(\mathbf{x}) = \frac{\underline{\lambda^C}(z^{-q}-1)}{\overline{\lambda^C}} < 0, \qquad (6.24)$$

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

$$\frac{\mathcal{D}V(\mathbf{x})}{V(\mathbf{x})} \leq \frac{\underline{\lambda^C}(z^{-q}-1)}{2\overline{\lambda^C}} := -\beta < 0. \qquad (6.25)$$

Let C denote the finite set

$$C = \{\mathbf{x} \in \mathcal{S} : \varphi(\mathbf{x}) = \sum_i a_i + \sum_i b_i \leq A\}. \qquad (6.26)$$

We have

$$\mathcal{D}V(\mathbf{x}) \leq -\beta V(\mathbf{x}) + \gamma \mathbb{I}_C(\mathbf{x}), \qquad (6.27)$$

with

$$\gamma := \max_{\mathbf{x}\in C} \mathcal{D}V(\mathbf{x}). \qquad (6.28)$$

Therefore $(\mathbf{X}_n)_{n\geq 0}$ is V-uniformly ergodic, by theorem 16.0.1 in [**MT09**]. □

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

$$\mathbb{P}[C_i^\pm(t+\delta t) - C_i^\pm(t) = 1 | \mathbf{X}'(t) = \mathbf{x}'] = \lambda_i^{C^\pm}\delta t + o(\delta t), \qquad (6.29)$$

instead of

$$\mathbb{P}[C_i^+(t+\delta t) - C_i^+(t) = 1 | \mathbf{X}'(t) = \mathbf{x}'] = \lambda_i^{C^+} a_i(t)\delta t + o(\delta t), \qquad (6.30)$$

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

We have the following result:

Theorem 6.3. *Set*

$$\Lambda^{C^\pm} := \sum_{i=1}^K \lambda_i^{C^\pm} \text{ and } \Lambda^{L^\pm} := \sum_{i=1}^K \lambda_i^{L^\pm}. \tag{6.31}$$

Under the condition

$$\lambda^{M^+} + \lambda^{M^-} + \Lambda^{C^+} + \Lambda^{C^-} > (\Lambda^{L^+} + \Lambda^{L^-})(1 + Kd_\infty), \tag{6.32}$$

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

$$\|U'^n(\mathbf{x}, .) - \nu'(.)\| \leq R_3 r_3^n V(\mathbf{x}), n \in \mathbb{N}, \mathbf{x} \in \mathcal{S}. \tag{6.33}$$

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

Proof. Let us prove the result for $(\mathbf{X}'_n)$. Inequality (6.20) is still valid by the same arguments, but this time the arrival rates are independent of $\mathbf{x}'$

$$\begin{aligned} \frac{\mathcal{D}V(\mathbf{x}')}{V(\mathbf{x}')} &\leq \frac{\lambda^{M^+} + \lambda^{M^-}}{\Lambda}(z^{-q} - 1) \\ &+ \frac{\Lambda^{L^+} + \Lambda^{L^-}}{\Lambda}(z^{q+Kd_\infty} - 1) \\ &+ \frac{\Lambda^{C^+} + \Lambda^{C^-}}{\Lambda}(z^{-q} - 1). \end{aligned} \tag{6.34}$$

Set

$$z =: 1 + \epsilon > 1. \tag{6.35}$$

A Taylor expansion in ϵ gives

$$\begin{aligned} \Lambda \frac{\mathcal{D}V(\mathbf{x})}{V(\mathbf{x})} &\leq (\lambda^{M^+} + \lambda^{M^-})(-q\epsilon) \\ &+ (\Lambda^{L^+} + \Lambda^{L^-})(q + Kd_\infty)\epsilon \\ &+ (\Lambda^{C^+} + \Lambda^{C^-})(-q\epsilon) + o(\epsilon). \end{aligned} \tag{6.36}$$

For $\epsilon > 0$ small enough, the sign of (6.36) is determined by the quantity

$$-(\lambda^{M^+} + \lambda^{M^-}) + (\Lambda^{L^+} + \Lambda^{L^-})(1 + Kd_\infty) - (\Lambda^{C^+} + \Lambda^{C^-}). \tag{6.37}$$

Hence, if (6.32) holds

$$\mathcal{D}V(\mathbf{x}) \quad \leq -\beta V(\mathbf{x}) \;\text{ for some } \beta > 0, \tag{6.38}$$

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

If for concreteness we set $q = 1$ share, and all the arrival rates are symmetric and do not depend on i, then condition (6.32) can be rewritten as

$$\lambda^M + K\lambda^C > K\lambda^L(1 + Kd_\infty). \tag{6.39}$$

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

6.4. Large-scale Limit of the Price Process. We are now able to answer the main question of this chapter. Let us define the process

$$e(t) \in \{1, \ldots, 2(2K + 1)\}$$

which indicates the last event

$$\{M^\pm, L_i^\pm, C_i^\pm\}_{i\in\{1,\ldots,K\}},$$

that has occurred before time t.

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

$$\mathbf{Y}(t) := (\mathbf{X}(t), e(t)) \tag{6.40}$$

which still satisfies the drift condition (2.34).

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

$$V(\mathbf{y}) \;:=\; q + \sum a_i + \sum |b_i| + e, \;\; \text{(continuous-time setting)} \tag{6.41}$$

$$V(\mathbf{y}) \;:=\; e^{\sum a_i + \sum |b_i| + e}. \;\; \text{(discrete-time setting)} \tag{6.42}$$

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

□

Given the state $\mathbf{X}_{n-1}$ of the order book at time $n-1$ and the event e_n, the price increment at time n can be determined. (See equation (5.3).) We define the sequence of random variables

$$\eta_n := \Psi(\mathbf{X}_{n-1}, e_n) := \Phi(\mathbf{Y}_n, \mathbf{Y}_{n-1}), \tag{6.43}$$

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

$$P_n := \sum_{k=1}^{n} \overline{\eta}_n = \sum_{k=1}^{n} \overline{\Phi}(\mathbf{Y}_k, \mathbf{Y}_{k-1}), \tag{6.44}$$

where

$$\overline{\eta}_k := \eta_k - \mathbb{E}_\mu[\eta_k] = \overline{\Phi}_k = \Phi_k - \mathbb{E}_\mu[\Phi_k], \tag{6.45}$$

and the asymptotic behavior of the rescaled-centered price process

$$\widetilde{P}^{(n)}(t) := \frac{P_{\lfloor nt \rfloor}}{\sqrt{n}}, \tag{6.46}$$

as n goes to infinity.

Theorem 6.4. *The series*

$$\sigma^2 = \mathbb{E}_\mu[\overline{\eta}_0^2] + 2\sum_{n=1}^{\infty} \mathbb{E}_\mu[\overline{\eta}_0 \overline{\eta}_n] \tag{6.47}$$

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

$$\widetilde{P}^{(n)}(t) \xrightarrow{n\to\infty} \sigma B(t), \tag{6.48}$$

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

Proof. The idea is to apply the functional central limit theorem for (stationary and ergodic) sequences of *weakly dependent* random variables with *finite variance*. Firstly, we note that the variance of the price increments η_n is finite since it is bounded by $K+1$. Secondly, the V-uniform ergodicity of $(\mathbf{Y}_n)$ is equivalent to

$$\|M^n(\mathbf{x}, .) - \mu(.)\| \leq R\rho^n V(\mathbf{x}), n \in \mathbb{N}, \tag{6.49}$$

for some $R < \infty$ and $\rho < 1$. This implies thanks to theorem 16.1.5 in [**MT09**][13] that for any $g^2, h^2 \leq V$, $k, n \in \mathbb{N}$, and any initial condition $\mathbf{y}$

$$|\mathbb{E}_{\mathbf{y}}[g(\mathbf{Y}_k)h(\mathbf{Y}_{k+n})] - \mathbb{E}_{\mathbf{y}}[g(\mathbf{Y}_k)]\mathbb{E}_{\mathbf{y}}[h(\mathbf{Y}_{k+n})]| \leq R\rho^n[1 + \rho^k V(\mathbf{y})], \tag{6.50}$$

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

$$|\mathbb{E}_{\mathbf{y}}[\overline{h}(\mathbf{Y}_k)\overline{g}(\mathbf{Y}_{k+n})]| \leq R_1\rho^n[1 + \rho^k V(\mathbf{y})] \tag{6.51}$$

for some $R_1 < \infty$, where $\overline{h} = h - \mathbb{E}_\mu[h], \overline{g} = g - \mathbb{E}_\mu[g]$. By taking the expectation over μ on both sides of (4.17) and noting that $\mathbb{E}_\mu[V(\mathbf{Y}_0)]$ is finite by theorem 14.3.7 in [**MT09**], we get

$$|\mathbb{E}_\mu[\overline{h}(\mathbf{Y}_k)\overline{g}(\mathbf{Y}_{k+n})]| \leq R_2\rho^n =: \rho(n), k, n \in \mathbb{N}. \tag{6.52}$$

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

$$\sum_n \rho(n) < \infty. \tag{6.53}$$

Theorems 19.2 and 19.3 in [**Bil99**] on functions of mixing processes[14] let us conclude that

$$\sigma^2 := \mathbb{E}_\mu[\overline{\eta}_0^2] + 2\sum_{n=1}^{\infty} \mathbb{E}_\mu[\overline{\eta}_0\overline{\eta}_n] \tag{6.54}$$

is well-defined—the series in (4.20) converges absolutely—and coincides with the asymptotic variance

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E}_\mu\left[\sum_{k=1}^{n} (\overline{\eta}_k)^2\right] = \sigma^2. \tag{6.55}$$

Moreover

$$\widetilde{P}^{(n)}(t) \overset{n\to\infty}{\longrightarrow} \sigma B(t), \tag{6.56}$$

where $(B(t))$ is a standard Brownian motion. The convergence in (4.22) happens in $D[0, \infty)$, the space of $\mathbb{R}$-valued càdlàg functions, equipped with the Skorohod topology. □

Remark 6.1. *In the large-scale limit, the mid-price P, the ask price $P^A = P + \frac{S}{2}$, and the bid price $P^B = P - \frac{S}{2}$ converge to the same process $(\sigma B(t))$.*

[13] We refer to §16.1.2 *"V-geometric mixing and V-uniform ergodicity"* in [**MT09**] for more details.

[14] See also theorem 4.4.1 in [**Whi02**] and discussion therein.

REMARK 6.2. *Theorem 4.2 is also true with constant cancellation rates under condition* (6.32). *In this case the result holds both in event time and physical time. Indeed, let* $(N(t))_{t\in\mathbb{R}_+}$ *denote a Poisson process with intensity* $\Lambda = \lambda^{M^\pm} + \Lambda^{L^\pm} + \sum_{i=1}^{K} \lambda_i^{C^\pm}$. *The price process in physical time* $(P_c(t))_{t\in\mathbb{R}_+}$ *can be linked to the price in event time* $(P_n)_{n\in\mathbb{N}}$ *by*

$$P_c(t) = P_{N(t)}. \tag{6.57}$$

Then

$$\frac{P_{\lfloor kt \rfloor}}{\sqrt{k}} \quad \overset{k\to\infty}{\longrightarrow} \sigma B(t) \ \text{ as in theorem 4.2,} \tag{6.58}$$

$$\text{and since } \tfrac{N(u)}{\Lambda u} \overset{u\to\infty}{\longrightarrow} 1 \text{ a.s.,}$$

$$\frac{P_c(kt)}{\sqrt{k}} \quad = \tfrac{P_{N(kt)}}{\sqrt{k}} \overset{k\to\infty}{\sim} \to \tfrac{P_{\lfloor \Lambda kt \rfloor}}{\sqrt{k}} \overset{k\to\infty}{\longrightarrow} \sqrt{\Lambda}\sigma B(t). \tag{6.59}$$

REMARK 6.3. Yet another specification of the cancellation process. *Another interesting specification of the cancellation process* $(C_i(t))$ *is to assume that the arrival rate is constant (for each i) but the canceled* volume *is proportional to the queue size* $|\mathbf{X}_i|$. *In this case, the treatments of the continuous time chain and its embedded discrete-time counterpart are equivalent, and theorems 6.1–4.2 can be obtained in an analogous manner to the proofs in this section.*

CHAPTER 3

Markovian Order Book Models II: Numerical Example

1. Numerical Example

In order to gain a better intuitive understanding of the "mechanics" of the model, we sketch in Algorithm 1 below the simulation procedure in pseudo-code (See also [**GO10**] for a similar description.) For simplicity, we take a symmetric order book model. We also let (usual notations):

$$\begin{aligned}
\boldsymbol{\lambda}^L &:= \left(\lambda_1^L, \ldots, \lambda_K^L\right), && (1.1)\\
\Lambda^L &:= \sum_{i=1}^{K} \lambda_i^L, && (1.2)\\
\boldsymbol{\lambda}^C(\mathbf{a}) &:= \left(\lambda_1^C a_1, \ldots, \lambda_K^C a_K\right), && (1.3)\\
\Lambda^C(\mathbf{a}) &:= \sum_{i=1}^{K} \lambda_i^C a_i, && (1.4)\\
\boldsymbol{\lambda}^C(\mathbf{b}) &:= \left(\lambda_1^C |b_1|, \ldots, \lambda_K^C |b_K|\right), && (1.5)\\
\Lambda^C(\mathbf{b}) &:= \sum_{i=1}^{K} \lambda_i^C |b_i|, && (1.6)\\
\Lambda(\mathbf{a}, \mathbf{b}) &:= 2(\lambda^M + \Lambda^L) + \Lambda^C(\mathbf{a}) + \Lambda^C(\mathbf{b}). && (1.7)
\end{aligned}$$

In order to put the simulation results and the data on the same footing, we relax the assumption of constant order sizes; we draw the order volumes from lognormal distributions.

The parameters of the model are estimated from tick by tick data as detailed in section 2. For concreteness[1], we use the parameters of the stock SCHN.PA (Schneider Electric) in March 2011 for the plots. They are summarized in tables 3 and 4.

[1]The results are qualitatively similar for all CAC 40 stocks.

Algorithm 1 Order book simulation.

Require: *Model parameters*— Arrival rates: $\lambda^M, \{\lambda_i^L\}_{i\in\{1,\dots K\}}, \{\lambda_i^C\}_{i\in\{1,\dots K\}}$, order book size: K, reservoirs: a_∞, b_∞, volume distribution parameters: $(\nu^M, s^M), (\nu^L, s^L), (\nu^C, s^C)$.
Simulation Parameters— Number of time steps: N.
Initialization— $t \leftarrow 0, \mathbf{X}(0) \leftarrow \mathbf{X}_{\text{init}}$.

1: **for** time step $n = 1, \dots, N$, **do**
2: Compute the best bid p^B and best ask p^A.
3: Compute $\Lambda^C(\mathbf{b}) = \sum_{i=1}^{K} \lambda_i^C |b_i|$, i.e. the weighted sum of shares at price levels from $p^A - K$ to $p^A - 1$.
4: Compute $\Lambda^C(\mathbf{a}) = \sum_{i=1}^{K} \lambda_i^C a_i$.
5: Draw the waiting time τ for the next event from an exponential distribution with parameter

$$\Lambda(\mathbf{a}, \mathbf{b}) = 2(\lambda^M + \Lambda^L) + \Lambda^C(\mathbf{a}) + \Lambda^C(\mathbf{b}).$$

6: Draw a new event according to the probability vector

$$\left(\lambda^M, \lambda^M, \Lambda^L, \Lambda^L, \Lambda^C(\mathbf{a}), \Lambda^C(\mathbf{b})\right) / \Lambda(\mathbf{a}, \mathbf{b}).$$

These probabilities correspond respectively to a buy market order, a sell market order, a buy limit order, a sell limit order, a cancellation of an existing sell order and a cancellation of an existing buy order.
7: Depending on the event type, draw the order volume from a lognormal distribution with parameters $(\nu^M, s^M), (\nu^L, s^L)$ or (ν^C, s^C).
8: If the selected event is a limit order, select the relative price level from $\{1, 2, \dots, K\}$ according to the probability vector

$$\left(\lambda_1^L, \dots, \lambda_K^L\right) / \Lambda^L.$$

9: If the selected event is a cancellation, select the relative price level at which to cancel an order from $\{1, 2, \dots, K\}$ according to the probability vector

$$\left(\lambda_1^C a_1, \dots, \lambda_K^C a_K\right) / \Lambda^C(\mathbf{a}).$$

(or $\lambda^C(\mathbf{b})/\Lambda^C(\mathbf{b})$ for the bid side.)
10: Update the order book state according to the selected event.
11: Enforce the boundary conditions:

$$\begin{aligned} a_i &= a_\infty, i \geq K + 1, \\ b_i &= b_\infty, i \geq K + 1. \end{aligned}$$

12: Increment the event time n by 1 and the physical time t by τ.
13: **end for**

Remark 1.1. *For the practical implementation, it is easier to work with an "absolute" price frame $\Delta p \times \{1 \dots L\}$ where $L \gg K$.*

Figure 1 represents the average depth profile, that is, the average number of outstanding shares at a distance of i ticks from the best opposite price. The agreement between the simulation and the data is fairly good (See panel (a) of figure 10 for a cross-sectional view on CAC 40 stocks.) We also plot the distribution of the spread in figure 2. Note that the simulated distribution is *tighter* than the actual one (this is systematic and is documented in panel (b) of figure 10.) Figure 3 shows the fast decay of the autocorrelation function of the price increments. Note the high negative autocorrelation of simulated trade prices relatively to the data. In accordance with the theoretical analysis, figures 4–6 illustrate the asymptotic normality of price increments.

The *signature plot* of the price time series is defined as the variance of price increments at lag h normalized by the lag, that is

$$\sigma_h^2 := \frac{\mathbb{V}\left[P(t+h) - P(t)\right]}{h}. \tag{1.8}$$

This function measures the variance of price increments per time unit. It is interesting in that it shows the transition from the variance at small time scales where micro-structure effects dominate, to the long-term variance. By theorem 4.2[2]

$$\lim_{h\to\infty} \sigma_h^2 = \sigma^2, \text{ for some fixed value } \sigma. \tag{1.9}$$

We verify this numerically in figure 7. Two remarks are in order regarding the signature plot:

Long-term variance— The simulated long-term variance is systematically lower than the variance computed from the data (This is documented in panel (c) of figure 10.) Intuitively, the depth of the order book is expected to increase from the best price towards the center of the book. In the absence of autocorrelation in trade signs, this would cause prices to wander less often far away from the current best as they hit a higher "resistance". We also suspect that actual prices exhibit locally more "drifting phases" than in a (symmetric) Markovian model where the expected price drift is null at all times. An interesting analysis of a simple order book model that allows time-varying arrival rates can be found in [**CS03**].

Short-term variance— The signature plot predicted by the model is too high at short time scales relative to the asymptotic variance, especially for

[2]Strictly spreaking, we proved the result in event-time.

traded prices. This is classically known to be due to bid-ask bounce. It is however remarkable that the signature plot of actual trade prices looks much flatter compared to the simulation (See figure 7.) This was discovered and discussed in detail by Bouchaud et al. in [**BGPW04**], and Lillo and Farmer in [**LF04**] (See also [**FGLM06**] and [**BFL09**].) They note that actual order signs exhibit positive long-ranged correlations. They also note that actual prices are diffusive—the signature plot is flat—even at small time scales. They solve this apparent paradox by showing that diffusivity results from two opposite effects: autocorrelation in trade signs induces persistence in the prices, just at the exact amount to counterbalance the mean reversion induced by the liquidity stored in the order book. This subtle equilibrium between liquidity takers and liquidity providers which guarantees price diffusivity at short lags, is not accounted for by the bare Markovian order book model we study, and one can speak about anomalous diffusion at short time scales for Markovian order book models [**SFGK03**]. Because of the absence of positive autocorrelation in trade signs in the model, this effect is magnified when one looks at trades. The next paragraph elaborates on this point.

1.1. Anomalous Diffusion at Short Time Scales. A qualitative understanding of the discrepancy between the model and the data signature plots at short time scales can be gleaned with the following heuristic argument. In what follows, we reason in trade time t. Denote by $P^{Tr}(t)$ the price of the trade at time t, and $\alpha(t)$ its sign:

$$\alpha(t) = 1, \text{ for a buyer initiated trade, i.e. a buy market order,} \tag{1.10}$$

and,

$$\alpha(t) = -1, \text{ for a seller initiated trade, i.e. a sell market order.} \tag{1.11}$$

We assume that the two signs are equally probable (symmetric model). But to make the argument valid for both the model (for which successive trade signs are independent) and the data (for which trade signs exhibit long memory) we do not assume independence of successive trade signs. Let also for a quantity Z

$$\Delta Z(t) := Z(t+1) - Z(t). \tag{1.12}$$

We have by definition

$$P^{Tr}(t) = P(t^-) + \frac{1}{2}\alpha(t)S(t^-), \tag{1.13}$$

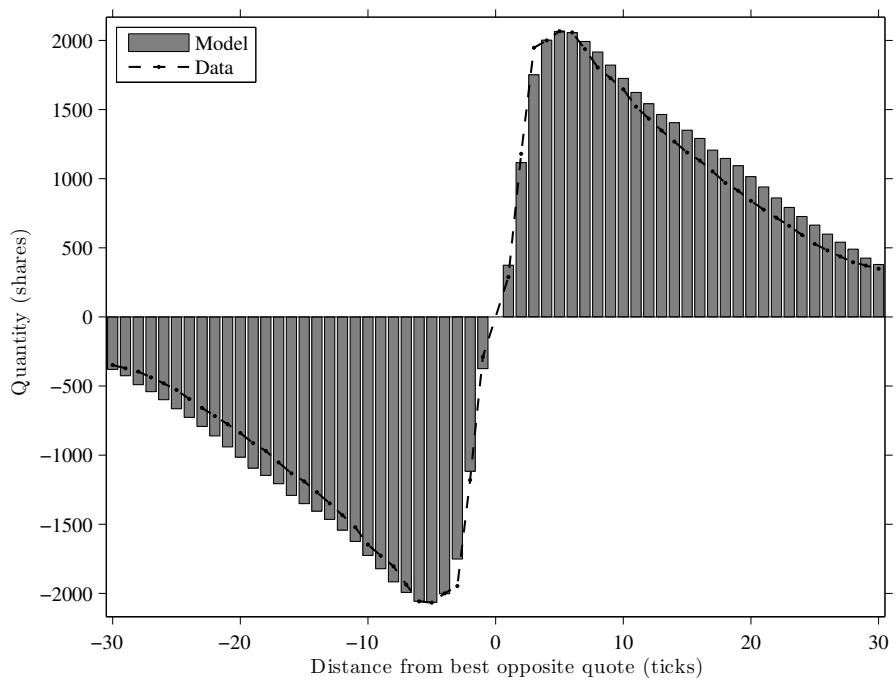


FIGURE 1. Average depth profile. Simulation parameters are summarized in tables 3 and 4.

where $P(t^-)$ and $S(t^-)$ are respectively the prevailing mid-price and spread just before the trade. From equation (1.13)

$$\begin{aligned}
\sigma_1^{Tr\,2} &:= \mathbb{V}[\Delta P^{Tr}(t)] \\
&= \mathbb{E}\left[\left(\Delta P^{Tr}(t)\right)^2\right] \\
&= \mathbb{E}\left[\left(\Delta P(t^-)\right)^2\right] \\
&+ \mathbb{E}\left[\Delta P(t^-)\Delta(\alpha(t)S(t^-))\right] \\
&+ \frac{1}{4}\mathbb{E}\left[\left(\Delta(\alpha(t)S(t^-))\right)^2\right]. \qquad (1.14)
\end{aligned}$$

The first term in the r.h.s. is the variance of mid-price increments σ_1^2. The second term represents the covariance of mid-price increments and the trade sign (weighted by the spread) and we assume it is negligible[3]. Let us focus

[3]This amounts to neglecting the correlation between trade signs and mid-quote movements, which is justified by the dominance of cancellations and limit orders in comparison to market orders in order book data.

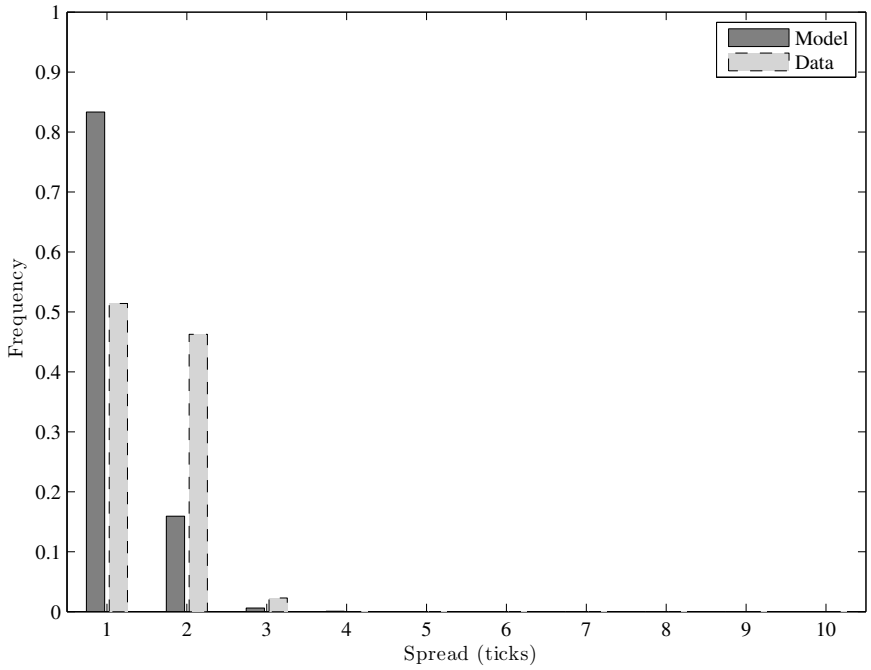


FIGURE 2. Probability distribution of the spread. Note that the model (dark gray) predicts a tighter spread than the data.

on the third term:

$$\Delta(\alpha(t)S(t^-)) = \alpha(t+1)\Delta S(t^-) + S(t^-)\Delta\alpha(t). \tag{1.15}$$

Then

$$\begin{aligned}\mathbb{E}\left[(\Delta(\alpha(t)S(t^-)))^2\right] &= \mathbb{E}\left[(\Delta\alpha(t))^2\right]\mathbb{E}\left[S(t^-)^2\right] \\ &+ 2\mathbb{E}\left[\alpha(t+1)\Delta S(t^-)S(t^-)\Delta\alpha(t)\right] \\ &+ \mathbb{E}\left[\alpha(t+1)^2\right]\mathbb{E}\left[(\Delta S(t^-))^2\right]. \end{aligned} \tag{1.16}$$

Again, we neglect the cross term[4] in the r.h.s. and we are left with

$$\begin{aligned}\mathbb{E}\left[(\Delta(\alpha(t)S(t^-)))^2\right] &\approx \mathbb{E}\left[(\Delta\alpha(t))^2\right]\mathbb{E}\left[S(t^-)^2\right] \\ &+ \mathbb{E}\left[(\Delta S(t^-))^2\right]. \end{aligned} \tag{1.17}$$

[4]This time, we are neglecting the correlation between trade signs and spread movements.

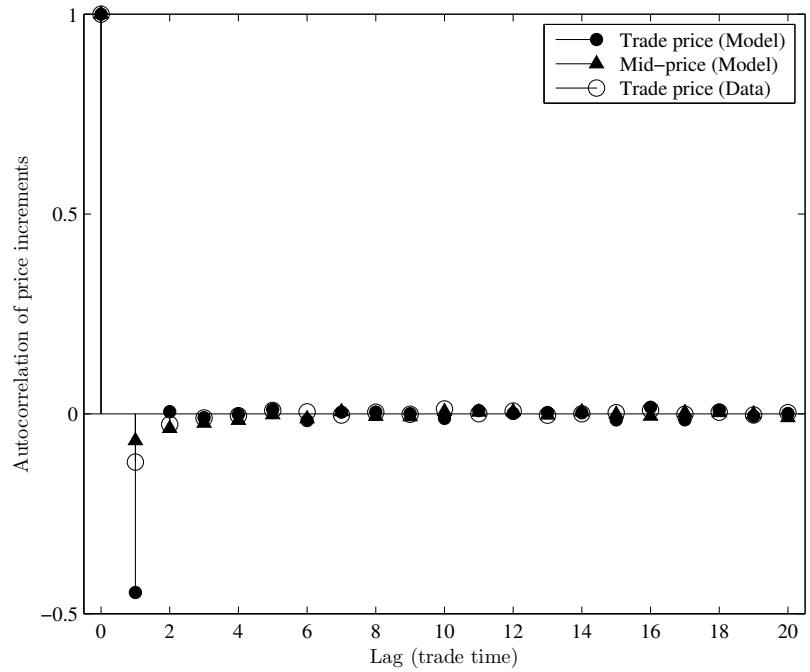


FIGURE 3. Autocorrelation of price increments. This figure shows the fast decay of the autocorrelation function, and the large negative autocorrelation of trades at the first lag.

But

$$\begin{aligned}\mathbb{E}\left[(\Delta\alpha(t))^2\right] &= \mathbb{E}\left[\alpha(t+1)^2\right] + \mathbb{E}\left[\alpha(t)^2\right] - 2\mathbb{E}\left[\alpha(t)\alpha(t+1)\right] \\ &= 2\left(1-\rho_1(\alpha)\right), \end{aligned} \tag{1.18}$$

where $\rho_1(\alpha)$ is the autocorrelation of trade signs at the first lag.

Finally[5]:

$$\sigma_1^{Tr2} \approx {\sigma_1}^2 + \frac{1}{2}\left(1-\rho_1(\alpha)\right)\mathbb{E}\left[S(t^-)^2\right] + \frac{1}{4}\mathbb{E}\left[(\Delta S(t^-))^2\right]. \tag{1.20}$$

Two effects are clear from equation (1.20):

[5]More generally, after n trades:

$$\sigma_n^{Tr2} \approx {\sigma_n}^2 + \frac{1}{2n}\left(1-\rho_n(\alpha)\right)\mathbb{E}\left[S(t^-)^2\right]. \tag{1.19}$$

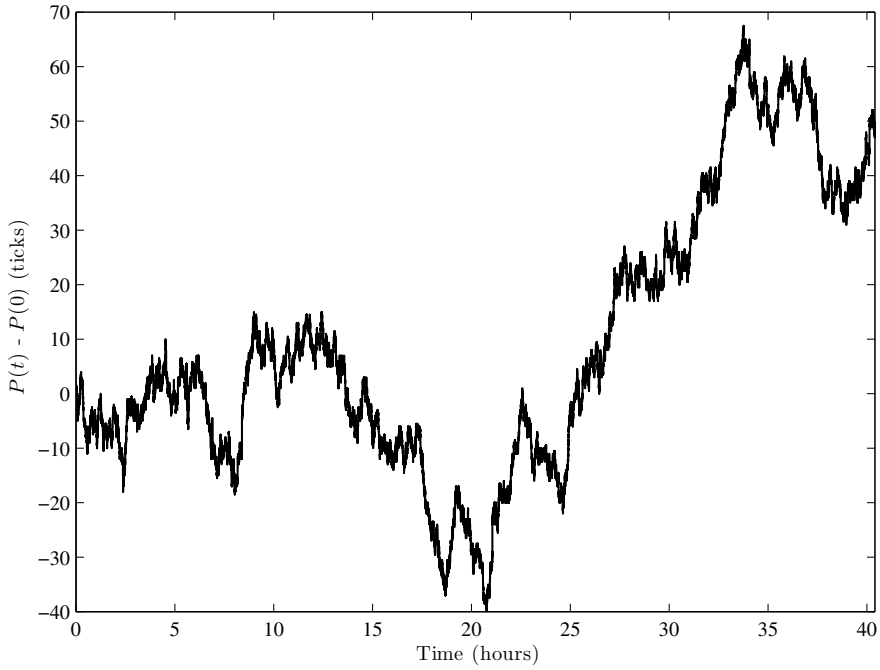


FIGURE 4. Price sample path. At large time scales, the price process is a Brownian motion.

(1) The trade price variance at short time scales is larger than the mid-price variance,
(2) Autocorrelation in trade signs dampens this discrepancy. This partially explains[6] why the trades signature plot obtained from the data is flatter than the model predictions: $\rho_1(\alpha)_{model} = 0$, while $\rho_1(\alpha)_{data} \approx 0.6$.

From a modeling perspective, a possible solution to recover the diffusivity even at very short time scales, is to incorporate long-ranged correlation in the order flow. Toth et al. [**TLD+11**] have investigated numerically this route using a "ϵ-intelligence" order book model. In this model, market orders signs are long-ranged correlated, that is, in trade time

$$\rho_n(\alpha) = \mathbb{E}\left[\alpha(t+n)\alpha(t)\right] \propto n^{-\gamma}, \qquad \gamma \in]0,1[. \tag{1.21}$$

[6]Interestingly, although the arguments that led to (1.20) are rather qualitative, a back of the envelope calculation with $\mathbb{E}\left[S^2\right] \in [1,9]$, gives a difference $\sigma^{Tr^2} - \sigma^2$ in the range $[0.5, 4.5]$; which has the same order of magnitude of the values obtained by simulation.

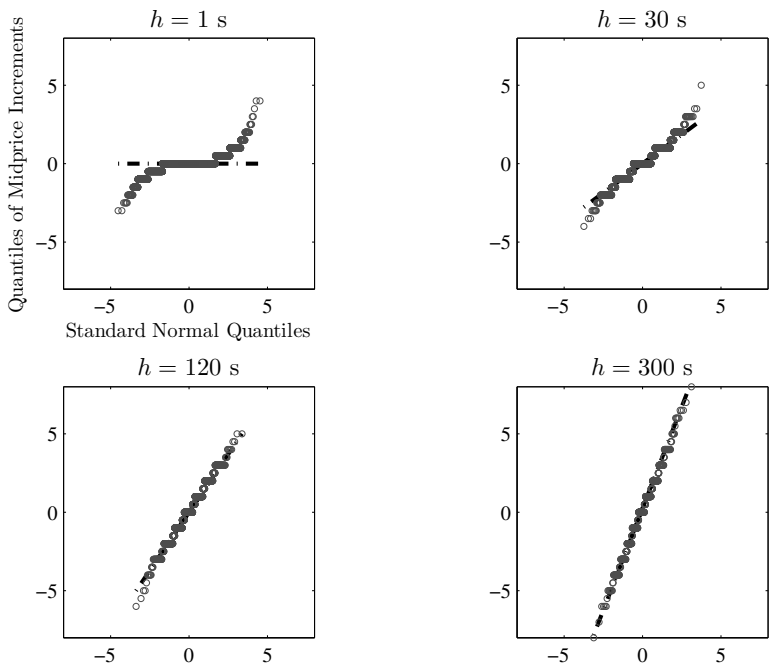


FIGURE 5. Q-Q plot of mid-price increments. h is the time lag in seconds. This figure illustrates the aggregational normality of price increments.

And the size of incoming market orders is a fraction f of the volume displayed at the best opposite quote, with f drawn from the distribution

$$P_\xi(f) = \xi(1-f)^{\xi-1}, \tag{1.22}$$

They show that, by fine tuning the additional parameters γ and ξ, one can ensure the diffusive behavior of the price both at mesoscopic ($\approx$ a few trades) and macroscopic ($\approx$ hundred trades) time scales[7].

[7]Note that Toth. el al. [**TLD+11**] model the "latent order book", not the actual observable order book. The former represents the *intended* volume at each price level p, that is, the volume that would be revealed should the price come close to p. So that the interpretation of their parameters, in particular the expected lifetime τ_{life} of an order, does not strictly match ours.

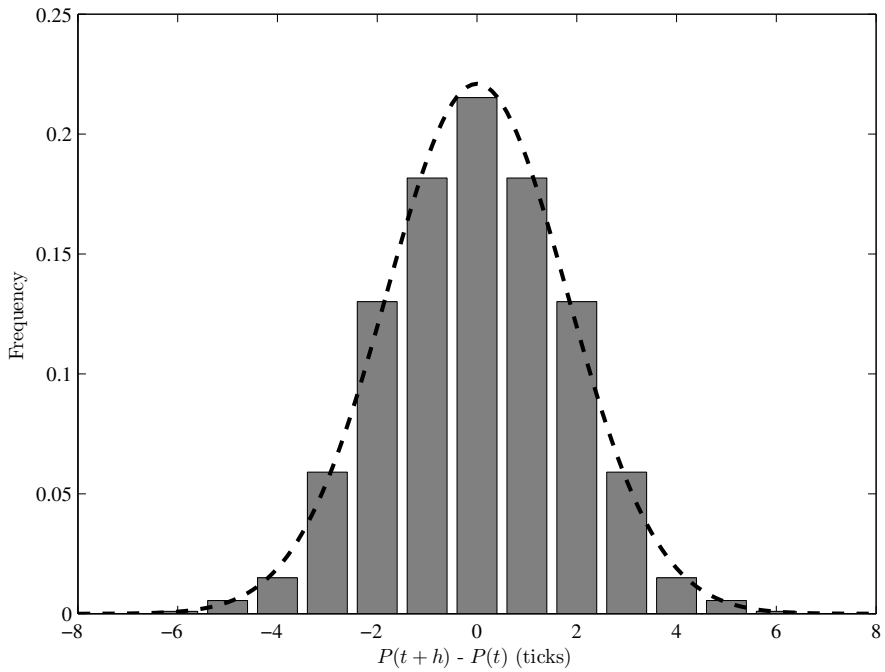


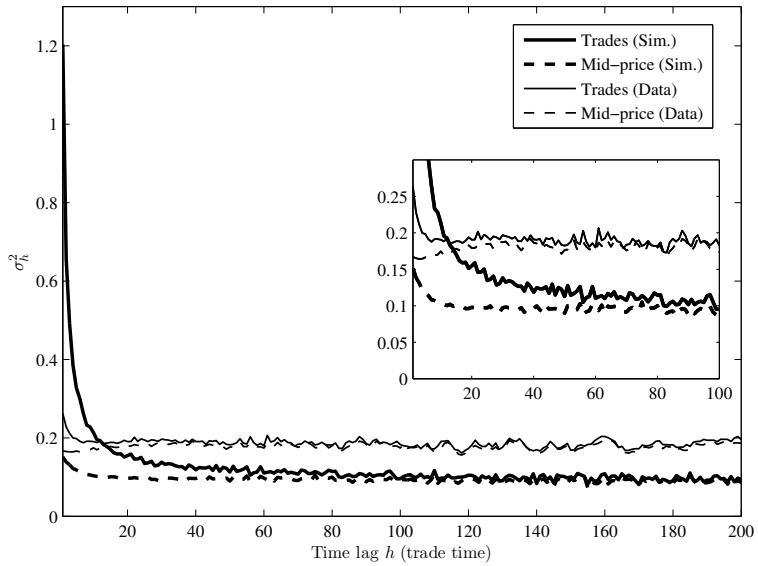
FIGURE 6. Probability distribution of price increments. Time lag h = 1000 events.

2. Model Parameters Estimation

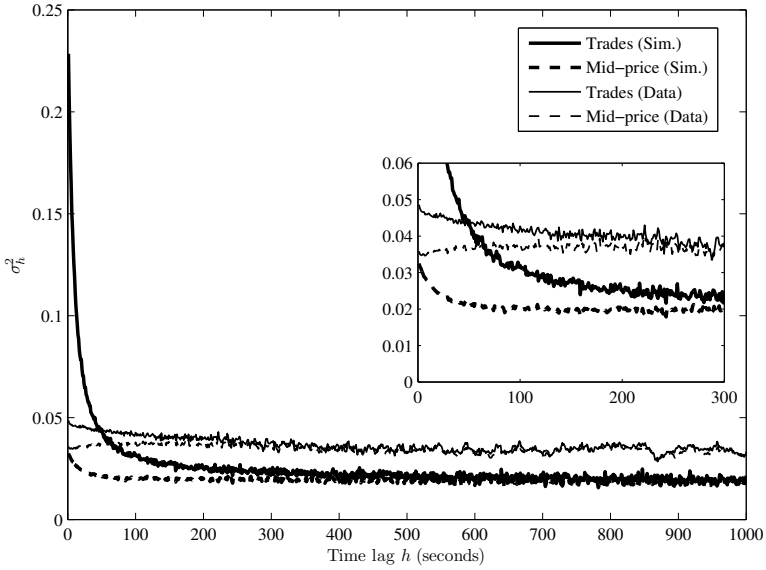
2.1. Description of the Data. For reproducibility, we summarize in tables 3 and 4 the parameters used to obtain figures 1–7. These correspond to estimating the model for the stock SCHN.PA (Schneider Electric). Our dataset consists of TRTH[8] data for the CAC 40 index constituents in March 2011 (23 trading days). We have tick by tick order book data up to 10 price levels, and trades. A snapshot of these files is given in tables 1 and 2. In order to avoid the diurnal seasonality in trading activity (and the impact of the US market open on European stocks), we somehow arbitrarily restrict our attention to the time window [9 : 30–14 : 00] CET.

2.2. Trades and Tick by Tick Data Processing. As one cannot distinguish market orders from cancellations in tick by tick data, and since the timestamps of the trades and tick by tick data files are asynchronous, we use

[8]Thomson Reuters Tick History.



(A) Trade time.



(B) Calendar time.

FIGURE 7. Signature plot: $\sigma_h^2 := \mathbb{V}\,[P(t+h) - P(t)]/h$. y axis unit is tick^2 per trade for panel (a) and tick^2.second^{-1} for panel (b). We used a 1,000,000 event simulation run for the model signature plots. Data signature plots are computed separately for each trading day [9 : 30–14 : 00] then averaged across 23 days. For calendar time signature plots, prices are sampled every second using the last tick rule. The inset is a zoom-in.

Timestamp	Side	Level	Price	Quantity
33480.158	B	1	121.1	480
33480.476	B	2	121.05	1636
33481.517	B	5	120.9	1318
33483.218	*B*	*1*	*121.1*	*420*
33484.254	B	1	121.1	556
33486.832	A	1	121.15	187
33489.014	B	2	121.05	1397
33490.473	*B*	*1*	*121.1*	*342*
33490.473	*B*	*1*	*121.1*	*304*
33490.473	*B*	*1*	*121.1*	*256*
33490.473	A	1	121.15	237

TABLE 1. Tick by tick data file sample. Note that the field "Level" does not necessarily correspond to the distance in ticks from the best opposite quote as there might be gaps in the book. Lines corresponding to the trades in table 2 are highlighted in *italics*.

Timestamp	Last	Last quantity
33483.097	121.1	60
33490.380	121.1	214
33490.380	121.1	38
33490.380	121.1	48

TABLE 2. Trades data file sample.

a matching procedure to reconstruct the order book events. In a nutshell, we proceed as follows for each stock and each trading day:

(1) Parse the tick by tick data file to compute order book state variations:
   - If the variation is positive (volume at one or more price levels has increased), then label the event as a limit order.
   - If the variation is negative (volume at one or more price levels has decreased), then label the event as a "likely market order".

- If no variation—this happens when there is just a renumbering in the field "Level" that does not affect the state of the book—do not count an event.

(2) Parse the trades file and for each trade:
  (a) Compare the trade price and volume to likely market orders whose timestamps are in $[t^{Tr} - \Delta t, t^{Tr} + \Delta t]$, where t^{Tr} is the trade timestamp and Δt is a predefined time window[9].
  (b) Match the trade to the first likely market order with the same price and volume and label the corresponding event as a market order—making sure the change in order book state happens at the best price limits.
  (c) Remaining negative variations are labeled as cancellations.

Doing so, we have an average matching rate of around 85% for CAC 40 stocks. As a byproduct, one gets the sign of each matched trade, that is, whether it is buyer or seller initiated.

2.3. Parameters Estimation. If T be the trading duration of interest each day ($T = 4.5$ hours—[9 : 30–14 : 00]—in our case.) Then

$$\widehat{\lambda^M} := \frac{\#\text{trades}}{2T}, \tag{2.1}$$

and

$$\widehat{\lambda_i^L} := \frac{1}{2T}\cdot \left(\#\text{buy limit orders arriving } i \text{ tick away from the best opposite quote} + \#\text{sell lim. orders etc.}\right). \tag{2.2}$$

For cancellations, we need to normalize the count by the average number of shares $\langle \mathbf{X}_i \rangle$ at distance i from the best opposite quote:

$$\widehat{\lambda_i^C} := \frac{1}{\langle \mathbf{X}_i \rangle}\frac{1}{2T}\cdot \left(\#\text{cancellation orders in the bid side arriving } i \text{ tick away from the best opposite quote} + \#\text{cancellation orders in the ask side etc.}\right), \tag{2.3}$$

[9]We set $\Delta t = 3$ s for CAC 40 stocks. We found that the median reporting delay for trades is −900 ms: on average, trades are reported 900 milliseconds *before* the change is recorded in tick by tick data.

K	30
a_∞	250
b_∞	250
(v^M, s^M)	$(4.00, 1.19)$
(v^L, s^L)	$(4.47, 0.83)$
(v^C, s^C)	$(4.48, 0.82)$
$\lambda^{M^\pm}$	0.1237

TABLE 3. Model parameters for the stock SCHN.PA (Schneider Electric) in March 2011 (23 trading days). Figures 8 and 9 are graphical representation of these parameters.

We then average $\widehat{\lambda^M}$, $\widehat{\lambda_i^L}$ and $\widehat{\lambda_i^L}$ across 23 trading days to get the final estimates. As for the volumes, we estimate by maximum likelihood the parameters $(\widehat{v}, \widehat{s})$ of a lognormal distribution separately for each order type. We depict the parameters in figures 8 and 9.

2.4. A Typical Parmaters Set.

$$\begin{aligned} \lambda^M &= 0.1, && \text{(order / second)} \\ \lambda^L &= 1, && \text{(order / second / tick)} \\ \lambda^C &= 0.2, && \text{(order / second / tick)} \\ K &= 10, && \text{(ticks)} \\ q &= 1, && \text{(share)} \\ a_\infty = b_\infty &= 5. && \text{(shares)} \end{aligned} \tag{2.4}$$

3. Results for CAC 40 Stocks

In order to get a cross-sectional view of the performance of the model on all CAC 40 stocks, we estimate the parameters separately for each stock and run a 100, 000 event simulation for each parameter set. We then compare in figure 10 the average depth, average spread and the long-term "volatility" measured directly from the data, to those obtained from the simulations. Dashed line is the identity function—It would correspond to a perfect match between model predictions and the data. Solid line is a linear regression $z_{\text{data}} = b_1 + b_2\, z_{\text{model}}$ for each quantity of interest z.

i (ticks)	$\langle \mathbf{X}_i \rangle$ (shares)	$\lambda_i^{L^\pm}$	$10^3.\lambda_i^{C^\pm}$
1	276	0.2842	0.8636
2	1129	0.5255	0.4635
3	1896	0.2971	0.1487
4	1924	0.2307	0.1096
5	1951	0.0826	0.0402
6	1966	0.0682	0.0341
7	1873	0.0631	0.0311
8	1786	0.0481	0.0237
9	1752	0.0462	0.0233
10	1691	0.0321	0.0178
11	1558	0.0178	0.0127
12	1435	0.0015	0.0012
13	1338	0.0001	0.0001
14	1238	0.0	0.0
15	1122	$\vdots$	$\vdots$
16	1036		
17	943		
18	850		
19	796		
20	716		
21	667		
22	621		
23	560		
24	490		
25	443		
26	400		
27	357		
28	317		
29	285	$\vdots$	$\vdots$
30	249	0.0	0.0

TABLE 4. Model parameters for the stock SCHN.PA (Schneider Electric) in March 2011 (23 trading days). Figures 8 and 9 are graphical representation of these parameters.

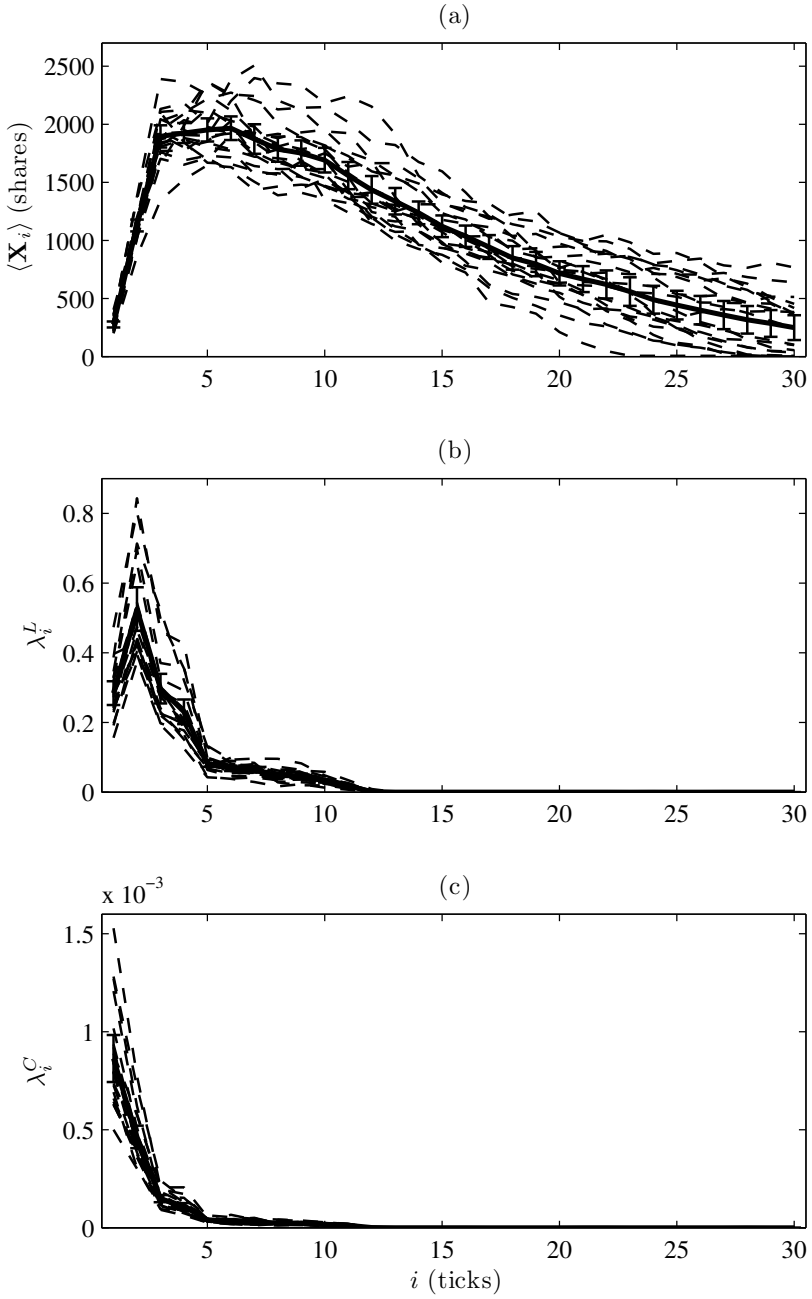


FIGURE 8. Model parameters: arrival rates and average depth profile (parameters as in table 4). Error bars indicate variability across different trading days.

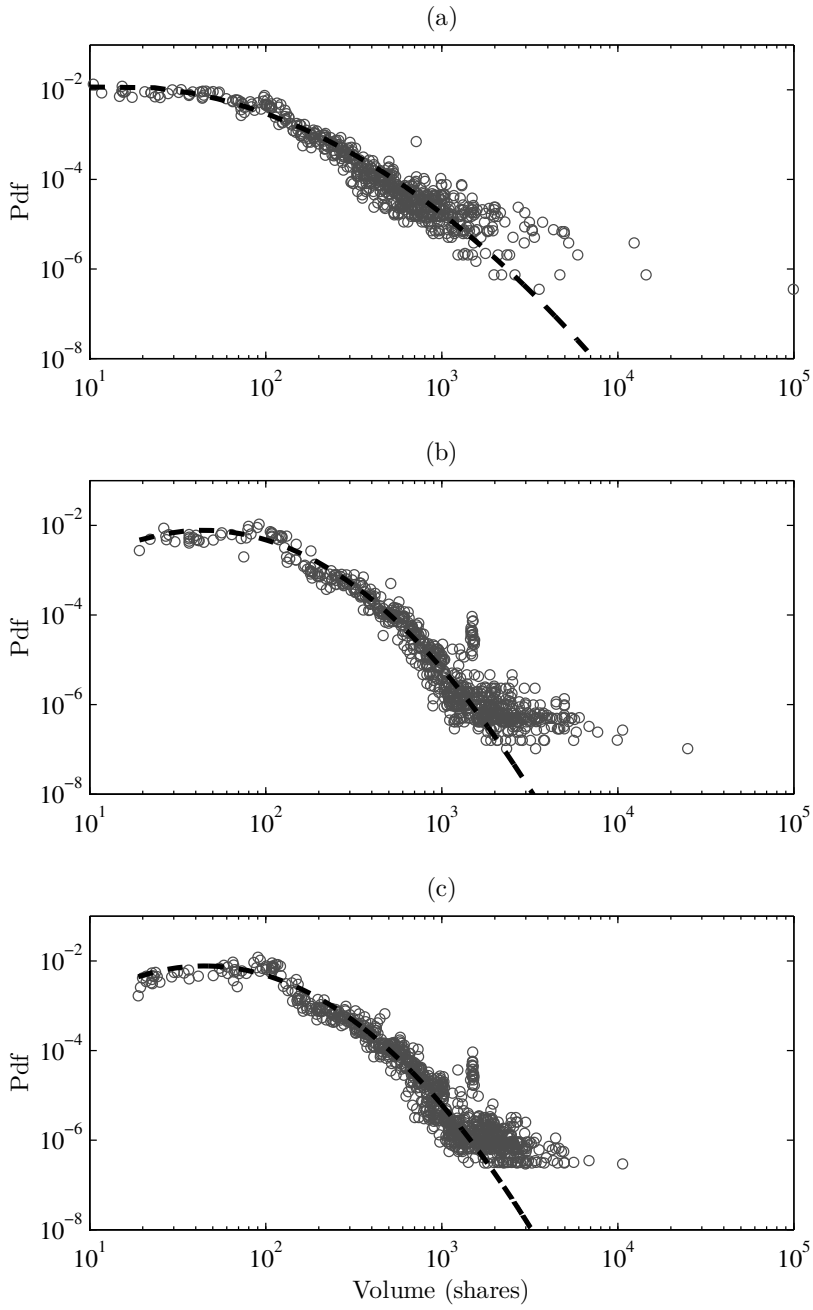


FIGURE 9. Model parameters: volume distribution. Panels $(a), (b)$ and (c) correspond respectively to market, limit and cancellation orders volumes. Dashed lines are lognormal fits (parameters as in table 3).

	b_1	b_2	R^2
$Log\,\langle A\rangle\,(5)$	$-0.42\,(\pm 0.11)$	$1.13\,(\pm 0.04)$	0.99
$Log\,\langle S\rangle$	$0.20\,(\pm 0.06)$	$1.16\,(\pm 0.07)$	0.97
σ_∞	$-0.012\,(\pm 0.05)$	$1.35\,(\pm 0.11)$	0.94

TABLE 5. CAC 40 stocks regression results.

Note that despite the good agreement between the average depth profiles (panel (a)), and although the model successfully predicts the relative magnitudes of the long-term variance σ_∞^2 and the average spread $\langle S\rangle$ for different stocks, it tends to systematically underestimate σ_∞^2 and $\langle S\rangle$. This may be related to the absence of autocorrelation in order signs in the model and the presence of more drifting phases in actual prices than in those obtained by simulation.

4. Conclusions

In the previous two chapters, we analyzed a simple Markovian order book model, in which elementary changes in the price and spread processes are explicitly linked to the instantaneous shape of the order book and the order flow parameters.

Two basic properties were investigated: the ergodicity of the order book and the large-scale limit of the price process. The first property, which we answered positively, is desirable in that it assures the stability of the order book in the long run, and gives a theoretical underpinning to statistical measurements on order book data. The scaling limit of the price process is, as anticipated, a Brownian motion. A key ingredient in this result is the convergence of the order book to its stationary state at an exponential rate, a property equivalent to a geometric mixing condition satisfied by the stationary version of the order book. This short memory effect, plus a constraint on the variance of price increments guarantee a diffusive limit at large time scales. Our assumptions are independent Poissonian order flows, proportional cancellation rates, and the presence of two reservoirs of liquidity K ticks away from the best quotes to guarantee that the spread does not diverge.[10]

[10]We believe this assumption can be relaxed under a balance condition on the arrival rates. One has however to consider an order book model with finite but *unbounded support*, and control not only the stability of the spread but also of all the gaps in the book.

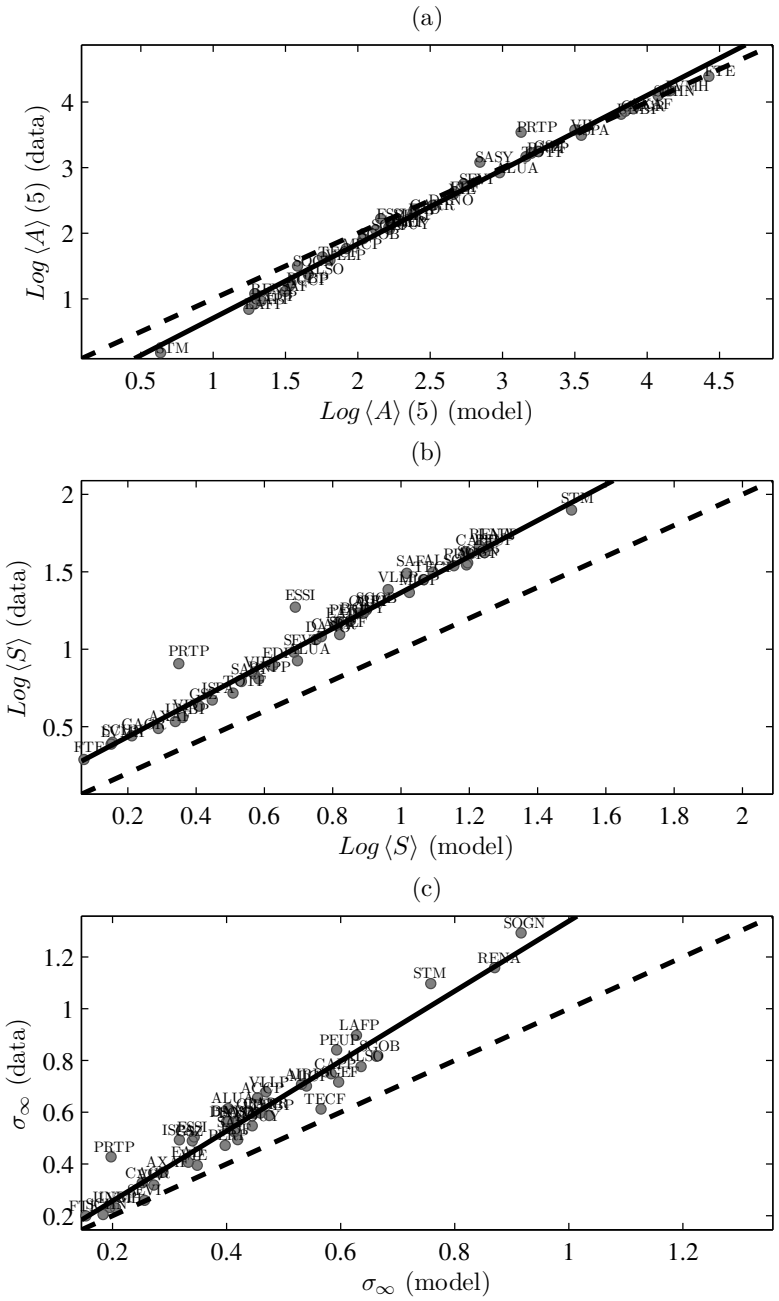


FIGURE 10. A cross-sectional comparison of liquidity and price diffusion characteristics between the model and data for CAC 40 stocks (March 2011).

We believe the results hold for a wide class of Markovian order book models: In general, one can state that price increments in a *stable* Markovian order book model are aggregationally Gaussian[11].

In a sense, this could offer a mathematical justification to the Bachelier model of asset prices, from a market microstructure perspective. In reality, the picture is however more subtle: even if the price process is asymptotically diffusive, at short time scales, the model produces stronger anti-correlation in traded prices than what is actually observed in the data. At those time scales, price diffusivity is arguably the result of a balance between persistent liquidity taking and anti-persistent liquidity providing.

We believe however that the approach presented here is interesting for clearly identifying conditions under which the asymptotic normality of price increments holds; and more importantly, for introducing a set of mathematical tools for further investigating the price dynamics in more sophisticated stochastic order book models. Indeed, using the same techniques, we are studying extensions of our results to the case of mutually exciting—and therefore dependent—order flows (point 1 below). This will be published elsewhere.

Our work can naturally be extended in several ways. In the following lines, we suggest some possible avenues to explore.

Firstly, actual order flows exhibit non-negligible *cross dependences*. As documented in [**MT11b**], market orders excite limit orders and vice versa. A possible solution for endogenously incorporating these dependences is the use of mutually exciting processes:

$$\begin{aligned} \lambda^M(t) &= \lambda^M(0) + \int_0^t \varphi^{MM}(t-s)dN^M(s) \\ &+ \int_0^t \varphi^{LM}(t-s)dN^L(s), \end{aligned} \tag{4.1}$$

and,

$$\begin{aligned} \lambda^L(t) &= \lambda^L(0) + \int_0^t \varphi^{LL}(t-s)dN^L(s) \\ &+ \int_0^t \varphi^{ML}(t-s)dN^M(s), \end{aligned} \tag{4.2}$$

[11]Rigorously, the convergence to the stationary state has to happen *fast enough*. That is, with an integrable convergence rate $\rho(n)$ as in (4.19).

This model has the additional advantage of capturing *clustering* in order arrivals (due to the self-excitation terms φ^{MM} and φ^{LL}), and for exponentially decaying kernels[12] can be cast into a Markovian setting.

Besides, *long-ranged correlation in order signs* is a very important feature of the data, as discussed in section 1. Analyzing this mathematically is more difficult since the model is no longer Markovian.

Moreover, it is natural to add another source of randomness on the rates themselves, for instance

$$d\lambda(t) = \theta(\overline{\lambda}(t) - \lambda(t))dt + \nu\sqrt{\lambda(t)}dW(t), \tag{4.3}$$

where $\overline{\lambda}$ is a (deterministic) background intensity to account for the U-shaped daily trading activity and θ, ν are the parameters of a CIR process. Such *stochastic arrival rates* would lead to stochastic volatility in the prices.

Although we argued that the simple Markovian order book model we study is stable and asymptotically diffusive, markets do show signs of fragility quite often and large jumps do occur in actual prices. Understanding how these macroscopic *jumps* (or departure from equilibrium) arise from events at the order book level, for instance via sudden evaporation of liquidity in one side of the book is much needed.

Finally, richer price dynamics (e.g. fat-tailed return distributions) can be obtained using *feedback loops* between the arrival rates and the price (or its volatility) as in [**PGPS06**].

These extensions may, however, render the model less amenable to mathematical analysis, and we leave the investigation of such interesting (but sometime difficult) questions for future research.

[12] $\varphi(u) = \alpha e^{-\beta u}$.

CHAPTER 4

On the Stability and Price Scaling Limit of a Hawkes Process-Based Order Book Model

1. Introduction

Since their introduction in [**Haw71b**], Hawkes processes have been applied in a wide range of research areas from seismology (originally), to credit risk [**EGG10**], financial contagion [**ASCDL13**] and more recently market microstructure modelling [**BDHM13a, BDHM13b, BM13, Bow07, Lar07, MT11b, MTP12, ZRA13**].

In market microstructure, and particularly order book modelling, the relevance of these processes comes at least from two empirical properties of (market, limit and cancellation) order flows at the microscopic level:

(1) Time clustering: order arrivals are highly clustered in time.
(2) Mutual dependence: order flow exhibit non-negligible cross dependences. For instance, as documented in [**MT11b**], market orders excite limit orders and *vice versa*.

At the microscopic level, point process-based microstructure models capture by construction the intrinsic discreteness of prices and volumes. A question of interest in this context is the microscopic to macroscopic transition in the price dynamics. This strand of research has attracted a lot of interest of late [**AJ13, BDHM13a, BDHM13b, BM13, CdL13, CdL12, HP13, ZRA13**].

In this chapter, we cast a Hawkes process-based order book model into a Markovian setting, and using techniques from the theory of Markov chains and stochastic stability [**MT09**], show that the order book is stable and leads to a diffusive price limit at large time scales.

1.1. Outline. Section 2 is a distillation of some mathematical results about Hawkes processes and Markov chains stochastic stability. Section 3

contains three auxiliary stability results which, apart from their own interest, are useful to prove the stability of the order book. Section 4 is an application to a particular order book model and is the main contribution of the chapter.

1.2. Notations. The following notations appear frequently throughout this chapter, and we recall them here for reference:

- (X_n): discrete-time process,
- $(X(t))$: continuous-time process,
- $|x| = \sum_{i=1}^{p} |x_i|$,
- $[\![1, p]\!] = \{1, 2, \dots p\}$.

2. Preliminary Remarks

We collect in this section several definitions and results that are useful for the rest of this chapter. The presentation is rather informal.

2.1. Point Processes.

Definition 2.1 (Point process). *A* point process *is an increasing sequence* $(T_n)_{n\in\mathbb{N}}$ *of positive random variables defined on a measurable space* $(\Omega, \mathcal{F}, \mathbb{P})$.

We will restrict our attention to processes that are *nonexplosive*, that is, for which $\lim_{n\to\infty} T_n = \infty$. To each realization (T_n) corresponds a counting function $(N(t))_{t\in\mathbb{R}^+}$ defined by

$$N(t) = n \text{ if } t \in [T_n, T_{n+1}[, n \geq 0. \tag{2.1}$$

$(N(t))$ is a right continuous step function with jumps of size 1 and carries the same information as the sequence (T_n), so that $(N(t))$ is also called a point process.

Definition 2.2 (Multivariate point process). *A* multivariate point process *(or marked point process) is a point process* (T_n) *for which a random variable* X_n *is associated to each* T_n*. The variables* X_n *take their values in a measurable space* $(E, \mathcal{E})$.

We will restrict our attention to the case where $E = \{1, \dots, M\}$, $M \in \mathbb{N}^*$. For each $m \in \{1, \dots, M\}$, we can define the counting processes

$$N_m(t) = \sum_{n\geq 1} \mathbb{I}(T_n \leq t)\mathbb{I}(X_n = i). \tag{2.2}$$

We also call the process

$$N(t) = (N_1(t), \ldots, N_M(t))$$

a multivariate point process.

Definition 2.3 (Intensity of a point process). *A point process* $(N(t))_{t\in\mathbb{R}^+}$ *can be completely characterized by its (conditional) intensity function,* $\lambda(t)$*, defined as*

$$\lambda(t) = \lim_{u\to 0} \frac{\mathbb{P}\left[N(t+u) - N(u) = 1|\mathcal{F}_t\right]}{u}, \tag{2.3}$$

where $\mathcal{F}_t$ *is the history of the process up to time t, that is, the specification of all points in* $[0, t]$*. Intuitively*

$$\mathbb{P}\left[N(t+u) - N(u) = 1|\mathcal{F}_t\right] = \lambda(t)\, u + o(u), \tag{2.4}$$

$$\mathbb{P}\left[N(t+u) - N(u) = 0|\mathcal{F}_t\right] = 1 - \lambda(t)u + o(u), \tag{2.5}$$

$$\mathbb{P}\left[N(t+u) - N(u) > 1|\mathcal{F}_t\right] = o(u). \tag{2.6}$$

This is naturally extended to the multivariate case by setting for each $m \in \{1, \ldots, M\}$

$$\lambda_m(t) = \lim_{u\to 0} \frac{\mathbb{P}\left[N_m(t+u) - N_m(u) = 1|\mathcal{F}_t\right]}{u}. \tag{2.7}$$

2.2. Hawkes Processes.

2.2.1. *Hawkes Process.*

Definition 2.4. *A Hawkes process* $(N(t))_{t\in\mathbb{R}_+}$ *is a point process whose intensity is specified by*

$$\lambda(t) = \mu + \alpha \int_0^t e^{-\beta(t-s)} dN(s) = \mu + \alpha \sum_{0\le s_i\le t} e^{-\beta(t-s_i)}, \tag{2.8}$$

for a triplet (μ, α, β) *of positive real numbers*[1].

The process thus defined is *self-excited*: it has a base intensity μ, plus exponentially decaying shocks due to previous jumps. The parameter α characterizes the scale of the excitation and β its decay in time.

Proposition 2.1. *The process* $X(t) = (N(t), \lambda(t))$ *is Markov.*

[1] A more general definition would have

$$\lambda(t) = \mu + \int_0^t \varphi(t-s) ds, \tag{2.9}$$

with an unspecified kernel $\varphi > 0$. But we only consider exponentially decaying kernels in this chapter.

Proof. From a straightforward calculation, we have for any $t_2 > t_1$

$$\begin{aligned}\lambda(t_2) &= \mu + \alpha \int_0^{t_2} e^{-\beta(t_2-s)} dN(s) \\ &= \mu + \alpha \int_0^{t_1} e^{-\beta(t_2-s)} dN(s) + \alpha \int_{t_1}^{t_2} e^{-\beta(t_2-s)} dN(s) \\ &= \mu + e^{-\beta(t_2-t_1)}(\lambda(t_1) - \mu) + \int_{t_1}^{t_2} e^{-\beta(t_2-s)} dN(s). \qquad (2.10)\end{aligned}$$

So that in order to compute $\lambda(t_2)$, we only need to know $\lambda(t_1)$ and $\{N(t) : t_1 \leq t \leq t_2\}$—the information contained in $\{N(t), \lambda(t) : 0 \leq t < t_1\}$ is irrelevant. Hence

$$\mathbb{P}\left[(N(t_2), \lambda(t_2)) \in A | \{N(t), \lambda(t) : t \in [0, t_1]\}\right] = \mathbb{P}\left[(N(t_2), \lambda(t_2)) \in A | N(t_1), \lambda(t_1)\right], \qquad (2.11)$$

for any measurable set $A \subset \mathbb{N} \times \mathbb{R}^+$, and X is Markov. □

2.2.2. *Multivariate Hawkes Process.*

Definition 2.5. *We say that $N = (N_1, \ldots, N_M)$ is a multivariate Hawkes process when*

$$\lambda_m(t) = \mu_m + \sum_{j=1}^{M} \alpha_{mj} \int_0^t e^{-\beta_{mj}(t-s)} dN_j(s). \qquad (2.12)$$

Proposition 2.2. *Let $Y^{ij}(t) = \alpha_{ij} \int_0^t e^{-\beta_{ij}(t-s)} dN_j(s)$, $1 \leq i, j \leq M$, and $Y(t) = \{Y^{ij}(t)\}_{1 \leq i,j \leq M}$. The process $X(t) = (N(t), Y(t))$ is Markov.*

Proof. Let $t_2 > t_1$. Since

$$Y^{mj}(t_2) = e^{-\beta_{mj}(t_2-t_1)} Y^{mj}(t_1) + \int_{t_1}^{t_2} e^{-\beta_{mj}(t_2-s)} dN_j(s), \qquad (2.13)$$

and

$$\lambda_m(t_2) = \mu_m + \sum_{j=1}^{M} Y^{mj}(t_2), \qquad (2.14)$$

the law of $(N(t_2), Y(t_2))$ conditional on $\{(N(t), Y(t)) : 0 \leq t \leq t_1\}$ is the same as when conditionning on $(N(t_1), Y(t_1))$ only—the information contained in $\{(N(t), Y(t)) : 0 \leq t < t_1\}$ is irrelevant, and X is Markov. □

2.2.3. *Stationarity.*

Definition 2.6. *A point process is stationary when for every $r \in \mathbb{N}^*$ and all bounded Borel subsets $A_1, \ldots, A_r$ of the real line, the joint distribution of*

$$\{N(A_1 + t), \ldots, N(A_r + t)\}$$

does not depend on t.

In [**HO74**], Hawkes and Oakes show that:

Proposition 2.3. *If*

$$\frac{\alpha}{\beta} < 1 \tag{2.15}$$

then there exists a (unique) stationary point process $(N(t))$, whose intensity is specified as in definition 2.4.

Brémaud and Massoulié generalize this to the multivariate case in [**BM96**]:

Proposition 2.4. *Let the matrix A be defined by*

$$A_{ij} = \frac{\alpha_{ij}}{\beta_{ij}}, \qquad 1 \leq i, j \leq M. \tag{2.16}$$

If

$$\rho(A) < 1$$

then there exists a (unique) stationary multivariate point process $N(t) = (N_1(t), \ldots, N_m(t))$ whose intensity is specified as in definition 2.5.
$\rho(A)$ is the spectral radius of the matrix A, that is, its largest eigenvalue.

2.3. The Embedded Discrete-time Hawkes Process. Throughout this chapter, we will mostly work with processes sampled in discrete time. We show in this section how to construct a discrete-time version $(X_n)_{n\in\mathbb{N}}$ out of a multivariate Hawkes process $X(t) = (N(t), Y(t))_{t\in\mathbb{R}^+}$, where Y is defined by

$$Y^{ij}(t) = \alpha_{ij} \int_0^t e^{-\beta_{ij}(t-s)} dN_j(s), \qquad 1 \leq i, j \leq M, \tag{2.17}$$

as in proposition 2.2.

First denote $(T_n)_{n\geq 1}$ the jump times of the process (and set $T_0 = 0$), and

$$X_n = X(T_n) = (N(T_n), Y(T_n)). \tag{2.18}$$

We define $E_n = E(T_n) \in \{1, \dots, M\}$ as the mark of the process. The value of E_n indicates which component of $N(t)$ has jumped at time T_n. We also define the waiting times $(\tau_n)_{n\geq 1}$ between two successive jumps as

$$\tau_n = T_{n+1} - T_n. \tag{2.19}$$

Given that $(N_n, Y_n) = (\xi, y)$, (N_{n+1}, Y_{n+1}) is generated as follows: Set

$$\tau_{n+1} = \min(\tau^1_{n+1}, \dots, \tau^M_{n+1}), \tag{2.20}$$

where conditional on $(N_n, Y_n) = (\xi, y)$, the distribution of $\tau^1_{n+1}, \dots, \tau^M_{n+1}$ is that of independent positive random variables whose marginal distributions are determined by hazard rates

$$h_m(t) := \mu_m + \sum_{j=1}^{M} y_{mj} e^{-\beta_{ij} t}, \qquad t \geq 0, 1 \leq m \leq M. \tag{2.21}$$

Then set

$$E_{n+1} = argmin_{1\leq m\leq M} \tau^m_{n+1}, \tag{2.22}$$

$$N_{n+1} = (\xi_1, \dots, \xi_{E_n} + 1, \dots, \xi_M), \tag{2.23}$$

and

$$Y^{mj}_{n+1} = y_{mj} e^{-\beta_{mj}\tau_{n+1}} + \alpha_{mj} \mathbb{I}(E_{n+1} = j). \tag{2.24}$$

2.4. Drift of a Discrte-time Markov Process.

Definition 2.7. *The drift operator $\mathcal{D}$ is defined to act on any nonnegative measurable function V by*

$$\mathcal{D}V(x) = \mathbb{E}[V(X_{n+1}) - V(X_n)|X_n = x]. \tag{2.25}$$

We will also use the notation

$$\mathcal{P}V(x) = \mathbb{E}[V(X_{n+1})|X_n = x], \tag{2.26}$$

hence

$$\mathcal{D}V(x) = \mathcal{P}V(x) - V(x). \tag{2.27}$$

As will be clear in the next section, the importance of this operator stems from the existence of criteria based on the drift to establish properties of the process. It can be interpreted as the analogous for a process to the derivative for a function[2].

[2]Cf. Dynkyn's formula or its discrete-time formulation.

2.5. A Digression on Stochastic Stability. Let $(X_n)_{n\in\mathbb{N}}$ be a Markov process on a sate space $\mathcal{S}$ and $(Q^n)_{n\in\mathbb{N}^*}$ its transition probability function, that is

$$Q^n(x, A) = \mathbb{P}\left[X_n \in A | X_0 = x\right], \tag{2.28}$$

for $x \in \mathcal{S}$ and A a measurable subset of $\mathcal{S}$.

2.5.1. *Ergodicity of a Markov process.* Ergodicity is a strong form of "stability": To rephrase [**MT09**], it means that *"there is an invariant regime described by a measure π such that if the process starts in this regime (that is, if X_0 has distribution π) then it remains in the regime. And moreover if the process starts in some other regime, then it converges in a strong probabilistic sense with π as a limiting distribution."*

Formally, a (aperiodic, irreducible) Markov process is *ergodic* if an invariant[3] probability measure π exists and

$$\lim_{n\to\infty} \|Q^n(x, .) - \pi(.)\| = 0, \forall x \in \mathcal{S}, \tag{2.30}$$

where $\|.\|$ designates for a signed measure ν the total variation norm[4] defined as

$$\|\nu\| := \sup_{f:|f|<1} |\nu(f)| = \sup_{A\in\mathcal{B}(\mathcal{S})} \nu(A) - \inf_{A\in\mathcal{B}(\mathcal{S})} \nu(A). \tag{2.32}$$

In (2.32), $\mathcal{B}(\mathcal{S})$ is the Borel σ-field generated by $\mathcal{S}$, and for a measurable function f on $\mathcal{S}$, $\nu(f) := \int_{\mathcal{S}} f d\nu$.

2.5.2. *V-uniform ergodicity.* We say that a Markov process is *V-uniformly ergodic* if there exists a coercive[5] function $V > 1$, an invariant distribution π, and constants $0 < r < 1$, and $R < \infty$ such that

$$\|Q^n(x, .) - \pi(.)\| \leq Rr^n V(x), x \in \mathcal{S}. \tag{2.33}$$

[3]That is, satisfying the invariance equations

$$\pi(A) = \int_{\mathcal{S}} \pi(dx) Q(x, A), \qquad A \in \mathcal{B}(\mathcal{S}). \tag{2.29}$$

[4]If the state space $\mathcal{S}$ is countable (this is *not* the case for $(X(t), Y(t))$ of proposition 2.2.), the convergence in total variation norm implies the more familiar pointwise convergence

$$\lim_{n\to\infty} |Q^n(x, y) - \pi(y)| = 0, \forall x, y \in \mathcal{S}. \tag{2.31}$$

[5]That is, a function such that $V(x) \to \infty$ as $|x| \to \infty$. The condition $V > 1$ is of course arbitrary and 1 can be replaced by any positive constant.

This is a strong form of ergodicity (note the geometric rate of convergence), and it can be characterized in terms of the drift operator $\mathcal{D}$. Indeed, it is shown in [**MT09, MT92**] that it is equivalent to the existence of a coercive function V (the "Lyapunov test function") such that

$$\mathcal{D}V(x) \leq -K_1 V(x) + K_2 \mathbb{I}_C(x) \quad \text{(Geometric drift condition.)} \tag{2.34}$$

for some positive constants K_1 and K_2, and $C \subset \mathcal{S}$ a compact set. (Theorem 16.0.1 in [**MT09**].) Condition (2.34) is equivalent to

$$\mathcal{P}V(x) \leq \theta V(x) + K_3 \mathbb{I}_C(x) \tag{2.35}$$

for some $0 < \theta < 1$. Intuitively, it says that the larger $V(X_n)$ the stronger X is pulled back towards the center of the state space $\mathcal{S}$.

Interestingly, it is possible to develop central limit theorems for functionals of V-uniformly ergodic Markov processes. This will be used to show that the price process in a stable Hawkes process-based order book model is asymptotically diffusive. Before that, we need the following auxiliary results.

3. Auxiliary Results

3.1. $V-$uniform Ergodicity of the Intensity of a Hawkes Process. Let $(N(t), \lambda(t))_{t\in\mathbb{R}^+}$ be a Hawkes process with parameters (μ, α, β), and $(N_n, \lambda_n)_{n\in\mathbb{N}}$ its embedded discrete-time process as constructed in section 2.3.

Proposition 3.1. *If $\alpha < \beta$, then the process $(\lambda_n)_{n\in\mathbb{N}}$ is $V-$uniformly ergodic, with*

$$V(\lambda) = e^{\gamma\lambda}, \tag{3.1}$$

and γ a suitably chosen positive number.

Proof. If $\tau_n = T_{n+1} - T_n$ be the waiting time between two successive jumps of $(X(t))$. There holds for $t' \in [T_n, T_{n+1}[$,

$$\lambda(t') = \lambda_n + (\lambda_n - \mu)e^{-\beta(t'-T_n)}. \tag{3.2}$$

The hazard rate associated to τ_n, conditional on $\lambda_n = \lambda \in \mathbb{R}^+$, is

$$h(t) := \mu + (\lambda - \mu)e^{-\beta t}, \tag{3.3}$$

and the p.d.f. of τ_n is

$$f(t) = h(t)e^{-\int_0^t h(s)ds} = \left(\mu + (\lambda - \mu)e^{-\beta t}\right) e^{-\mu t - \frac{\lambda-\mu}{\beta}(1-e^{-\beta t})}. \tag{3.4}$$

Let

$$V(\lambda) := e^{\gamma\lambda} \tag{3.5}$$

be a Lyapunov test function with $\gamma > 0$ an arbitrary parameter. Then

$$\begin{aligned}
\mathbb{E}\left[V\left(\lambda_{n+1}\right)|\lambda_n = \lambda\right] &= \int_0^\infty V(\lambda(t^+))f(t)dt \\
&= \int_0^\infty V\left(\mu + (\lambda-\mu)e^{-\beta t} + \alpha\right) \\
&\times \left(\mu + (\lambda-\mu)e^{-\beta t}\right) e^{-\mu t - \frac{\lambda-\mu}{\beta}(1-e^{-\beta t})} dt \\
&= \int_0^\infty e^{\gamma\left(\mu+(\lambda-\mu)e^{-\beta t}+\alpha\right)} \\
&\times \left(\mu + (\lambda-\mu)e^{-\beta t}\right) e^{-\mu t - \frac{\lambda-\mu}{\beta}(1-e^{-\beta t})} dt
\end{aligned} \tag{3.6}$$

Hence

$$\begin{aligned}
\frac{\mathcal{P}V(\lambda)}{V(\lambda)} &= e^{-\gamma\lambda}\,\mathbb{E}\left[V\left(\lambda_{n+1}\right)|\lambda_n = \lambda\right] \\
&= \int_0^\infty e^{-\gamma(\lambda-\mu)(1-e^{-\beta t})+\gamma\alpha}\left(\mu + (\lambda-\mu)e^{-\beta t}\right) e^{-\mu t - \frac{\lambda-\mu}{\beta}(1-e^{-\beta t})} dt \\
&= e^{\gamma\alpha}\mu \int_0^\infty e^{-\gamma(1+\frac{1}{\beta})(\lambda-\mu)(1-e^{-\beta t})-\mu t} dt \\
&+ e^{\gamma\alpha}(\lambda-\mu)\int_0^\infty e^{-\gamma(1+\frac{1}{\beta})(\lambda-\mu)(1-e^{-\beta t})-(\beta+\mu)t} dt.
\end{aligned} \tag{3.7}$$

Using lemma 5.1, we get

$$\begin{aligned}
\frac{\mathcal{P}V(\lambda)}{V(\lambda)} &= e^{\gamma\alpha}\mu\, \mathcal{I}\left((\gamma + \frac{1}{\beta})(\lambda-\mu), \beta, \mu\right) \\
&+ e^{\gamma\alpha}(\lambda-\mu)\, \mathcal{I}\left((\gamma + \frac{1}{\beta})(\lambda-\mu), \beta, \beta+\mu\right),
\end{aligned} \tag{3.8}$$

where

$$\mathcal{I}(a,b,c) := \int_0^\infty e^{-a(1-e^{-bt})-ct} dt. \tag{3.9}$$

Then

$$\begin{aligned}
\lim_{\lambda\to\infty} \frac{\mathcal{P}V(\lambda)}{V(\lambda)} &= 0 + \frac{e^{\gamma\alpha}}{\beta(\gamma+\frac{1}{\beta})} \\
&= \frac{e^{\gamma\alpha}}{1+\gamma\beta}.
\end{aligned} \tag{3.10}$$

And

$$\begin{aligned}\lim_{\lambda\to\infty}\frac{\mathcal{D}V(\lambda)}{V(\lambda)} &= \lim_{\lambda\to\infty}\frac{\mathcal{P}V(\lambda)}{V(\lambda)}-1\\ &= \frac{e^{\gamma\alpha}}{1+\gamma\beta}-1\\ &= \frac{e^{\gamma\alpha}-1-\gamma\beta}{1+\gamma\beta}.\end{aligned} \tag{3.11}$$

A Taylor expansion in γ around 0 yields

$$\lim_{\lambda\to\infty}\frac{\mathcal{D}V(\lambda)}{V(\lambda)} = \gamma(\alpha-\beta)+o(\gamma), \tag{3.12}$$

which has the sign of $\alpha-\beta$. Finally, if $\alpha<\beta$, one can choose $\gamma>0$, $\lambda_0\in\mathbb{R}_+^*$ and $\kappa>0$ such that $\forall\lambda>\lambda_0$

$$\mathcal{D}V(\lambda)\le -\kappa V(\lambda), \tag{3.13}$$

and the V-uniform ergodicity of (λ_n) follows. □

3.2. $V-$uniform Ergodicity of the Intensity of a Multivariate Hawkes Process. Consider now a multivariate setting. Let $X(t)=(N(t),Y(t))$ be a M-variate Hawkes process with parameters

$$\mu=(\mu_1,\dots,\mu_M)^t, \tag{3.14}$$

$$\alpha=(\alpha_{ij})_{1\le i,j\le M}, \tag{3.15}$$

and

$$\beta=(\beta_{ij})_{1\le i,j\le M}. \tag{3.16}$$

Define also

$$\alpha_{\max}=\max\{\alpha_{ij}\}_{1\le i,j\le M}\in\mathbb{R}_+, \tag{3.17}$$

and

$$\beta_{\min}=\min\{\beta_{ij}\}_{1\le i,j\le M}\in\mathbb{R}_+^*,\qquad \beta_{\max}=\max\{\beta_{ij}\}_{1\le i,j\le M}\in\overline{\mathbb{R}}_+^*. \tag{3.18}$$

We recall that $Y(t)=(Y^{ij})_{1\le i,j\le M}$ is defined by

$$Y^{ij}(t)=\alpha_{ij}\int_0^t e^{-\beta_{ij}s}dN(s). \tag{3.19}$$

As in the univariate case, let (N_n) and (Y_n) be the discrete time processes

$$\begin{aligned}N_n &= N(T_n^+),\\ \text{and } Y_n &= Y(T_n^+),\end{aligned} \tag{3.20}$$

sampled at the jump times (T_n) of (X). We have the following stability result for (Y_n).

PROPOSITION 3.2. *If*

$$\frac{M\alpha_{\max}}{\beta_{\min}} e^{1-\frac{M\alpha_{\max}}{\beta_{\max}}} < 1, \tag{3.21}$$

then the intensity of a multivariate Hawkes process is V-uniformly ergodic, with

$$V(y) = e^{\gamma \sum_{1\le k,l\le M} y^{kl}}, \tag{3.22}$$

and γ a suitably chosen positive number.

PROOF. As in the univariate case, let

$$V(y) = e^{\gamma \sum_{1\le k,l\le M} y^{kl}}. \tag{3.23}$$

Define the hazard rates

$$h_i(t) = \mu_i + \sum_{j=1}^{M} e^{-\beta_{ij}t} y_{ij}, \qquad 1 \le i \le M, \tag{3.24}$$

and

$$h(t) = \sum_{i=1}^{M} h_i(t). \tag{3.25}$$

We first note that, conditional on $\tau_{n+1} = t$, the probability that the next jump is on $N^i,\ i \in \{1,\dots,M\}$, is

$$\mathbb{P}\left[E_{n+1} = i | Y_n = y, \tau_{n+1} = t\right] = \frac{h_i(t)}{h(t)}. \tag{3.26}$$

We have then

$$\begin{aligned}
\mathbb{E}\left[V(Y_{n+1})|Y_n = y\right] &= \int_0^\infty \sum_{i=1}^{M} e^{\gamma \sum_{1\le k,l\le M}\left(e^{-\beta_{k,l}t} y_{kl} + \mathbb{I}(l=i)\alpha_{k,l}\right)} \frac{h_i(t)}{h(t)} \times h(t) e^{-\int_0^t h(s)ds} dt \\
&= \int_0^\infty \sum_{i=1}^{M} e^{\gamma \sum_{k=1}^{M} \alpha_{ki} + \gamma \sum_{1\le k,l\le M} e^{-\beta_{k,l}t} y_{kl}} \\
&\times \left(\mu_i + \sum_{j=1}^{M} y_{ij} e^{-\beta_{ij}t}\right) e^{-\sum_{k=1}^{M} \mu_k t - \sum_{1\le k,l\le M}(1-e^{-\beta_{kl}t})\frac{y_{kl}}{\beta_{kl}}} dt.
\end{aligned} \tag{3.27}$$

Dividing by $V(y)$ and rearranging the terms we get

$$\begin{aligned}\frac{\mathcal{P}V(y)}{V(y)} &= \int_0^\infty \sum_{i=1}^M e^{\gamma\sum_{k=1}^M \alpha_{ki}} e^{\sum_{1\le k,l\le M}(\gamma+\frac{1}{\beta_{kl}})(1-e^{1-\beta_{k,l}t})y_{kl}} \left(\mu_i + \sum_{j=1}^M y_{ij}e^{-\beta_{ij}t}\right) dt \\ &= \sum_{i=1}^M e^{\gamma\sum_{k=1}^M \alpha_{ki}} \mu_i \mathcal{I}_{M^2}\left(\left((\gamma+\frac{1}{\beta_{kl}})y_{kl}\right)_{1\le k,l\le M}; (\beta_{kl})_{1\le k,l\le M}; \sum_{k=1}^M \mu_k\right) \\ &+ \sum_{i=1}^M e^{\gamma\sum_{k=1}^M \alpha_{ki}} \sum_{j=1}^M y_{ij}\mathcal{I}_{M^2}\left(\left((\gamma+\frac{1}{\beta_{kl}})y_{kl}\right)_{1\le k,l\le n}; \left(\beta^{kl}\right)_{1\le k,l\le M}; \sum_{k=1}^n \mu_k + \beta_{ij}\right),\end{aligned} \tag{3.28}$$

where

$$\mathcal{I}_p(a_1,\dots,a_p;b_1,\dots,b_p;c) := \int_0^\infty e^{-a_1(1-e^{-b_1t})-\dots-a_p(1-e^{-b_pt})-ct}dt \tag{3.29}$$

is defined in lemma 5.2. The first term in the r.h.s of (3.28) vanishes when $|y|\to\infty$ by lemma 5.2. Again using lemma 5.2, as $|y|\to\infty$, $\forall 1\le i,j\le M$,

$$\begin{aligned}\mathcal{I}_{M^2}\left(\left((\gamma+\frac{1}{\beta_{kl}})y_{kl}\right)_{1\le k,l\le M}; (\beta_{kl})_{1\le k,l\le M}; \sum_{k=1}^M \mu_k + \beta_{ij}\right) &\le \frac{1}{\beta_{\min}\sum_{1\le k,l\le M}(\gamma+\frac{1}{\beta_{kl}})y_{kl}} \\ &\le \frac{1}{\beta_{\min}(\gamma+\frac{1}{\beta_{\max}})\sum_{1\le k,l\le M} y_{kl}}.\end{aligned} \tag{3.30}$$

Hence, the second term in the r.h.s of (3.28) is bounded by

$$\frac{\sum_{i=1}^M e^{\gamma\sum_{k=1}^M \alpha^{ki}}|y|}{\beta_{\min}(\gamma+\frac{1}{\beta_{\max}})|y|} \le \frac{e^{M\alpha_{\max}\gamma}}{\beta_{\min}(\gamma+\frac{1}{\beta_{\max}})}.$$

And for large $|y|$ we have

$$\frac{\mathcal{P}V(y)}{V(y)} \le \frac{e^{M\alpha_{\max}\gamma}}{\beta_{\min}(\gamma+\frac{1}{\beta_{\max}})}. \tag{3.31}$$

In order to conclude the proof, it is enough to show that there exists a suitably chosen $\gamma > 0$ such that

$$h(\gamma) = \frac{e^{M\alpha_{\max}\gamma}}{\beta_{\min}(\gamma+\frac{1}{\beta_{\max}})} < 1. \tag{3.32}$$

Minimizing h with respect to γ, the minimum is reached at

$$\gamma^* = \frac{1}{M\alpha_{\max}} - \frac{1}{\beta_{\max}} > 0. \tag{3.33}$$

and is equal to

$$h(\gamma^*) = \frac{M\alpha_{\max}}{\beta^{min}} e^{1-\frac{M\alpha_{\max}}{\beta_{\max}}}. \tag{3.34}$$

Note that for γ^* to be positive (and V to be coercive) we need

$$\alpha_{\max} \leq \frac{\beta_{\max}}{M}, \tag{3.35}$$

which we assume. Finally if

$$\frac{M\alpha_{\max}}{\beta^{min}} e^{1-\frac{M\alpha_{\max}}{\beta_{\max}}} < 1, \tag{3.36}$$

then $(Y_n)_{n\in\mathbb{N}}$ is $V-$uniformly ergodic. □

Remark 3.1. *Note that for $M = 1$ the condition is*

$$\frac{\alpha}{\beta} e^{1-\frac{\alpha}{\beta}} < 1, \tag{3.37}$$

which is satisfied i.i.f.

$$\frac{\alpha}{\beta} < 1. \tag{3.38}$$

($x \mapsto x(1-e^x)$ is strictly increasing from 0 to 1 on $[0,1]$). We get the result in the univariate case.

Remark 3.2. *A sufficient condition is*

$$\alpha_{\max} < \frac{\beta_{\min}}{M}. \tag{3.39}$$

Remark 3.3. *Stability condition* (3.36) *is not sharp: It is too stringent on the parameters $\alpha_{i,j}$ and $\beta_{i,j}$, and we suspect the stationarity condition of proposition 2.4 to be sufficient for V-uniform ergodicty.*

3.3. V-uniform Ergodicity of a "Birth-death" Hawkes Process. Let $(N_1(t), N_2(t))$ be a bivariate Hawkes process with intensities:

$$\lambda_1(t) = \mu_1 + \alpha_{11}\int_0^t e^{-\beta_{11}s} dN_1(s) + \alpha_{12}\int_0^t e^{-\beta_{12}s} dN_2(s), \tag{3.40}$$

$$\lambda_2(t) = \mu_2 + \alpha_{21}\int_0^t e^{-\beta_{21}s} dN_1(s) + \alpha_{22}\int_0^t e^{-\beta_{22}s} dN_2(s), \tag{3.41}$$

and define the queue $(X(t))$ by

- $X(t) \to X(t)+1$ when $N_1(t)$ jumps. This happens with (infinitesimal) probability $\lambda_1(t)dt$.
- $X(t) \to X(t)-1$ when $N_2(t)$ jumps and $X(t) \neq 0$. This happens with probability $\lambda_2(t)dt$.

- $X(t) \to X(t) - 1$ with probability $\lambda_3 X(t)dt$ for a constant $\lambda_3 > 0$. This corresponds to a proportional death rate, or in the context of order book modelling, to a proportional cancellation rate.

We also denote by $N_3(t)$ a counting process with intensity $\lambda_3 X(t)$ that jumps by 1 when $X(t)$ jumps by -1 due to a "cancellation".

The queue $X(t)$, albeit peculiar, is the building block of the order book model we present in the next section: N_1 represents the flow of limit orders, N_2 that of market orders and N_3 cancellations.

The following result is the key to the proof of the stability of the order book.

Proposition 3.3. *Provided $\beta_{\min}$ is large (specified precisely below), (X_n, Y_n) is V-uniformly ergodic, where*

$$V(x, y) = e^{\omega x + \gamma \sum_{1\le k,l\le 2} y_{kl}}, \tag{3.42}$$

and $\omega > 0$ and $\gamma > 0$.

Proof. As usual we write

$$\begin{aligned}
\frac{\mathcal{P}V(x,y)}{V(x,y)} &= e^{\omega+\gamma(\alpha_{11}+\alpha_{12})} \int_0^\infty e^{-\sum_{1\le k,l\le 2}(\gamma+\frac{1}{\beta_{kl}})(1-e^{-\beta_{kl}t})y_{kl}-(\mu_1+\mu_2)t-\lambda_3 xt}(\mu_1 + y_{11}e^{-\beta_{11}t} + y_{12}e^{-\beta_{12}t})dt \\
&+ e^{-\omega+\gamma(\alpha_{21}+\alpha_{22})} \int_0^\infty e^{-\sum_{1\le k,l\le 2}(\gamma+\frac{1}{\beta_{kl}})(1-e^{-\beta_{kl}t})y_{kl}-(\mu_1+\mu_2)t-\lambda_3 xt}(\mu_2 + y_{21}e^{-\beta_{21}t} + y_{22}e^{-\beta_{22}t})dt \\
&+ e^{-\omega} \int_0^\infty e^{-\sum_{1\le k,l\le 2}(\gamma+\frac{1}{\beta_{kl}})(1-e^{-\beta_{kl}t})y_{kl}-(\mu_1+\mu_2)t-\lambda_3 xt}\ \lambda_3 x\, dt \\
&= (e^{\omega+\gamma(\alpha_{11}+\alpha_{12})}\mu_1 + e^{\omega+\gamma(\alpha_{21}+\alpha_{22})}\mu_2 + e^{-\omega}\lambda_3 x) \\
&\times \int_0^\infty e^{-\sum_{1\le k,l\le 2}(\gamma+\frac{1}{\beta_{kl}})(1-e^{-\beta_{kl}t})y_{kl}-(\mu_1+\mu_2)t-\lambda_3 xt} dt \\
&+ e^{\omega+\gamma(\alpha_{11}+\alpha_{21})} y_{11} \int_0^\infty e^{-\sum_{1\le k,l\le 2}(\gamma+\frac{1}{\beta_{kl}})(1-e^{-\beta_{kl}t})y_{kl}-(\mu_1+\mu_2)t-\lambda_3 xt-\beta_{11}t} dt \\
&+ e^{\omega+\gamma(\alpha_{11}+\alpha_{21})} y_{12} \int_0^\infty e^{-\sum_{1\le k,l\le 2}(\gamma+\frac{1}{\beta_{kl}})(1-e^{-\beta_{kl}t})y_{kl}-(\mu_1+\mu_2)t-\lambda_3 xt-\beta_{12}t} dt \\
&+ e^{-\omega+\gamma(\alpha_{12}+\alpha_{22})} y_{21} \int_0^\infty e^{-\sum_{1\le k,l\le 2}(\gamma+\frac{1}{\beta_{kl}})(1-e^{-\beta_{kl}t})y_{kl}-(\mu_1+\mu_2)t-\lambda_3 xt-\beta_{21}t} dt \\
&+ e^{-\omega+\gamma(\alpha_{12}+\alpha_{22})} y_{22} \int_0^\infty e^{-\sum_{1\le k,l\le 2}(\gamma+\frac{1}{\beta_{kl}})(1-e^{-\beta_{kl}t})y_{kl}-(\mu_1+\mu_2)t-\lambda_3 xt-\beta_{22}t} dt.
\end{aligned} \tag{3.43}$$

As $|x| + |y| \to \infty$,

$$\frac{\mathcal{P}V(x,y)}{V(x,y)} \leq e^{-\omega} + \frac{2}{\beta_{\min}(\gamma + \frac{1}{\beta_{\max}})}(e^{\omega+\gamma(\alpha_{11}+\alpha_{21})} + e^{-\omega+\gamma(\alpha_{12}+\alpha_{22})}). \tag{3.44}$$

This quantity can be made smaller than 1 if $\beta_{\min}$ is large enough, hence the stated result. □

Remark 3.4. *Intuitively, a large β corresponds to a short memory for the process (X_n, Y_n).*

4. Application in Order Book Modelling

4.1. Model Setup. We present a stylized order book model whose dynamics is governed by Hawkes processes. We have already discussed at length a similar Poissonian order book model in [**AJ13**], so the description provided here is brief.

We shall assume that each side of the order book is fully described by a finite number of limits K, ranging from 1 to K ticks away from the best available opposite quote. We use the notation

$$X(t) = (a(t); b(t)) = (a_1(t), \ldots, a_K(t); b_1(t), \ldots, b_K(t)), \tag{4.1}$$

where $a = (a_1, \ldots, a_K)$ designates the ask side of the order book and a_i the number of shares available i ticks away from the best opposite quote, and $b = (b_1, \ldots, b_K)$ designates the bid side of the book.

3 types of events can happen:

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

Arrival of limit and market orders are described by 4 self and mutually exciting Hawkes processes:

- $L^{\pm}(t)$: arrival of a limit order, with intensity $\lambda^{L^{\pm}}(t)$;
- $M^{\pm}(t)$: arrival of new market order, with intensity $\lambda^{M^{\pm}}(t)$.

Cancellations are modelled by a (doubly stochastic) Poisson process whose intensity is proportional to the number of shares at each side of the order book, that is

$$\lambda^{C^{\pm}}|x^{\pm}|. \tag{4.2}$$

We denote by q the size of any new incoming order, and the superscript "+" (respectively "−") refers to the ask (respectively bid) side of the book. Buy limit orders $L^-(t)$ arrive below the ask price $P^A(t)$, and sell limit orders $L^+(t)$ arrive above the bid price $P^B(t)$.

Once a limit order arrives, its position is chosen randomly from 1 to K. Similarly once a cancellation order arrives, the order to be cancelled is chosen randomly among the outstanding orders.

Furthermore, we impose constant boundary conditions outside the moving frame of size $2K$: Every time the moving frame leaves a price level, the number of shares at that level is set to a_∞ (or b_∞ depending on the side of the book). a_∞ and b_∞ represent two "reservoirs of liquidity".

Our choice of a finite moving frame and constant boundary conditions has three motivations: Firstly, it assures that the order book does not empty and that P^A, P^B are always well defined. Secondly, it keeps the spread $S = P^A - P^B$ and the increments of P^A, P^B and $P = (P^A + P^B)/2$ bounded—This will be important when addressing the scaling limit of the price. Thirdly, it makes the model Markovian as we do not keep track of the price levels that have been visited (then left) by the moving frame at some prior time.

Figure 1 is a schematic representation of the order book.

4.2. Stability of the Order Book. We first specify the notations for the 4-variate Hawkes process. We set

$$\lambda_i(t) = \mu_i + \sum_{j=1}^{4} \alpha_{ij} e^{-\beta_{ij} s} dN_j(s), \qquad i \in [\![1, 4]\!], \tag{4.3}$$

and by convention the index 1 corresponds to L^+, 2 to M^+, 3 to L^- and 4 to M^-.

Proposition 4.1. *Provided $\beta_{\min}$ is large (specified precisely below), the order book (X_n, Y_n) is V-uniformly ergodic, where*

$$V(x, y) = e^{\omega \sum_{i=1}^{K} x_i^{\pm} + \gamma \sum_{1 \le k,l \le 4} y_{kl}}, \tag{4.4}$$

and $\omega > 0$ and $\gamma > 0$.

Proof. We follow the same pattern as the proof of proposition 3.3, and only modify it to account for the fact that the order book is formed from

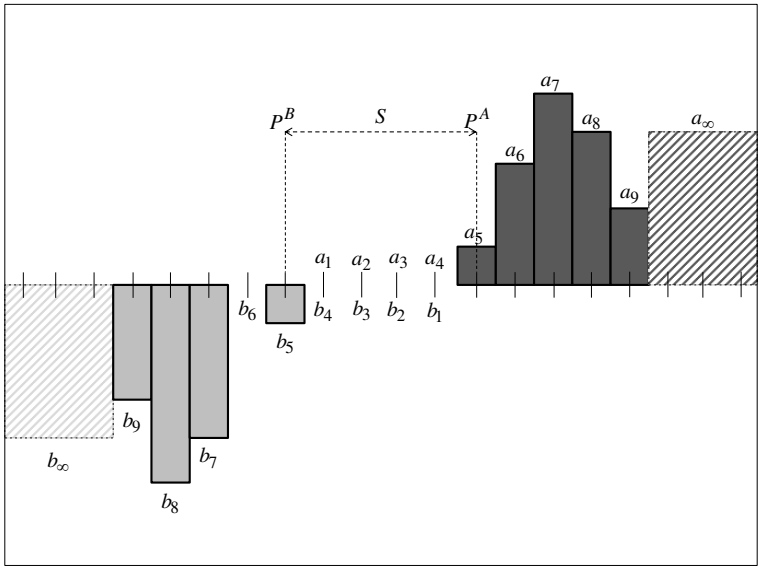


FIGURE 1. Order book dynamics: in this example, $K = 9$, $q = 1$, $a_\infty = 4$, $b_\infty = -4$. The shape of the order book is such that $a(t) = (0, 0, 0, 0, 1, 3, 5, 4, 2)$ and $b(t) = (0, 0, 0, 0, -1, 0, -4, -5, -3)$. The spread $S(t) = 5$ ticks. Assume that at time $t' > t$ a sell market order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

multiple queues, and the role of the boundary conditions a_∞ and b_∞:

$$
\begin{aligned}
\frac{\mathcal{P}V(x,y)}{V(x,y)} \quad \leq \quad & e^{\omega q + \omega a_\infty + \gamma \sum_{k=1}^4 \alpha_{k1}} \int_0^\infty e^{-\sum_{1\leq k,l\leq 4}(\gamma + \frac{1}{\beta_{kl}})(1-e^{-\beta_{kl}t})y_{kl} - \sum_{k=1}^4 \mu_k t - \lambda^{C^+} \sum_{k=1}^K x_k^+ t} \\
\times \quad & \left(\mu_1 + \sum_{j=1}^4 y_{1j} e^{-\beta_{1j}t}\right) dt \\
+ \quad & e^{-\omega q + \gamma \sum_{k=1}^4 \alpha_{k2}} \int_0^\infty e^{-\sum_{1\leq k,l\leq 4}(\gamma + \frac{1}{\beta_{kl}})(1-e^{-\beta_{kl}t})y_{kl} - \sum_{k=1}^4 \mu_k t - \lambda^{C^+} \sum_{k=1}^K x_k^+ t} \\
\times \quad & \left(\mu_2 + \sum_{j=1}^4 y_{2j} e^{-\beta_{2j}t}\right) dt \\
+ \quad & e^{-\omega q} \int_0^\infty e^{-\sum_{1\leq k,l\leq 4}(\gamma + \frac{1}{\beta_{kl}})(1-e^{-\beta_{kl}t})y_{kl} - \sum_{k=1}^4 \mu_k t - \lambda^{C^+} \sum_{k=1}^K x_k^+ t} \left(\lambda^{C^+} \sum_{k=1}^K x_k^+\right) dt \\
+ \quad & \text{similar terms for the bid side of the book} \qquad (4.5)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\mathcal{P}V(x,y)}{V(x,y)} &\leq \left(e^{\omega q+\omega a_\infty+\gamma\sum_{k=1}^4\alpha_{k1}}\mu_1 + e^{-\omega q+\omega a_\infty+\gamma\sum_{k=1}^4\alpha_{k2}}\mu_2 + e^{-\omega q}\lambda^{C^+}\sum_{k=1}^K x_k^+\right)\\
&\times \int_0^\infty e^{-\sum_{1\leq k,l\leq 4}(\gamma+\frac{1}{\beta_{kl}})(1-e^{-\beta_{kl}t})y_{kl}-\sum_{k=1}^4\mu_k t-\lambda^{C^+}\sum_{k=1}^K x_k^+ t}dt\\
&+ e^{\omega q+\omega a_\infty+\gamma\sum_{k=1}^4\alpha_{k1}}\sum_{j=1}^4 y_{1j}\int_0^\infty e^{-\sum_{1\leq k,l\leq 4}(\gamma+\frac{1}{\beta_{kl}})(1-e^{-\beta_{kl}t})y_{kl}-\sum_{k=1}^4\mu_k t-\lambda^{C^+}\sum_{k=1}^K x_k^+ t-\beta_{1j}}dt\\
&+ e^{-\omega q+\gamma\sum_{k=1}^4\alpha_{k2}}\sum_{j=1}^4 y_{2j}\int_0^\infty e^{-\sum_{1\leq k,l\leq 4}(\gamma+\frac{1}{\beta_{kl}})(1-e^{-\beta_{kl}t})y_{kl}-\sum_{k=1}^4\mu_k t-\lambda^{C^+}\sum_{k=1}^K x_k^+ t-\beta_{2j}}dt\\
&+ \text{similar terms for the bid side of the book.} \qquad (4.6)
\end{aligned}
$$

Again, as $|x|+|y|\to\infty$,

$$
\begin{aligned}
\frac{\mathcal{P}V(x,y)}{V(x,y)} &\leq e^{-wq}\\
&+ \frac{4}{\beta_{\min}(\gamma+\frac{1}{\beta_{\max}})}\\
&\times \left(e^{\omega q+\omega a_\infty+\gamma\sum_{k=1}^4\alpha_{k1}} + e^{-\omega q+\gamma\sum_{k=1}^4\alpha_{k2}} + e^{\omega q+\omega b_\infty+\gamma\sum_{k=1}^4\alpha_{k3}} + e^{-\omega q+\gamma\sum_{k=1}^4\alpha_{k4}}\right).
\end{aligned} \qquad (4.7)
$$

This quantity can be made smaller than 1 if $\beta_{\min}$ is large enough, and this concludes the proof of the proposition. □

4.3. Large-scale Limit of the Price Process. Given the state (X_{n-1}, Y_{n-1}) of the order book at time $n-1$ and the event E_n, the price increment at time n can be determined. We define the sequence of random variables

$$\eta_n = \Psi(X_{n-1}, Y_{n-1}, E_n), = \Phi(Z_n, Z_{n-1}), \qquad (4.8)$$

as the price increment at time n, where

$$Z_n = (X_n, Y_n). \qquad (4.9)$$

Ψ is a deterministic function giving the elementary "price-impact" of event E_n on the order book at state X_{n-1}. Let μ be the stationary distribution of (Z_n), and M its transition probability function. We are interested in the random sums

$$P_n := \sum_{k=1}^n \overline{\eta}_k = \sum_{k=1}^n \overline{\Phi}(Z_k, Z_{k-1}), \qquad (4.10)$$

where

$$\overline{\eta}_k := \eta_k - \mathbb{E}_\mu[\eta_k] = \overline{\Phi}_k = \Phi_k - \mathbb{E}_\mu[\Phi_k], \qquad (4.11)$$

and the asymptotic behavior of the rescaled-centered price process

$$\widetilde{P}^{(n)}(t) := \frac{P_{\lfloor nt \rfloor}}{\sqrt{n}}, \tag{4.12}$$

as n goes to infinity.

Proposition 4.2. *In event time, the large-scale limit of the price process is a Brownian motion. Formally, the series*

$$\sigma^2 = \mathbb{E}_\mu[\overline{\eta}_0^2] + 2\sum_{n=1}^{\infty} \mathbb{E}_\mu[\overline{\eta}_0 \overline{\eta}_n] \tag{4.13}$$

converges absolutely, and

$$\widetilde{P}^{(n)}(t) \overset{n\to\infty}{\longrightarrow} \sigma B(t), \tag{4.14}$$

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

Proof. This is an application of the functional central limit theorem for (stationary and ergodic) sequences of weakly dependent random variables with finite variance, and is identical to the proof of theorem 6.1 in [**AJ13**]. Firstly, we note that the variance of the price increments η_n is finite since it is bounded by $K + 1$. Secondly, the V-uniform ergodicity of (Z_n) is equivalent to

$$||M^n(z, .) - \mu(.)|| \leq R\rho^n V(z), n \in \mathbb{N}, \tag{4.15}$$

for some $R < \infty$ and $\rho < 1$. This implies thanks to theorem 16.1.5 in [**MT09**][6] that for any $g^2, h^2 \leq V$, $k, n \in \mathbb{N}$, and any initial condition z

$$|\mathbb{E}_z[g(Z_k)h(Z_{n+k})] - \mathbb{E}_z[g(Z_k)]\mathbb{E}_z[h(Z_k)]| \leq R\rho^n[1 + \rho^k V(z)], \tag{4.16}$$

where $\mathbb{E}_z[.]$ means $\mathbb{E}[.|Z_0 = z]$. This in turn implies

$$|\mathbb{E}_z[\overline{h}(Z_k)\overline{g}(Z_{k+n})]| \leq R_1\rho^n[1 + \rho^k V(z)] \tag{4.17}$$

for some $R_1 < \infty$, where $\overline{h} = h - \mathbb{E}_\mu[h], \overline{g} = g - \mathbb{E}_\mu[g]$. By taking the expectation over μ on both sides of (4.17) and noting that $\mathbb{E}_\mu[V(Z_0)]$ is finite by theorem 14.3.7 in [**MT09**], we get

$$|\mathbb{E}_\mu[\overline{h}(Z_k)\overline{g}(Z_{k+n})]| \leq R_2\rho^n = \rho(n), k, n \in \mathbb{N}. \tag{4.18}$$

Hence the stationary version of (Z_n) satisfies a geometric mixing condition, and in particular

$$\sum_n \rho(n) < \infty. \tag{4.19}$$

[6]We refer to §16.1.2 *"V-geometric mixing and V-uniform ergodicity"* in [**MT09**] for more details.

Theorems 19.2 and 19.3 in [**Bil99**] on functions of mixing processes let us conclude that

$$\sigma^2 := \mathbb{E}_\mu[\overline{\eta}_0^2] + 2\sum_{n=1}^{\infty} \mathbb{E}_\mu[\overline{\eta}_0\overline{\eta}_n] \tag{4.20}$$

is well-defined—the series in (4.20) converges absolutely—and coincides with the asymptotic variance

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E}_\mu\left[\sum_{k=1}^{n} (\overline{\eta}_k)^2\right] = \sigma^2. \tag{4.21}$$

Moreover

$$\widetilde{P}^{(n)}(t) \overset{n\to\infty}{\longrightarrow} \sigma B(t), \tag{4.22}$$

where $(B(t))$ is a standard Brownian motion. The convergence in (4.22) happens in $D[0,\infty)$, the space of $\mathbb{R}$-valued càdlàg functions, equipped with the Skorohod topology. □

5. Technical Lemmas

Lemma 5.1. *Let $a, b, c > 0$ be three positive real numbers. Then*

$$\begin{aligned} \mathcal{I}(a,b,c) &= \int_0^\infty e^{-a(1-e^{-bt})-ct}dt \\ &= \frac{(-a)^{-c/b}}{b}\, e^{-a}\, \left(\Gamma(\frac{c}{b}) - \Gamma(\frac{c}{b}, -a)\right), \end{aligned} \tag{5.1}$$

(5.2)

where the Gamma function *is defined for all complex numbers p such that $\mathcal{R}[p] > 0$ as*

$$\Gamma(p) = \int_0^\infty t^{p-1}e^{-t}dt, \tag{5.3}$$

and the incomplete Gamma function *is defined for all $p \in \mathbb{C}$, $\mathcal{R}[p] > 0$ and all $z \in \mathbb{C}$ as*

$$\Gamma(p,z) = \int_z^\infty t^{p-1}e^{-t}dt. \tag{5.4}$$

In particular, for all $b > 0, c > 0$

$$\lim_{a\to\infty} \mathcal{I}(a,b,c) = 0, \tag{5.5}$$

and

$$\lim_{a\to\infty} a\, \mathcal{I}(a,b,c) = \frac{1}{b}. \tag{5.6}$$

Proof. This representation and the limits can be obtained with a symbolic computation system such as Mathematica. □

Lemma 5.2. *More generally, if*

$$\mathcal{I}_p(a_1,\dots,a_p;b_1,\dots,b_p;c) = \int_0^\infty e^{-a_1(1-e^{-b_1 t})-\cdots-a_p(1-e^{-b_p t})-ct}dt, \tag{5.7}$$

with $a_i > 0, b_i > 0\ \forall i \in [\![1,p]\!]$*, and* $c > 0$*. Let*

$$b_{min} = \min\{b_i\}_{1\le i\le p}. \tag{5.8}$$

Then

$$\begin{aligned}\mathcal{I}_p(a_1,\dots,a_p;b_1,\dots,b_p;c) &\le \int_0^\infty e^{-\sum_{i=1}^p a_i\,(1-e^{-b_{min}t})-ct}\\ &= \mathcal{I}(|a|,b_{min},c),\end{aligned} \tag{5.9}$$

whith

$$|a| = \sum_{i=1}^p a_i. \tag{5.10}$$

Hence

$$\mathcal{I}_p(a;b;c) \le \frac{1}{b_{min}|a|},\quad \text{as } |a|\to\infty. \tag{5.11}$$

CHAPTER 5

Numerical Results and Applications

This chapter is geared towards practical aspects of stochastic order book modelling, namely the identification of a realistic point process-based order book model and its use as a workbench for algorithmic trading strategies assessment.

1. Introduction

1.1. Notations. We note for the intensity of multivariate Hawkes processes,

$$\lambda_i(t) = \mu_i + \int_0^t \sum_{j=1}^{M} \varphi_{ij}(t-s)dN_j(s), \qquad 1 \leq i \leq M. \tag{1.1}$$

or in vector form,

$$\boldsymbol{\lambda}(t) = \boldsymbol{\mu} + \int_0^t \boldsymbol{\varphi}(t-s)d\boldsymbol{N}(s), \tag{1.2}$$

and for exponentially decaying kernels,

$$\varphi_{ij}(t) = \sum_{p=1}^{P} \alpha_{ij}^p e^{-\beta_{ij}^p t}, \qquad P \geq 1, \tag{1.3}$$

$$\lambda_i(t) = \mu_i + \int_0^t \sum_{j=1}^{M} \sum_{p=1}^{P} \alpha_{ij}^p e^{-\beta_{ij}^p (t-s)} dN_j(s), 1 \leq i \leq M. \tag{1.4}$$

1.2. Classic Results. Firstly, the stationarity condition for Hawkes processes is

$$\rho\left(\int_0^\infty \boldsymbol{\varphi}(t)dt\right) < 1, \tag{1.5}$$

where ρ designates the spectral radius of a matrix, that is, the maximum of the absolute values of its eigenvalues. A proof of this result can be found in [**BM96**]. Stationarity was defined in chapter 4 (def. 2.6).

Secondly, taking the expectation of (1.2) and inverting, yields the following expression for the average intensity

$$\overline{\boldsymbol{\lambda}} = \mathbb{E}\left[\boldsymbol{\lambda}(t)\right] = \left(\boldsymbol{I} - \int_0^\infty \boldsymbol{\varphi}(t)dt\right)^{-1} \boldsymbol{\mu}. \tag{1.6}$$

2. Maximum Likelihood Estimation of Multivariate Hawkes Processes

The standard method for estimating the parameters of Hawkes processes is *via* the maximization of the likelihood function [**Oza79**]. We describe the method in this section and apply it to order book data.

First recall that the likelihood of a (regular univariate) point process, for an observation $0 < t_1, \cdots < t_n \leq T$, reads (A proof can be found in [**SM91**] or [**Rub72**].)

$$\mathcal{L}(t_1, \ldots, t_n) = e^{-\int_0^T \lambda(s)ds} \prod_{i=1}^n \lambda(t_i). \tag{2.1}$$

$\mathcal{L}(t_1, \ldots, t_n)$ is the *joint probability density* of observing n jumps exactly at times $t_1, \ldots, t_n$. The product $\prod_{i=1}^n \lambda(t_i)$ represents the probability of the process jumping at instants $t_1, \ldots, t_n$, while the term $e^{-\int_0^T \lambda(s)ds}$ corresponds to the probability of the process staying idle at any other instant in $]0, T]$.

The log-likelihood is therefore

$$\begin{aligned} \ln \mathcal{L}(t_1, \ldots, t_n) &= -\int_0^T \lambda(s)ds + \sum_{i=1}^n \ln \lambda(t_i) \\ &= -\int_0^T \lambda(s)ds + \int_0^T \ln \lambda(s)dN(s). \end{aligned} \tag{2.2}$$

2.1. Likelihood of Univariate Hawkes Processes. Consider now the specific case of a univariate Hawkes process. The intensity reads

$$\lambda(t) = \mu + \int_{-\infty}^t \sum_{p=1}^P \alpha^p e^{-\beta^p(t-s)} dN(s). \tag{2.3}$$

We are considering decay kernels of the form

$$\varphi(t) = \sum_{p=1}^P \alpha^p e^{-\beta^p t}. \tag{2.4}$$

Although this may seem restrictive at first, we shall show that this parametric form is rich enough to virtually approximate *any* kernel φ of interest (section 6.2).

The computations involved in equation (2.2) can be carried out explicitly. First set $P = 1$ (Thanks to the linearity of integration, the general case follows readily.)

$$\lambda(s) = \mu + \sum_{t_k < s} \alpha e^{-\beta(s-t_k)}. \tag{2.5}$$

We start with the first integral in (2.2). Assuming $t_n = T$ (This can always be made, as in practice we are only given an observation $t_1, \ldots, t_n$ and T is not specified [**Oza79**].),

$$\int_0^T \lambda(s)ds = \mu t_1 + \sum_{i=2}^{n} \int_{t_{i-1}}^{t_i} \lambda(s)ds. \tag{2.6}$$

$$\int_{t_{i-1}}^{t_i} \lambda(s)ds \quad = \quad \mu(t_i - t_{i-1}) + \alpha \int_{t_{i-1}}^{t_i} \sum_{t_k < s} e^{-\beta(s-t_k)}ds. \tag{2.7}$$

$$\begin{aligned} \int_{t_{i-1}}^{t_i} \sum_{t_k < s} e^{-\beta(s-t_k)}ds &= \int_{t_{i-1}}^{t_i} \sum_{t_k \le t_{i-1}} e^{-\beta(s-t_k)}ds \\ &= \sum_{t_k \le t_{i-1}} \int_{t_{i-1}}^{t_i} e^{-\beta(s-t_k)}ds \\ &= \frac{1}{\beta} \sum_{t_k \le t_{i-1}} \left[e^{-\beta(t_{i-1}-t_k)} - e^{-\beta(t_i-t_k)} \right]. \end{aligned} \tag{2.8}$$

Then

$$\sum_{i=2}^{n} \int_{t_{i-1}}^{t_i} \lambda(s)ds = \mu(t_n - t_1) + \sum_{i=2}^{n} \frac{\alpha}{\beta} \sum_{t_k \le t_{i-1}} \left[e^{-\beta(t_{i-1}-t_k)} - e^{-\beta(t_i-t_k)} \right]. \tag{2.9}$$

A careful look at the summand shows that the sum in the r.h.s. of (2.9) is telescopic. After simplification, it remains

$$\begin{aligned} \sum_{i=2}^{n} \int_{t_{i-1}}^{t_i} \lambda(s)ds &= \mu(t_n - t_1) + \frac{\alpha}{\beta} \left(\sum_{i=2}^{n} 1 - \sum_{k=1}^{n-1} e^{-\beta(t_n-t_k)} \right) \\ &= \mu(t_n - t_1) + \sum_{k=1}^{n-1} \frac{\alpha}{\beta} \left(1 - e^{-\beta(t_n-t_k)} \right), \end{aligned} \tag{2.10}$$

whence

$$\begin{aligned}\int_0^T \lambda(s)ds &= \mu t_n + \sum_{k=1}^{n-1} \frac{\alpha}{\beta}\left(1 - e^{-\beta(t_n - t_k)}\right) \\ &= \mu t_n + \sum_{k=1}^{n} \frac{\alpha}{\beta}\left(1 - e^{-\beta(t_n - t_k)}\right). \end{aligned} \tag{2.11}$$

Finally

$$\ln \mathcal{L}(t_1, \ldots, t_n) = -\mu t_n + \sum_{k=1}^{n} \frac{\alpha}{\beta}\left(e^{-\beta(t_n - t_k)} - 1\right) + \sum_{i=1}^{n} \ln\left(\mu + \alpha A(i)\right) \tag{2.12}$$

with

$$A(i) = \sum_{t_k < t_i} e^{-\beta(t_i - t_k)}, \quad \text{for } i \geq 2, \tag{2.13}$$

and

$$A(1) = 0. \tag{2.14}$$

More generally for $P \geq 1$, it holds,

Proposition 2.1 (Log-likelihood of a univariate Hawkes process).

$$\ln \mathcal{L}(t_1, \ldots, t_n) = -\mu t_n + \sum_{k=1}^{n} \sum_{p=1}^{P} \frac{\alpha^p}{\beta^p}\left(e^{-\beta^p(t_n - t_k)} - 1\right) + \sum_{i=1}^{n} \ln\left(\mu + \sum_{p=1}^{P} \alpha^p A^p(i)\right), \tag{2.15}$$

where for all $p \in [\![1, P]\!]$,

$$A^p(i) = \sum_{t_k < t_i} e^{-\beta^p(t_i - t_k)}, \ i \geq 2, \tag{2.16}$$

and

$$A^p(1) = 0. \tag{2.17}$$

2.2. Likelihood of Multivariate Hawkes Processes. Consider now the case of a M-variate Hawkes process, $M \geq 1$. The intensity vector $\lambda = (\lambda_1, \ldots, \lambda_M)^\top$ is specified by

$$\begin{aligned}\lambda_m(t) &= \mu_m + \sum_{j=1}^{M} \int_{-\infty}^{t} \sum_{p=1}^{P} \alpha_{mj}^p e^{-\beta_{mj}^p(t-s)} dN_j(s) \\ &= \mu_m + \sum_{j=1}^{M} \sum_{t_k^j < s} \sum_{p=1}^{P} \alpha_{mj}^p e^{-\beta_{mj}^p(t - t_k^j)}, \qquad m \in [\![1, M]\!], \end{aligned} \tag{2.18}$$

where t_k^j is the k-th jump in the process $(N_j(t))$, $j \in [\![1, M]\!]$.

We will denote the jump times of the multivariate process either as

$$\{t_k^j\}_{1\leq k\leq n^j, 1\leq j\leq M},$$

t_k^j being the k-th jump in the process $(N_j(t))$. Or equivalently as

$$\{t_i\}_{1\leq i\leq n},$$

in which case the marks $\{z_i\}_{1\leq i\leq n} \in [\![1, M]\!]$, indicating which component has jumped at time t_i, must also be given.

The log-likelihood of a multivariate Hawkes process is the sum of the log-likelihoods of each component

$$\ln \mathcal{L}\left(\{t_i, z_i\}_{1\leq i\leq n}\right) = \sum_{m=1}^{M} \ln \mathcal{L}_m\left(\{t_i, z_i\}_{1\leq i\leq n}\right), \tag{2.19}$$

where

$$\ln \mathcal{L}_m\left(\{t_i, z_i\}_{1\leq i\leq n}\right) = -\int_0^T \lambda_m(s)ds + \int_0^T \ln \lambda_m(s)dN_m(s). \tag{2.20}$$

As in the univariate case, the difficult term to compute is the first integral. We proceed similarly. We write ($P = 1$ to simplify notations)

$$\int_0^T \lambda_m(s)ds = \mu_m t_1 + \sum_{i=2}^{n} \int_{t_{i-1}}^{t_i} \lambda_m(s)ds. \tag{2.21}$$

$$\begin{aligned}\int_{t_{i-1}}^{t_i} \lambda_m(s)ds &= \mu_m(t_i - t_{i-1}) + \sum_{j=1}^{M} \int_{t_{i-1}}^{t_i} \sum_{t_k^j<s} \alpha_{mj} e^{-\beta_{mj}(s-t_k^j)} ds \\ &= \mu_m(t_i - t_{i-1}) + \sum_{j=1}^{M} \sum_{t_k^j\leq t_{i-1}} \frac{\alpha_{mj}}{\beta_{mj}} \left[e^{-\beta_{mj}(t_{i-1}-t_k^j)} - e^{-\beta_{mj}(t_i-t_k^j)}\right].\end{aligned} \tag{2.22}$$

Hence

$$\int_0^T \lambda_m(s)ds = \mu_m t_n + \sum_{j=1}^{M} \sum_{i=2}^{n} \sum_{t_k^j\leq t_{i-1}} \frac{\alpha_{mj}}{\beta_{mj}} \left[e^{-\beta_{mj}(t_{i-1}-t_k^j)} - e^{-\beta_{mj}(t_i-t_k^j)}\right] \tag{2.23}$$

The sum

$$\sum_{i=2}^{n} \sum_{t_k^j\leq t_{i-1}} \frac{\alpha_{mj}}{\beta_{mj}} \left[e^{-\beta_{mj}(t_{i-1}-t_k^j)} - e^{-\beta_{mj}(t_i-t_k^j)}\right] \tag{2.24}$$

is telescopic and equals after simplification

$$\sum_{t_k^j} \frac{\alpha_{mj}}{\beta_{mj}} \left(1 - e^{-\beta_{mj}(t_n - t_k^j)}\right). \tag{2.25}$$

Therefore,

$$\int_0^T \lambda_m(s)ds \quad = \quad \mu_m t_n + \sum_{j=1}^{M} \sum_{t_k^j} \frac{\alpha_{mj}}{\beta_{mj}} \left(1 - e^{-\beta_{mj}(t_n - t_k^j)}\right). \tag{2.26}$$

More generally for $P \geq 1$, we state,

Proposition 2.2 (Log-likelihood of a multivariate Hawkes process).

$$\begin{aligned} \ln \mathcal{L}_m \left(\{t_i, z_i\}_{1 \leq i \leq n}\right) \quad &= \quad -\mu_m t_n + \sum_{j=1}^{M} \sum_{t_k^j} \sum_{p=1}^{P} \frac{\alpha_{mj}^p}{\beta_{mj}^p} \left(e^{-\beta_{mj}^p (t_n - t_k^j)} - 1\right) \\ &+ \quad \sum_{t_k^m} \ln \left(\mu_m + \sum_{j=1}^{M} \sum_{p=1}^{P} \alpha_{mj}^p A_{mj}^p(k)\right), \end{aligned} \tag{2.27}$$

where

$$A_{mj}^p(k) = \sum_{t_l^j < t_k^m} e^{-\beta_{mj}^p (t_k^m - t_l^j)}, \ i \geq 2, \tag{2.28}$$

and

$$A_{mj}^p(1) = 0. \tag{2.29}$$

Formula (2.27) is the backbone of maximum likelihood estimation of linear Hawkes processes with exponential kernels.

Remark 2.1. *(Recursive expression of $A(k)$* [**MT11a, Oga81**]*) In order to accelerate the computation of the likelihood, it is custom to use the following recursive relanship:*

$$A_{mj}^p(k) = e^{-\beta_{mj}^p (t_k^m - t_{k-1}^m)} A_{mj}^p(k-1) + \sum_{t_{k-1}^m \leq t_l^j < t_k^m} e^{-\beta_{mj}^p (t_k^m - t_l^j)}, \qquad k \geq 2, \tag{2.30}$$

with the initial condition

$$A_{mj}^p(1) = 0. \tag{2.31}$$

In a univariate setting, this simply reads

$$A^p(k) = e^{-\beta^p (t_k - t_{k-1})} \left(A^p(k-1) + 1\right). \tag{2.32}$$

Given a data set $t_1, \ldots, t_n$ and $z_1, \ldots, z_n$, the maximum likelihood estimation method (MLE) consists in maximizing $\mathcal{L}(\{t_i\}, \{z_i\})$, or equivalently its logarithm $\ln \mathcal{L}(\{t_i\}, \{z_i\})$, with respect to the model parameters $\boldsymbol{\mu}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\beta}$.

The rationale is that one maximizes the probability of observing the actual sample $t_1, \ldots, t_n, z_1, \ldots, z_n$. Formal foundations of the method for the identification of Hawkes processes (albeit in the univariate case only) can be found in [**Oga78**] (theorems 2 and 5). There, it is shown (under technical conditions) that the MLE estimator is *consistent*, that is tends to the actual parameters values, and asymptotically *normal* (around the parameters) as the sample size tends to infinity.

Separability of the log-likelihood of multivariate Hawkes processes. We note that for each $m \in [\![1, M]\!]$, the *partial* log-likelihood $\ln \mathcal{L}_m$ (formula (2.27)), depends on μ_m, $(\alpha^p_{mj})_{1 \le j \le M, 1 \le p \le P}$ and $(\beta^p_{mj})_{1 \le j \le M, 1 \le p \le P}$ only. There is no dependence on $(\mu_{m'})_{m' \neq m}$, $(\alpha^p_{m'j})_{m' \neq m, 1 \le j \le M, 1 \le p \le P}$ nor $(\beta^p_{m'j})_{m' \neq m, 1 \le j \le M, 1 \le p \le P}$.

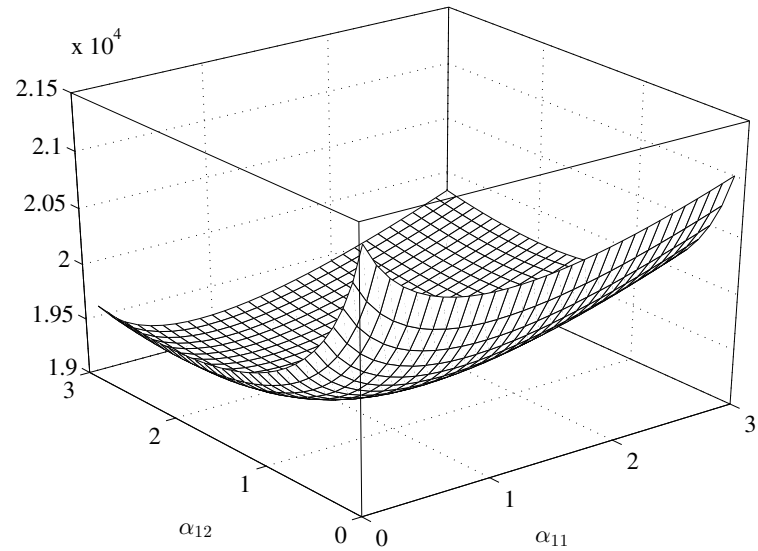
Thus in order to maximize $\ln \mathcal{L}$ (a function of $M + 2M^2P$ variables), one only needs to maximize the partial log-likelihood $\ln \mathcal{L}_m$ (a function of 1+2MP variables) separately for each $m \in [\![1, M]\!]$. This substantially reduces the complexity of the maximization.

2.3. Illustration. Figure 1.

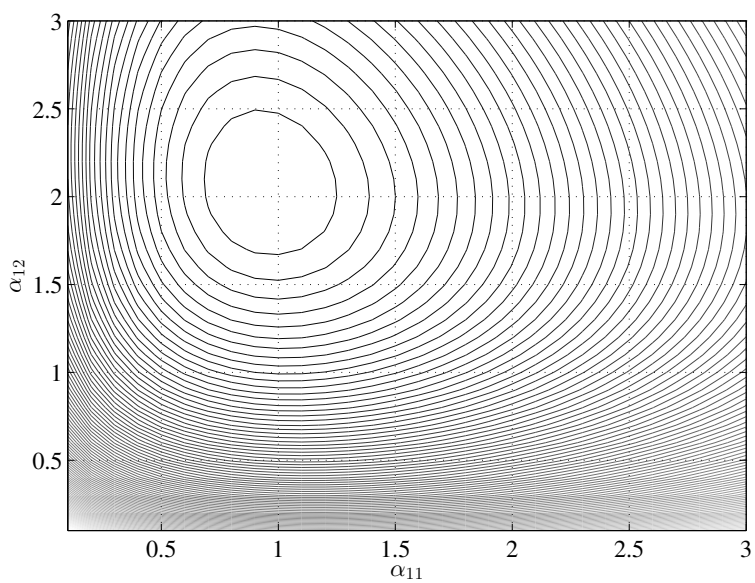
3. Goodness of Fit Assessment

Equally important to the identification of a point process is the assessment of the model's ability to mimic the data. We propose three approaches.

3.1. A Pragmatic Assessment. The first approach is the easiest to implement. First estimate the parameters by MLE. Then simulate the process with the obtained parameters (details about the simulation can be found in section). Finally Q-Q-plot the inter-arrival times of the simulated sample against those of the data. If the obtained plot follows the 45° line $y = x$, then the fit is satisfactory. Incidentally, this method takes into account the noise inherent to the simulation.



(A) Surface plot.



(B) Contour plot.

FIGURE 1. Negative of the partial log-likelihood, $-\ln \mathcal{L}_1$, for a 10000 points simulated sample as a function of $(\alpha_{11}, \alpha_{12})$. The parameters values are $M = 2$, $P = 1$, $\boldsymbol{\mu} = (0.1, 0.2)^\top$, $\boldsymbol{\alpha} = \begin{pmatrix} 1.0 & 2.0 \\ 3.0 & 4.0 \end{pmatrix}$, $\boldsymbol{\beta} = \begin{pmatrix} 10.0 & 20.0 \\ 30.0 & 40.0 \end{pmatrix}$. There is clearly a minimum at $\left(\alpha_{11}^*, \alpha_{12}^*\right) = (1.0, 2.0)$.

3.2. A Theoretically Sounder Assessment: the Time-rescaling Theorem. We first state the underlying theorem.

PROPOSITION 1 (Time-rescaling). *Let $\boldsymbol{N}(t) = (N_1(t), \ldots, N_M(t))$ be a M-variate Hawkes process. Fix $m \in [\![1, M]\!]$ and define the sequence $(\tau_k^m)_{2 \le k \le n^m}$ of random variables by*

$$\tau_k^m = \int_{t_{k-1}^m}^{t_k^m} \lambda_m(s) ds. \tag{3.1}$$

Then $(\tau_k^m)_{2 \le k \le n^m}$ are i.i.d. exponential *random variables with* parameter 1.

PROOF. Can be found in [**BBVK01**] (univariate) or [**Bow07**] (multivariate setting). □

The principle of the method is to estimate the parameters by MLE, compute the quantities (τ_k^m), then Q-Q-plot them against a standard exponential random variable. To compute the (τ_k^m) in practice, we made use of equation (2.22)

$$\int_{t_{i-1}}^{t_i} \lambda_m(s) ds \quad = \quad \mu_m(t_i - t_{i-1}) + \sum_{j=1}^{M} \sum_{t_k^j \le t_{i-1}} \frac{\alpha_{mj}}{\beta_{mj}} \left[e^{-\beta_{mj}(t_{i-1} - t_k^j)} - e^{-\beta_{mj}(t_i - t_k^j)} \right]$$

derived earlier for the integrated intensity, and the additivity of integration.

3.3. An Even Sharper Assessment. If we transform the (τ_k^m) to

$$\xi_k^m = 1 - e^{-\tau_k^m}, \tag{3.2}$$

then $(\xi_k^m)_{2 \le k \le n^m}$ are independent *uniform* random variables on the interval $]0, 1[$. It is known that the k-th *order statistic* $\xi_{(k)}^m$ of a uniform random variable follows a Beta distribution with parameters k and $n^m - k + 1$

$$f(\xi \mid k, n^m - k + 1) = \frac{n^m!}{(n^m - k)!(k-1)!} \xi^{k-1} (1 - \xi)^{n^m - k}, \qquad 0 < \xi < 1. \tag{3.3}$$

In (3.3), we stress that the notation n^m designates the total number of jumps in the porcess $(N_m(t))$—the m is *not* for exponentiation.

We can set 95% confidence bounds on $\xi_{(k)}^m$ using the 2.5% and 97.5% quantiles of the Beta distribution (These are available in most statistical software packages.) Alternatively, as noted in [**BBVK01**], for medium to large sized samples, we can use the Gaussian approximation of the Binomial

distribution in equation (3.3). The 95% confidence intervals thus obtained are

$$\xi^m_{(k)} \pm 1.96\sqrt{\frac{\xi^m_{(k)}(1-\xi^m_{(k)})}{n^m}}. \tag{3.4}$$

The idea is then, for each $m \in [\![1, M]\!]$, to compute $(\tau^m_k)_{2\le k\le n^m}$ as in section 3.2, transform them into (ξ^m_k), order the latter into $(\xi^m_{(k)})$, and plot the result against the CDF of the uniform distribution $b_k = \frac{k-\frac{1}{2}}{n^m}$, with (local) confidence bounds defined by equation (3.4).

Detailed examples of goodness of fit assessment are given in section 4. A fourth diagnosis method relying on the *covariance function* is thoroughly discussed in section 5.

4. Application to Tick By Tick Order Book Data

We refer to chapter 3, section 2 for the description of the tick by tick data we use and its processing. We estimate a 4-variate Hawkes process to the flow of:

(1) buy market orders $M^+(t)$,
(2) sell market orders $M^-(t)$,
(3) sell limit orders $L^+(t)$,
(4) buy limit orders $L^-(t)$.

In figures 2 and 3 the resulting goodness of fit plots are displayed, and in figure 4 the estimated decay kernel.

The following practical remarks are in order:

- To achieve the fit accuracy of figures 2 and 3, we set the number of exponentials to $P = 4$.
- Since the time resolution of the data is 1 millisecond, we add a random $\Delta t = 10^{-3}U$ second, with $U \sim$ Uniform $]0, 1[$ to points with the same time stamp (around 15% of the points for liquid CAC 40 stocks.)
- We consider limit orders arriving 5 ticks away from the best opposite quote, lumping together limit orders at each side of the book in one order flow. More sophisticated models could have a point process attached to each limit.

- We illustrate with the particular stock TOTF.PA (TOTAL), but the quality of the fit is similar across *all* CAC 40 stocks.
- We do not consider cancellations. Apart from reducing the dimensionality of the model from 6 to 4 order flows, this is motivated by the theoretical analysis of chapters 2 and 4, where it is shown that *proportional* cancellation rates are important for the stability of the order book.
- We use *warm-start* optimization strategies to accelerate the estimation. We first optimize over a subset (30 min) of the data then use the result as a starting point for the optimization over the full dataset (2 hours).

5. Non-parametric Estimation of Multivariate Hawkes Processes

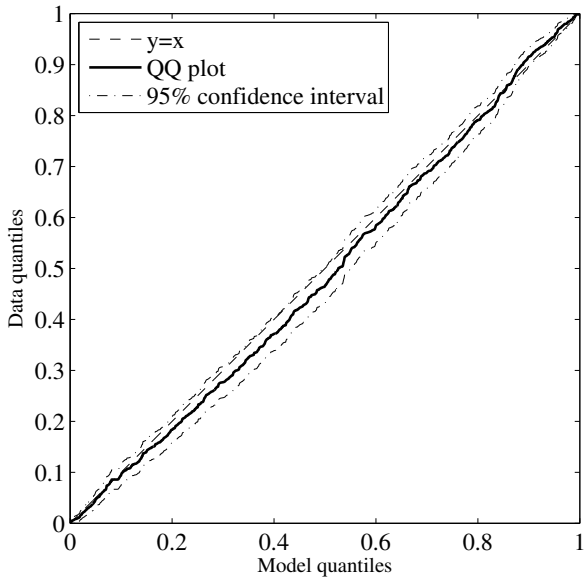
Although the results obtained by MLE are (very) satisfactory, we digress on a different estimation method recently proposed by Bacry and Muzy in [**BM13**]. We do this for the power of the method (when applicable) and the elegance of its underlying principle. While MLE exploits the structure of inter-event durations, the proposed method relies on the properties of counts of events, and is in a sense complementary to MLE. The idea dates back to the original papers [**Haw71a, Haw71b**] and is based on the *covariance density matrix*.

5.1. The Covariance Density Matrix. With usual notations, let $\boldsymbol{N}(t) = (N_1(t), \ldots, N_M(t))^\top$ be a M-variate regular[1] point process.

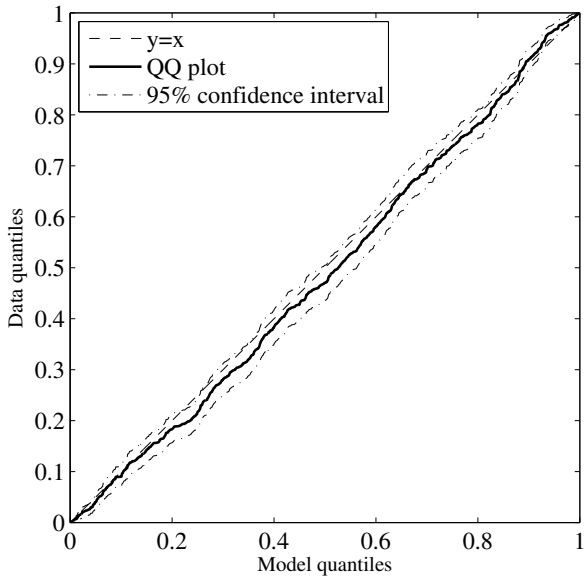
Definition 5.1 (Covariance density matrix). *The* covariance density matrix *of* $(\boldsymbol{N}(t))$ *is the* $M \times M$ *matrix-valued function defined, for all* $x \in \mathbb{R}^*$, *by*

$$\begin{aligned} C_{ij}(x) &= \mathbb{E}\left[\frac{dN_i(t+x)}{dx}\frac{dN_j(t)}{dt}\right] - \mathbb{E}\left[\frac{dN_i(t+x)}{dx}\right]\mathbb{E}\left[\frac{dN_j(t)}{dt}\right] \\ &= \mathbb{E}\left[\frac{dN_i(t+x)}{dx}\frac{dN_j(t)}{dt}\right] - \overline{\lambda}_i\overline{\lambda}_j, \qquad 1 \le i, j \le M, \end{aligned} \tag{5.1}$$

[1]That is, the probability of multiple jumps occuring at the same time is null, $Prob\,[dN_i(t) > 1] = o(dt), \qquad 1 \le i \le M.$

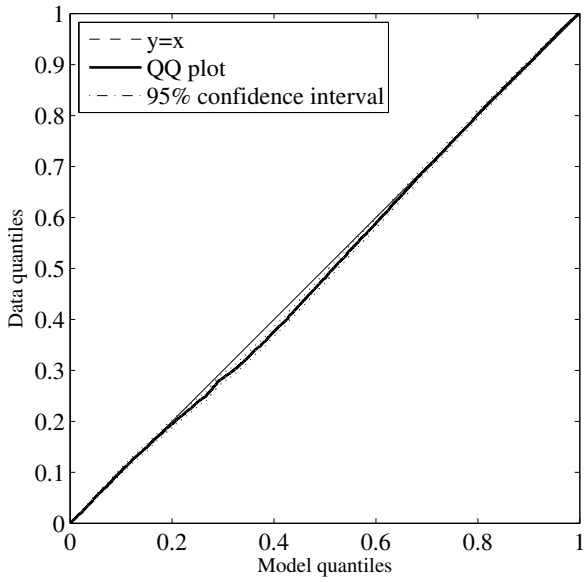


(A) Buy market orders.

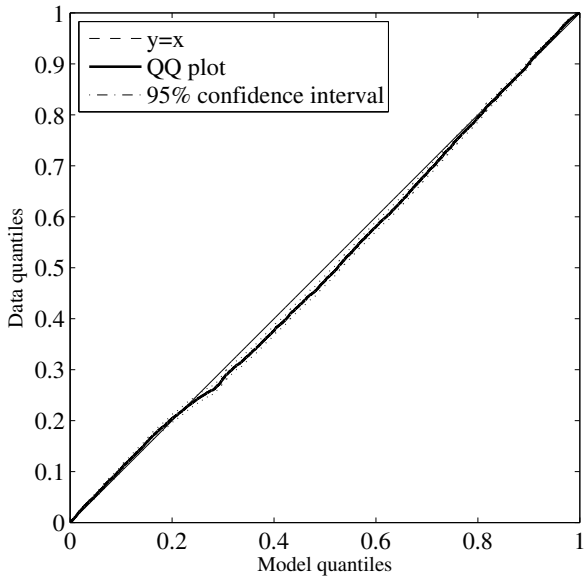


(B) Sell market orders.

FIGURE 2. Joint maximum likelihood estimation of 4 order flows (TOTF.PA, June 2011.) Continued in figure 3.



(A) Sell limit orders.



(B) Buy limit orders.

FIGURE 3. Joint maximum likelihood estimation of 4 order flows (TOTF.PA, June 2011.) Continuation of figure 2.

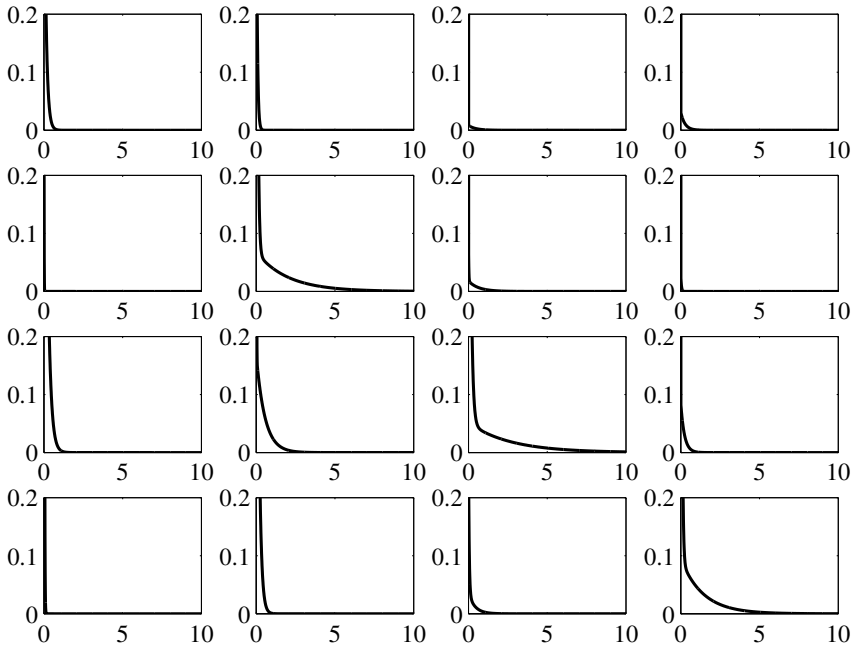


FIGURE 4. Maximum likelihood estimated kernel (TOTF.PA, June 2011.) This 4×4 matrix represents the decay kernels $\{\varphi_{ij}(t)\}_{1 \le i,j \le 4}$. x axis is time in seconds.

or equivalently,

$$\begin{aligned} \boldsymbol{C}(x) &= \mathbb{E}\left[\frac{d\boldsymbol{N}(t+x)}{dx}\frac{d\boldsymbol{N}^{\top}(t)}{dt}\right] - \mathbb{E}\left[\frac{d\boldsymbol{N}(t+x)}{dx}\right]\mathbb{E}\left[\frac{d\boldsymbol{N}^{\top}(t)}{dt}\right] \\ &= \mathbb{E}\left[\frac{d\boldsymbol{N}(t+x)}{dx}\frac{d\boldsymbol{N}^{\top}(t)}{dt}\right] - \overline{\boldsymbol{\lambda}}\,\overline{\boldsymbol{\lambda}}^{\top}, \end{aligned} \tag{5.2}$$

where

$$\begin{aligned} \overline{\boldsymbol{\lambda}} &= \mathbb{E}\left[\boldsymbol{\lambda}(t)\right] \\ &= (\overline{\lambda}_1, \ldots, \overline{\lambda}_M)^{\top}, \end{aligned} \tag{5.3}$$

is the mean intensity of the point process.

In definition 5.1, dx and dt are better thought of as small (indeed, infinitesimal) time increments. The quantity $dN_i(t)$ is the number of jumps of

(N_i) occurring in the interval $[t, t+dt[$. For small time intervals, the process being regular, this quantity is either 0 or 1.

Since

$$\mathbb{E}\left[(dN_i(t))^2\right] = [dN_i(t)] = \overline{\lambda}_i dt, \tag{5.4}$$

the diagonal terms $C_{ii}(x)$, $1 \leq i \leq M$, are singular at $x = 0$. It is convenient to isolate the singularity, defining the *complete covariance density* for all $x \in \mathbb{R}$ as

$$\boldsymbol{C}^{(c)}(x) = \boldsymbol{D}\delta(x) + \boldsymbol{C}(x), \tag{5.5}$$

whith

$$\boldsymbol{D} = diag(\overline{\boldsymbol{\lambda}}). \tag{5.6}$$

$\delta(x)$ is Dirac's delta function, and $\boldsymbol{C}(x)$ continuous at the origin. The upper script (c) stands for "complete". Note that for all $x \in \mathbb{R}$

$$C_{ij}(-x) = C_{ji}(x), \qquad 1 \leq i, j \leq M, \tag{5.7}$$

or in more compact notation

$$\boldsymbol{C}(-x) = \boldsymbol{C}^{\top}(x). \tag{5.8}$$

Thus, knowing $\boldsymbol{C}$ on $\mathbb{R}^+$ is enough to identify it.

Example 5.1 (Univariate Poisson process with constant intensity μ). *Clearly in this case, for all $x \in \mathbb{R}$,*

$$\begin{aligned} C^{(c)}(x) &= \mu\delta(x), \\ C(x) &= 0. \end{aligned} \tag{5.9}$$

5.2. The Hawkes-Wiener-Hopf Integral Equation. The following lemma will be used.

Lemma 1. *If $\mathcal{H}(t) = \sigma(N(s), s \leq t)$ denotes the history of the point process N up to time t, and X a $\mathcal{H}(t)$-measurable random variable, then*

$$\mathbb{E}\left[\frac{dN(t)}{dt}X\right] = \mathbb{E}\left[\lambda(t)X\right]. \tag{5.10}$$

Proof. By definition of λ

$$\mathbb{E}\left[dN(t)|\mathcal{H}_t\right] = \lambda(t)dt. \tag{5.11}$$

$$\begin{aligned}
\mathbb{E}\left[\lambda(t)X\right] &= \mathbb{E}\left[\mathbb{E}\left[\frac{dN(t)}{dt}|\mathcal{H}_t\right]X\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\frac{dN(t)}{dt}X|\mathcal{H}_t\right]\right] \\
&= \mathbb{E}\left[\frac{dN(t)}{dt}X\right]. \qquad (5.12)
\end{aligned}$$

□

We now particularize to the case of (multivariate) Hawkes processes. We write for $x > 0$

$$\boldsymbol{C}(x) = \mathbb{E}\left[\frac{d\boldsymbol{N}(t+x)}{dx}\frac{d\boldsymbol{N}^\top(t)}{dt}\right] - \overline{\boldsymbol{\lambda}}\,\overline{\boldsymbol{\lambda}}^\top. \qquad (5.13)$$

$$\begin{aligned}
\mathbb{E}\left[\frac{d\boldsymbol{N}(t+x)}{dx}\frac{d\boldsymbol{N}^\top(t)}{dt}\right] &= \mathbb{E}\left[\boldsymbol{\lambda}(t+x)\frac{d\boldsymbol{N}^\top(t)}{dt}\right] \\
&= \mathbb{E}\left[\left(\boldsymbol{\mu} + \int_{-\infty}^{t+x}\boldsymbol{\varphi}(t+x-u)d\boldsymbol{N}(u)\right)\frac{d\boldsymbol{N}^\top(t)}{dt}\right] \\
&= \boldsymbol{\mu}\overline{\boldsymbol{\lambda}}^\top + \mathbb{E}\left[\int_{-\infty}^{t+x}\boldsymbol{\varphi}(t+x-u)d\boldsymbol{N}(u)\,\frac{d\boldsymbol{N}^\top(t)}{dt}\right]. \qquad (5.14)
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\left[\int_{-\infty}^{t+x}\boldsymbol{\varphi}(t+x-u)d\boldsymbol{N}(u)\,\frac{d\boldsymbol{N}^\top(t)}{dt}\right] &= \mathbb{E}\left[\int_{-\infty}^{x}\boldsymbol{\varphi}(x-v)d\boldsymbol{N}(t+v)\,\frac{d\boldsymbol{N}^\top(t)}{dt}\right] \\
&= \int_{-\infty}^{x}\boldsymbol{\varphi}(x-v)\mathbb{E}\left[\frac{d\boldsymbol{N}(t+v)}{dv}\,\frac{d\boldsymbol{N}^\top(t)}{dt}\right]dv \\
&= \int_{-\infty}^{x}\boldsymbol{\varphi}(x-v)\left(\boldsymbol{C}^{(c)}(v) - \overline{\boldsymbol{\lambda}}\,\overline{\boldsymbol{\lambda}}^\top\right)dv \\
&= \boldsymbol{\varphi}(x)\boldsymbol{D} + \int_{-\infty}^{x}\boldsymbol{\varphi}(x-v)\boldsymbol{C}(v)dv \\
&- \int_{-\infty}^{x}\boldsymbol{\varphi}(x-v)dv\,\overline{\boldsymbol{\lambda}}\,\overline{\boldsymbol{\lambda}}^\top. \qquad (5.15)
\end{aligned}$$

$$\boldsymbol{C}(x) = \left(\boldsymbol{\mu} - \int_{-\infty}^{x}\boldsymbol{\varphi}(x-v)dv\,\overline{\boldsymbol{\lambda}} - \overline{\boldsymbol{\lambda}}\right)\overline{\boldsymbol{\lambda}}^\top + \boldsymbol{\varphi}(x)\boldsymbol{D} + \int_{-\infty}^{x}\boldsymbol{\varphi}(x-v)\boldsymbol{C}(v)dv. \qquad (5.16)$$

Since

$$\boldsymbol{\mu} - \int_0^\infty \boldsymbol{\varphi}(t)dt\, \overline{\boldsymbol{\lambda}} - \overline{\boldsymbol{\lambda}} = 0, \tag{5.17}$$

we get, $x > 0$,

$$\boldsymbol{C}(x) \quad = \quad \boldsymbol{\varphi}(x)\boldsymbol{D} + \int_{-\infty}^x \boldsymbol{\varphi}(x-v)\boldsymbol{C}(v)dv. \tag{5.18}$$

This is an integral equation linking $\boldsymbol{C}$ and $\boldsymbol{\varphi}$. We shall rewrite it in a classical integral equation form. Firstly

$$\boldsymbol{\varphi}(x) \quad = \quad \boldsymbol{C}(x)\boldsymbol{D}^{-1} - \int_{-\infty}^x \boldsymbol{\varphi}(x-v)\boldsymbol{C}(v)\boldsymbol{D}^{-1}dv.$$

If we define

$$\boldsymbol{G}(x) = \boldsymbol{C}(x)\boldsymbol{D}^{-1}, \tag{5.19}$$

$$\boldsymbol{\varphi}(x) \quad = \quad \boldsymbol{G}(x) - \int_{-\infty}^x \boldsymbol{\varphi}(x-v)\boldsymbol{G}(v)dv. \tag{5.20}$$

Let the transposes of $\boldsymbol{G}$ and $\boldsymbol{\varphi}$ be

$$\boldsymbol{K}(x) = \boldsymbol{G}^\top(x). \tag{5.21}$$

$$\boldsymbol{\phi}(x) = \boldsymbol{\varphi}^\top(x). \tag{5.22}$$

Then

$$\boldsymbol{\phi}(x) \quad = \quad \boldsymbol{K}(x) - \int_{-\infty}^x \boldsymbol{K}(v)\boldsymbol{\phi}(x-v)dv. \tag{5.23}$$

A last change of variables in the integral yields the desired result.

Proposition 5.1 ([**Haw71a, Haw71b**]). *With the above notations, it holds for all $x > 0$,*

$$\boldsymbol{\phi}(x) \quad = \quad \boldsymbol{K}(x) - \int_0^\infty \boldsymbol{K}(x-t)\boldsymbol{\phi}(t)dt. \tag{5.24}$$

We call this the *Hawkes-Wiener-Hopf integral equation* (HWH). It is a Fredholm integral equation of the second kind (actually a linear system thereof[2]). As such, it lends itself to a variety of solution methods [**DM85, PTVF07**]. We choose the simplest: Nyström's.

In order to get a practical sense of the functions at play, we illustrate, before we proceed, with the following example.

[2]Note that both $\boldsymbol{K}$ and $\boldsymbol{\phi}$ are M by M matrices. In particular, the product inside the integral is not commutative.

EXAMPLE 5.2. *Consider the univariate exponential case*

$$K(x) = Ae^{-B|x|}. \tag{5.25}$$

$$\phi(x) = K(x) - \int_0^\infty K(x-t)\phi(t)dt. \tag{5.26}$$

With slightly more general notation

$$\phi(x) = F(x) + \kappa \int_0^\infty K(x-t)\phi(t)dt. \tag{5.27}$$

$$\kappa = -1, \qquad F(x) = K(x). \tag{5.28}$$

This is close to the Lalesco-Picard integral equation. We rewrite

$$\phi(x) = F(x) + \kappa A e^{-Bx} \int_0^x e^{Bt}\phi(t)dt + \kappa A e^{Bx} \int_x^\infty e^{-Bt}\phi(t)dt. \tag{5.29}$$

Differentiation twice with respect to x yields

$$\phi'(x) = F'(x) - \kappa A B e^{-Bx} \int_0^x e^{Bt}\phi(t)dt + \kappa A e^{Bx} \int_x^\infty e^{-Bt}\phi(t)dt, \tag{5.30}$$

$$\phi''(x) = F''(x) - B^2 F(x) + B(-2\kappa A + B)\phi(x). \tag{5.31}$$

With $\kappa = -1$ *and* $F(x) = K(x) = Ae^{-Bx}$, $x > 0$

$$\phi''(x) = B(2A + B)\phi(x). \tag{5.32}$$

Which with the boundary condition

$$\lim_{x\to\infty} \phi(x) = 0, \tag{5.33}$$

admits the solution

$$\phi(x) = \alpha e^{-\beta x}, \tag{5.34}$$

$$\beta = \sqrt{B^2 + 2AB}. \tag{5.35}$$

Now using equation (5.26) *at 0*

$$\alpha = \phi(0) = F(0) + A\alpha \int_0^\infty e^{-(B+\beta)t}, \tag{5.36}$$

we get

$$\alpha = \frac{A(B+\beta)}{A+B+\beta}. \tag{5.37}$$

We can therefore state.

PROPOSITION 5.2. *Suppose the covariance density of a univariate Hawkes process writes*

$$C(x) = \overline{\lambda}\, A e^{-B|x|}, \qquad x \in \mathbb{R}, A \geq 0, B > 0. \tag{5.38}$$

$\overline{\lambda}$ *being the empirical mean intensity. Then, the process's kernel is*

$$\varphi(x) = \alpha e^{-\beta x}, \qquad x \geq 0, \tag{5.39}$$

$$\varphi(x) = 0, \qquad x < 0, \tag{5.40}$$

with

$$\beta = \sqrt{B^2 + 2AB} \tag{5.41}$$

$$\alpha = \frac{A(B + \sqrt{B^2 + 2AB})}{A + B + \sqrt{B^2 + 2AB}} = \frac{A(B+\beta)}{A+B+\beta}. \tag{5.42}$$

These are the inverse of equation (16) in [**Haw71b**]. When appropriate, that is, when the empirical covariance is well fit by *one* exponential[3], formulae (5.39)–(5.42) make the identification of a univariate Hawkes process[4] immediate. As an aside, they could serve as a benchmark for numerical solutions of equation (5.24).

5.3. Nyström's Method. Consider the system of integral equations (5.24), $x > 0$,

$$\boldsymbol{\phi}(x) = \boldsymbol{F}(x) + \kappa \int_0^\infty \boldsymbol{K}(x-t)\boldsymbol{\phi}(t)dt. \tag{5.48}$$

[3]This is rarely the case in financial datasets. Typical covariances are *power-law* tailed.

[4]It is easy to check that the stability condition

$$\alpha < \beta, \tag{5.43}$$

is always met for the Hawkes process thus defined. Ideed

$$\frac{A(B+\beta)}{A+B+\beta} < \beta, \tag{5.44}$$

is equivalent to

$$\beta^2 + B\beta - AB > 0. \tag{5.45}$$

Since

$$\beta \geq B, \tag{5.46}$$

$$\begin{aligned} \beta^2 + B\beta - AB &\geq \beta^2 + B^2 - AB \\ &= 2B^2 + AB \\ &> 0. \end{aligned} \tag{5.47}$$

$\boldsymbol{\phi}$ is the unknown, a M by M matrix-valued function. Both $\boldsymbol{K}$ and $\boldsymbol{F}$ are given M by M matrix-valued functions. $\boldsymbol{\phi}$ is defined on $\mathbb{R}_+$ (By causality of the kernel, it is null outside.) $\boldsymbol{F}$ is defined on $\mathbb{R}_+$ and $\boldsymbol{K}$ on $\mathbb{R}$. κ is a constant.

Let $\{t_m\}_{0\le m\le N}$ and $\{w_m\}_{0\le m\le N}$ be the nodes and weights of a certain quadrature rule on $[0,\infty[$. For all $0 \le m \le N$,

$$\boldsymbol{\phi}(t_m) \quad = \quad \boldsymbol{F}(t_m) + \kappa \int_0^\infty \boldsymbol{K}(t_m - t)\boldsymbol{\phi}(t)dt. \tag{5.49}$$

Discretize the integral,

$$\boldsymbol{\phi}(t_m) \quad = \quad \boldsymbol{F}(t_m) + \kappa \sum_{n=0}^{N} w_n \boldsymbol{K}(t_m - t_n)\boldsymbol{\phi}(t_n). \tag{5.50}$$

Rearrange,

$$\sum_{n=1}^{N} \left[\boldsymbol{I}\delta_{mn} - \kappa w_n \boldsymbol{K}(t_m - t_n)\right] \boldsymbol{\phi}(t_n) \quad = \quad \boldsymbol{F}(t_m), \qquad 0 \le m \le N. \tag{5.51}$$

Or, in matrix notation

$$\begin{pmatrix} \boldsymbol{I} - \kappa w_0 \boldsymbol{K}(0) & -\kappa w_1 \boldsymbol{K}(t_0 - t_1) & \cdots & -\kappa w_N \boldsymbol{K}(t_0 - t_N) \\ -\kappa w_0 \boldsymbol{K}(t_1 - t_0) & I - \kappa w_1 \boldsymbol{K}(0) & \cdots & -\kappa w_N \boldsymbol{K}(t_1 - t_N) \\ \vdots & \vdots & \ddots & \vdots \\ -\kappa w_0 \boldsymbol{K}(0) & -\kappa w_1 \boldsymbol{K}(t_N - t_1) & \cdots & I - \kappa w_N \boldsymbol{K}(0) \end{pmatrix} \begin{pmatrix} \boldsymbol{\phi}(t_0) \\ \boldsymbol{\phi}(t_1) \\ \vdots \\ \boldsymbol{\phi}(t_N) \end{pmatrix} = \begin{pmatrix} \boldsymbol{F}(t_0) \\ \boldsymbol{F}(t_1) \\ \vdots \\ \boldsymbol{F}(t_N) \end{pmatrix}. \tag{5.52}$$

This is a $M(N+1)$ by $M(N+1)$ matrix equation in

$$\begin{pmatrix} \boldsymbol{\phi}(t_0) \\ \boldsymbol{\phi}(t_1) \\ \vdots \\ \boldsymbol{\phi}(t_N) \end{pmatrix},$$

and can be readily solved by standard techniques.

In particular, using the trapezoïdal rule with step size $h > 0$ as the quadrature rule, and the original notations of the HWH equation (5.24)

$$\boldsymbol{F}(x) \quad = \quad \boldsymbol{K}(x), \tag{5.53}$$

$$\kappa \quad = \quad -1, \tag{5.54}$$

the linear system reads

$$\begin{pmatrix} \boldsymbol{I} + \frac{h}{2}\boldsymbol{K}(0) & h\boldsymbol{K}(t_0 - t_1) & \cdots & \frac{h}{2}\boldsymbol{K}(t_0 - t_N) \\ \frac{h}{2}\boldsymbol{K}(t_1 - t_0) & I + h\boldsymbol{K}(0) & \cdots & \frac{h}{2}\boldsymbol{K}(t_1 - t_N) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{h}{2}\boldsymbol{K}(t_N - t_0) & h\boldsymbol{K}(t_N - t_1) & \cdots & I + \frac{h}{2}\boldsymbol{K}(0) \end{pmatrix} \begin{pmatrix} \boldsymbol{\phi}(t_0) \\ \boldsymbol{\phi}(t_1) \\ \vdots \\ \boldsymbol{\phi}(t_N) \end{pmatrix} = \begin{pmatrix} \boldsymbol{K}(t_0) \\ \boldsymbol{K}(t_1) \\ \vdots \\ \boldsymbol{K}(t_N) \end{pmatrix}. \quad (5.55)$$

Note that, to evaluate $\boldsymbol{K}$ for negative t, we can use property (5.8)

$$\boldsymbol{K}(-x) = \boldsymbol{K}^\top(x).$$

The value of t_N, which determines the domain of integration, should be chosen large enough for the approximation of the integral to be precise. In practise, we choose it such that the integral of the empirical density (or its rescaled transpose $\boldsymbol{K}$) reaches a plateau before t_N.

Because of the unbounded integration domain, and the fact that we are dealing with a system of integral equations, existence and uniqueness of solutions to the HWH equation are *a priori* not guaranteed. Nor are the stabiliy and convergence of the proposed numerical method, obvious. In practice, the method is successful on a wide range of datasets, but fails for some "difficult" cases (more on this below.) A systematic study of the properties of equation (5.24) and its numerical solution is therefore warranted[5]. These considerations are, however, beyond the scope of this thesis.

5.4. Empirical Covariance Density. Estimating the empirical covariance density matrix is quite straightforward. One chooses a small sampling interval $\Delta > 0$, then sets

$$\widehat{C}_{ij}(x) = \frac{1}{\Delta^2}\mathrm{Cov}\left[N_i(t + x + \Delta) - N_i(t + x), N_j(t + \Delta) - N_j(t)\right], \quad (5.56)$$

Cov being the sample's empirical covariance. Our experimentation with market data shows that this estimate is stable, and for reasonably small Δ, independent of the sampling interval. We note that the Fast Fourier Transform [**PTVF07**] can be used to speed up the evaluation of equations (5.56).

5.5. Estimating the Base Rate $\boldsymbol{\mu}$. Once $\boldsymbol{\varphi}$ has been estimated as the (numerical) solution to equation (5.24), we make use of the identity

$$\boldsymbol{\mu} = \left(\boldsymbol{I} - \int_0^\infty \boldsymbol{\varphi}(t)dt\right)\widehat{\boldsymbol{\lambda}}, \quad (5.57)$$

[5]Functional transforms might be the right path to this end [**Haw71a**].

to find $\boldsymbol{\mu}$.

5.6. Principle of the Non-parametric Kernel Estimation Method.

(1) Estimate the empirical mean intensity

$$\widehat{\lambda}_i = \frac{\#\text{ Jumps in } N_i \text{ in } [0,T]}{T}, \qquad i \in [\![1, M]\!]. \tag{5.58}$$

(2) Estimate the empirical covariance density $\widehat{\boldsymbol{C}}$.
(3) Rescale it to $\boldsymbol{K} = \widehat{\boldsymbol{C}}\,\mathbf{Diag}(\widehat{\boldsymbol{\lambda}})^{-1}$.
(4) Solve the HWH equation for $\boldsymbol{\phi} = \boldsymbol{\varphi}^\top$.
(5) Set

$$\boldsymbol{\mu} = \left(\boldsymbol{I} - \int_0^\infty \boldsymbol{\varphi}(t)dt\right)\widehat{\boldsymbol{\lambda}}. \tag{5.59}$$

5.7. Illustration. Figure 6.

6. Simulation *via* Markovian Projection

In this section, a (new) technique for the fast simulation of multivariate Hawkes processes with arbitrary decay kernels is presented.

6.1. Ogata's Thinning Simulation Method. The standard method for the simulation of multivariate Hawkes processes is Ogata's modified thinning algorithm [**Oga81**] (See also Lewis and Schedler [**LS79**].) The idea is to simulate a non-homogeneous univariate Poisson process $\{t_n^*\}$ whose intensity $\lambda^*(t)$ dominates the total intensity

$$\lambda^{\text{sum}}(t) = \sum_{i=1}^{M} \lambda_i(t), \tag{6.1}$$

that is,

$$\lambda^*(t) \geq \lambda^{\text{sum}}(t), \text{ forall } t, \tag{6.2}$$

and attach a mark $j_0 \in [\![1, M+1]\!]$ to each point according to the probabilities

$$\left(\frac{\lambda_1}{\lambda^*}, \frac{\lambda_2}{\lambda^*}, \ldots, \frac{\lambda_M}{\lambda^*}, \frac{\lambda^* - \lambda^{\text{sum}}}{\lambda^*}\right). \tag{6.3}$$

If the mark j_0 equals $M + 1$, the point is *rejected*. By proposition 1 in [**Oga81**], the remaining points form a point-process that has the desired intensity $\boldsymbol{\lambda}(t)$.

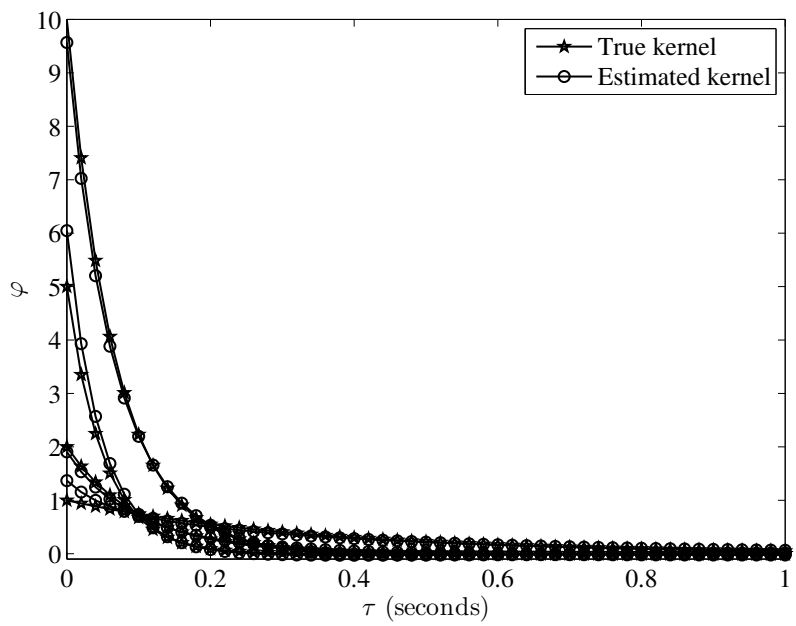


FIGURE 5. Non-parametric kernel estimation. Artificial data, 10000 points sample, is generated with the parameters $M = 2$, $P = 1$, $\boldsymbol{\mu} = (0.1, 0.2)^\top$, $\boldsymbol{\alpha} = \begin{pmatrix} 5.0 & 10.0 \\ 1.0 & 2.0 \end{pmatrix}$, $\boldsymbol{\beta} = \begin{pmatrix} 20.0 & 15.0 \\ 3.0 & 10.0 \end{pmatrix}$. The covariance density matrix is then estimated, and the Hawkes-Wiener-Hopf equation solved by Nystöm's method. The number of integration nodes is $N = 3000$ and $t_N = 60$ seconds. The estimated base intensity is $\widehat{\boldsymbol{\mu}} = (0.0967, 0.1932)^\top$.

For Hawkes processes, since the kernel $\boldsymbol{\varphi}(t)$ is decreasing (and therefore $\lambda^{\text{sum}}(t)$ after a jump), a natural choice for the dominant intensity $\lambda^*(t)$ is the value of λ^{sum} immediately after a jump. One therefore gets a piecewise constant λ^* between jumps (figure 8).

We implement this idea in algorithm 2. Note that we heavily exploit the Markovian structure of the point process (chapter 4, proposition 2.2). We comment on this below.

Recall that

$$\boldsymbol{\alpha} = \left(\alpha_{ij}^p\right)_{1 \le i,j \le M, 1 \le p \le P}, \tag{6.4}$$

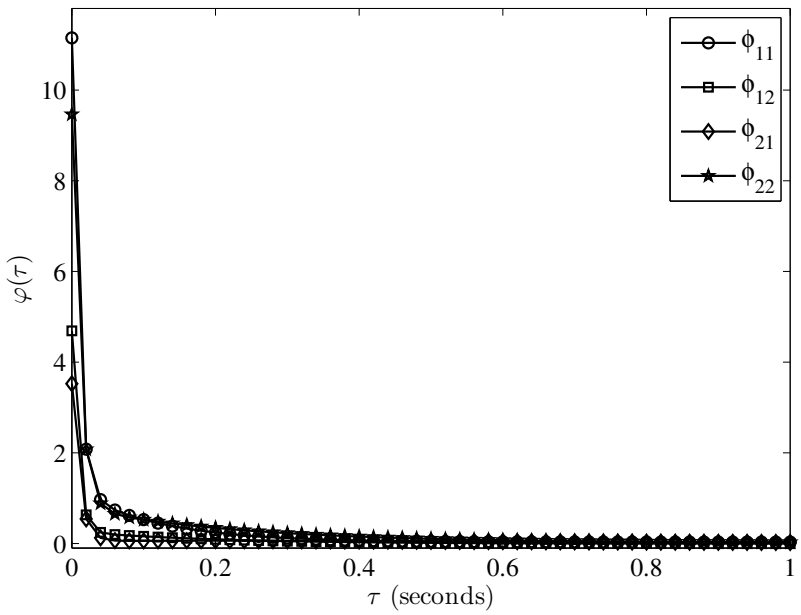


FIGURE 6. Non-parametric kernel estimation (Buy and sell market orders, TOTF.PA, June 2011.)

$$\boldsymbol{\beta} = \left(\beta_{ij}^{p}\right)_{1\leq i,j\leq M, 1\leq p\leq P}, \tag{6.5}$$

are three dimensional matrices. i, j, p are respectively the row, column and page index. The three dimensional matrix defined as

$$Y_{ij}^{p}(t) \quad = \quad \int_0^t e^{-\beta_{ij}^{p}(t-s)} dN_j(s), \qquad 1 \leq i, j \leq M, 1 \leq p \leq P. \tag{6.6}$$

is the workhorse of the algorithm.

$\boldsymbol{Y}$ jumps by 1 with each jump of the process. More precisely, if $z_n \in [\![1, M]\!]$ is the mark of the jump,

$$Y_{iz_n}^{p}(t_{n+1}) \quad = \quad Y_{iz_n}^{p}(t_{n+1}) + 1, \qquad 1 \leq i \leq M, 1 \leq p \leq P. \tag{6.7}$$

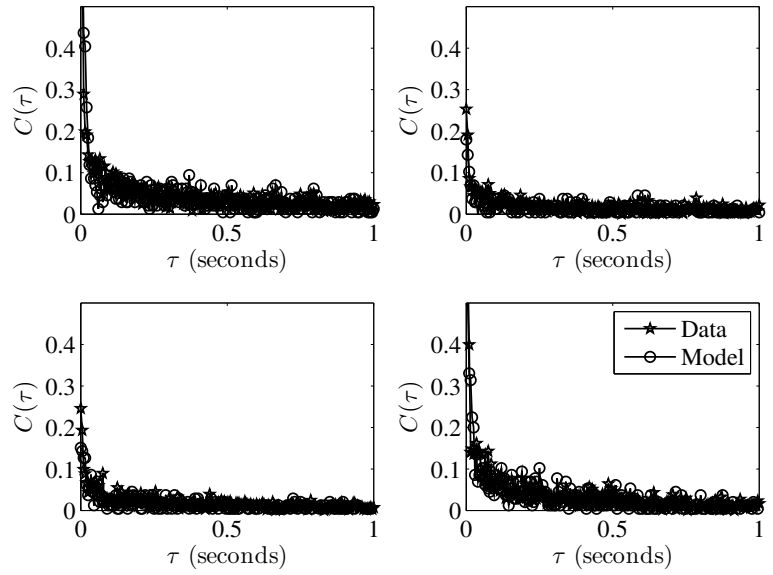
Only a "slice" of $\boldsymbol{Y}$ jumps by 1, it corresponds the plan $j = z_n$ (figure 9).

Betwen t_n and t_{n+1}, $\boldsymbol{Y}$ decays as[6]

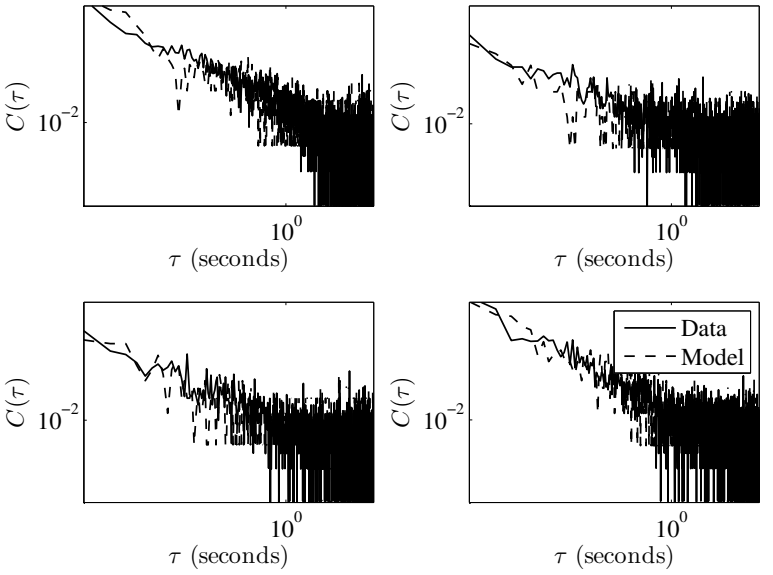
$$\boldsymbol{Y}(t_{n+1}) \quad = \quad e^{-\boldsymbol{\beta}(t_{n+1}-t_n)} \boldsymbol{Y}(t_{n+1}). \tag{6.8}$$

These are the key recursive relationships we use in the simulation.

[6]The exponential and multiplication are meant element by element.



(A)



(B)

FIGURE 7. Data versus model covariance density (Buy and sell market orders, TOTF.PA, June 2011.)

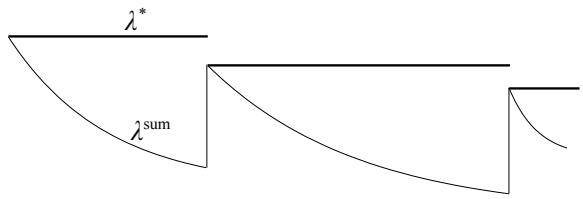


FIGURE 8. Schematic representation of the dominant intensity $\lambda^*(t)$.

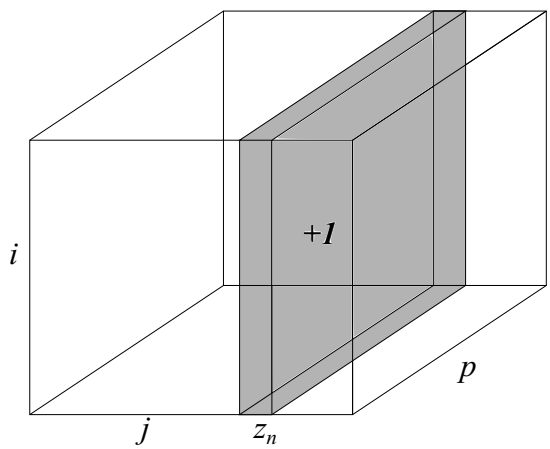


FIGURE 9. Illustration for equation (6.7).

6.2. General Kernel. Suppose now we are given a general decay kernel $\boldsymbol{\varphi}$, a power-law,

$$\varphi_{ij}(t) = \frac{a_{ij}}{(b_{ij} + t)^{c_{ij}}}, \qquad a_{ij}, b_{ij}, c_{ij} > 0, \tag{6.9}$$

say. The problem at hand is not Markovian. But it can be very well approximated by one. The family of functions

$$\left\{\alpha e^{-\beta t}\right\} \tag{6.10}$$

is rich enough to virtually approximate any kernel of interest (on a finite but arbitrarily large interval $[0, T]$ and with arbitrary accuracy). The idea is therefore to project the kernel φ on the "basis functions",

$$\left\{\alpha e^{-\beta t}\right\} \tag{6.11}$$

for instance by least squares fitting, then use algorithm 2 to simulate the process.

Algorithm 2 Multivariate Hawkes process simulation.

Require: $N, M, P, \boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\beta}$

$n \leftarrow 1$
$\boldsymbol{Y} \leftarrow \mathbf{0}$
$\lambda^* \leftarrow \sum_{i=1}^{M} \mu_i$
while $n \leq N$, **do**
  Generate the inter-event time according to an exponential r.v. with mean λ^* $\tau \sim \mathrm{Exp}(1/\lambda^*)$
  $newT \leftarrow newT + \tau$
  $\boldsymbol{Y} \leftarrow e^{-\beta\tau}\boldsymbol{Y}$ {eq. 6.8}
  for $i = 2$ to M **do**
    $\lambda_i = \mu_i + \sum_{1 \leq j \leq M, 1 \leq p \leq P} \alpha_{ij}^p Y_{ij}^p$
  end for
  $\lambda^{\mathrm{sum}} = \sum_{1 \leq i \leq M} \lambda_i$
  Choose randomly the event type j_0 in $[\![1, M+1]\!]$ according to the probability vector

$$(\lambda_1, \lambda_2, \ldots, \lambda_M, \lambda^* - \lambda^{\mathrm{sum}}) / \lambda^*$$

  if $j_0 \leq M$ **then**
    {*Keep the point*}
    $t_n \leftarrow newT$
    $z_n \leftarrow j_0$
    $\boldsymbol{Y}(:, j_0, :) \leftarrow \boldsymbol{Y}(:, j_0, :) + 1$ {*This is a shorthand notation of eq.* (6.7).}
    $\lambda^* \leftarrow \sum_{1 \leq i \leq M} \mu_i + \sum_{1 \leq i, j \leq M, 1 \leq p \leq P} \alpha_{ij}^p Y_{ij}^p$
    $n \leftarrow n + 1$
  else
    {*Reject the point*}
    $\lambda^* \leftarrow \lambda^{\mathrm{sum}}$
  end if
end while

Interestingly, in the case of power-law decaying kernels, this has a nice financial interpretation. In econophysics, power-law behaviour is known to emerge from the aggregation of heterogeneous agents with different time

	a	b	c
φ_{11}	1.1678	2.0049	3.1901
φ_{12}	1.1014	2.0511	3.1021
φ_{21}	1.1591	2.0395	3.1803
φ_{22}	1.0822	2.0690	3.0730

TABLE 1. Power-law kernel parameters. Each triplet (a, b, c) was randomly generated by $(1, 2, 3) + 0.2(U_1, U_2, U_3)$, with $U_i \sim$ Uniform$]0, 1[$.

horizons. Each term in the exponential sum

$$\frac{a}{(b+t)^c} \approx \sum_{p=1}^{P} \alpha_p e^{-\beta_p t}, \tag{6.12}$$

represents an agent (or group of agents) with a characteristic time of $\frac{1}{\beta_p}$. In a sense, Markovian projection identifies the most important groups. Remarkably, at least from a computing perspective, you only need about $\log T$ exponentials to approximate a power-law over the range $[0, T]$. This is the key insight behind the algorithm.

Note that although we restricted the parameters to be positive, we can imagine them taking complex values (for fancier shaped kernels), exploit the Markovian structure to simulate, then return to the real line.

The speedup, from $O(N^2)$ to $O(N \log N)$, is dramatic (figure 12).

7. Applications

In the previous sections, it was shown how to estimate and efficiently simulate a stochastic order book model based on Hawkes processes. We now turn to the measurement of two quantities of practical importance using the estimated model. The flexibility of the order book model allows the assessment of a wide range of algorithmic trading strategies, and this section is a proof-of-concept for the possibilities such models may offer.

7.1. Probability of Fill of a Limit Order. The first quantity of interest is the *probability of fill* of a limit order. Its estimation from market data can be challenging (See e.g. [**LMZ02**]) and we propose an indirect (and

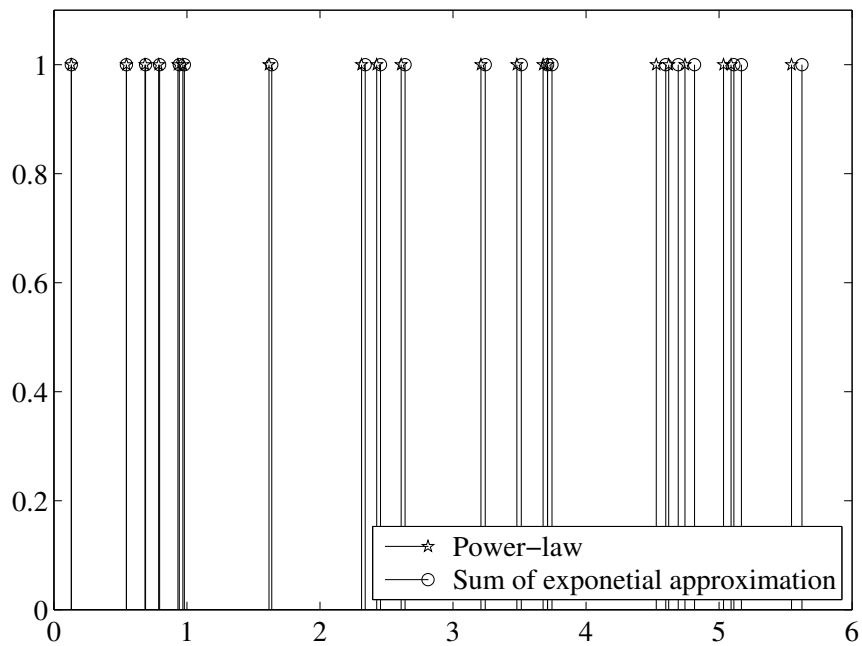


FIGURE 10. Jump times comparison.

original) approach for its estimation. To do this, we simulate an order book model with 7 order flows:

(1) $M^+(t)$: buy market orders,
(2) $M^-(t)$: sell market orders,
(3) $L^+(t)$: sell limit orders,
(4) $L^-(t)$: buy limit orders,
(5) $L_A^+(t)$: *An additional agent's sell limit orders*,
(6) $C^+(t)$: cancellation of a sell limit order,
(7) $C^-(t)$: cancellation of a buy limit order.

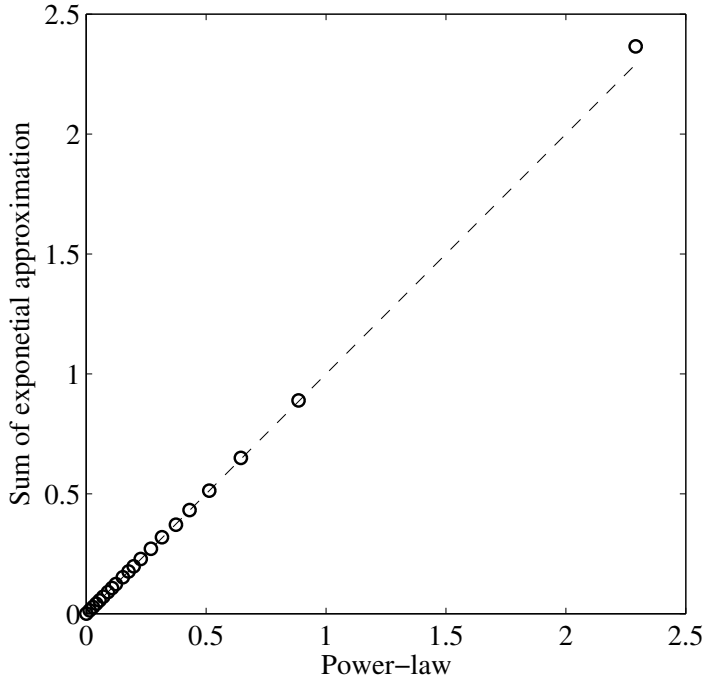


FIGURE 11. Interarrival durations comparison (Q-Q plot).

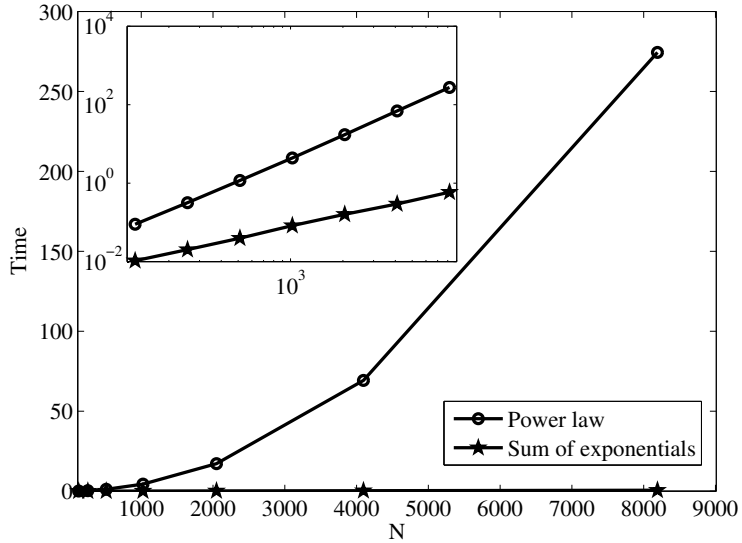


FIGURE 12. Timing comparison. The time axis is in arbitrary units.

The first 5 order flows form a Hawkes process whose decay kernel matrix reads

$$\boldsymbol{\varphi} = \begin{pmatrix} \varphi_{11} & \varphi_{12} & \varphi_{13} & \varphi_{14} & \underline{\varphi_{13}} \\ \varphi_{21} & \varphi_{22} & \varphi_{23} & \varphi_{24} & \underline{\varphi_{23}} \\ \varphi_{31} & \varphi_{32} & \varphi_{33} & \varphi_{34} & \underline{\varphi_{33}} \\ \varphi_{41} & \varphi_{42} & \varphi_{43} & \varphi_{44} & \underline{\varphi_{43}} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{7.1}$$

Note that we account for the effect of the agent's trading on the order flow (underlined terms). This would be impossible when merely replaying market data. The agent is submitting sell limit orders at a constant Poissonian rate, and affecting the order flow similarly to any other sell limit order. The result of the simulation is displayed in figure 13.

7.2. Execution Shortfall of a Limit Order. The second quantity is related to the cost of executing a limit order [**NKPS05**]. Consider the so-called "submit and leave" strategy: A agent submits a limit order at price P with a certain volume v, waits T seconds for it to be (partially) executed, and finally submits a market order with the remaining shares. A question of interest is the determination of the optimal price and waiting time for such a simple strategy.

If we define the (algebraic) *shortfall* of this strategy (started at t_0) as[7]

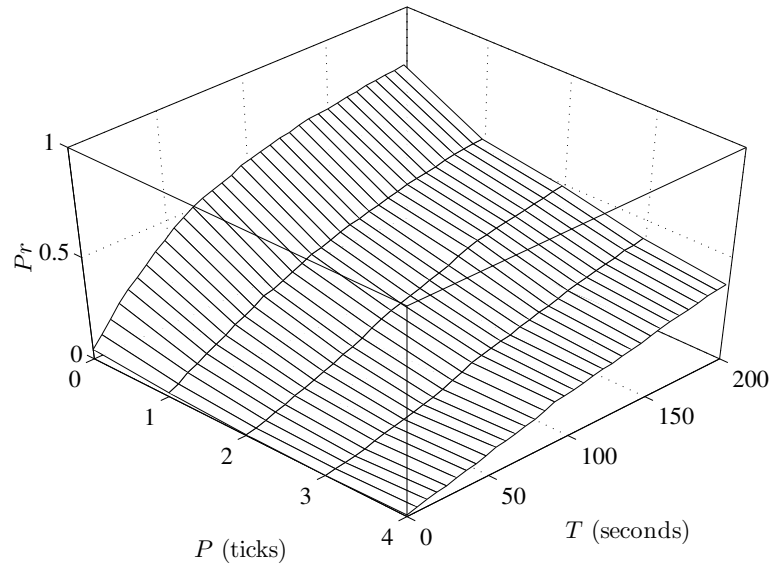
$$S = \text{Execution Price} - \text{Prevailing mid-price at } t_0, \tag{7.3}$$

expressed in ticks. Then estimates of the average shortfall $\langle S \rangle$ and its variance are valuable indicators for the optimal choice. We run this strategy on simulated data and display the result in figure 14. For instance, it is clear, for figure 14's parameter set, that it is not optimal to place the order deep in the book.

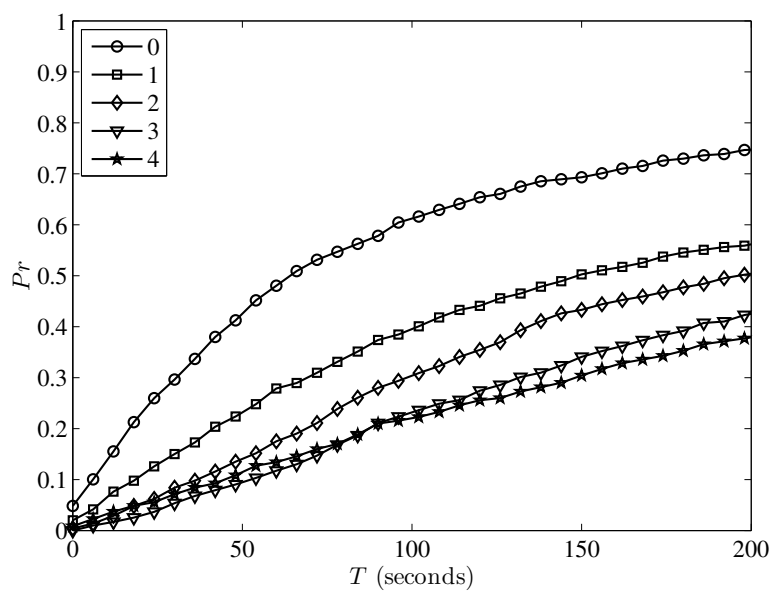
[7]It is also possible to normalize by the mid-price

$$S = \frac{\text{Execution Price} - \text{Prevailing mid-price at } t_0}{\text{Prevailing mid-price at } t_0}, \tag{7.2}$$

to get the "return" of the strategy.

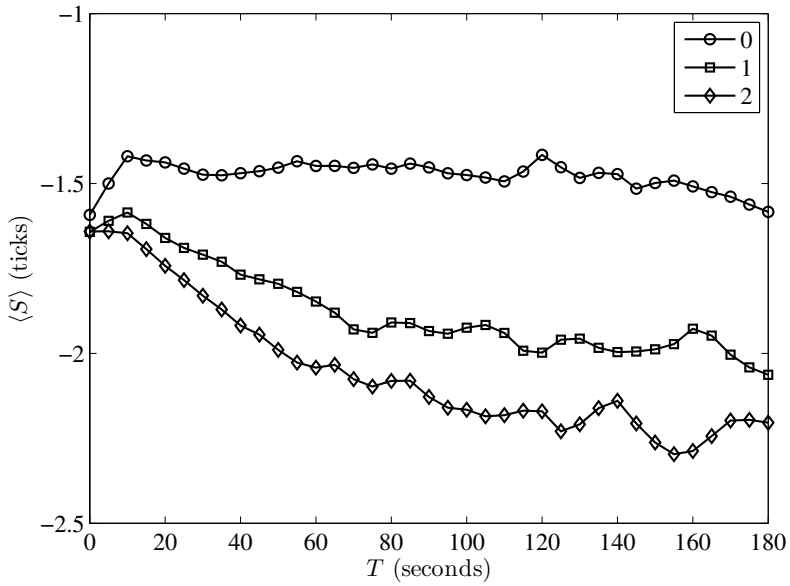


(A)

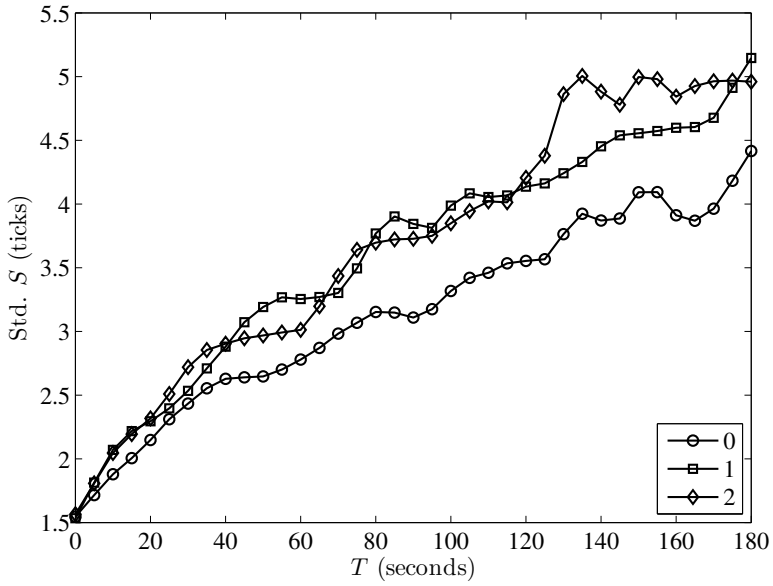


(B)

FIGURE 13. Probability of fill of a sell limit order of size $v = 50$ shares. In panel (B), the legend indicates the price in ticks relative to the best offer. For instance, 0 means that the limit order is placed with its price set to the best ask, 1 at ask+1 tick etc. Model parameters are estimated for the stock TOTF.PA, June 2011.



(A)



(B)

FIGURE 14. Execution shortfall (A) and its standard deviation (B) for a limit order of size $v = 1000$ shares. Model paramters are estimated for the stock TOTF.PA, June 2011.

Bibliography

[AJ13] F. Abergel and A. Jedidi, *A mathematical approach to order book modeling*, International Journal of Theoretical and Applied Finance (2013).

[ASCDL13] Y. Aït-Sahalia, J. Cacho-Diaz, and R.J.A. Laeven, *Modelling financial contagion using mutually exciting Hawkes processes*, Preprint (2013).

[BBVK01] E. Brown, R. Barbieri, V. Ventura, and R.E. Kass, *The time-rescaling theorem and its application to neural spike train data analysis*, Neural Computation (2001).

[BDHM13a] E. Bacry, S. Delattre, M. Hoffmann, and J.F. Muzy, *Modelling microstructure noise with mutually exciting point processes*, Quantitative Finance (2013).

[BDHM13b] _____, *Scaling limits for Hawkes processes and application to financial statistics*, Stochastic Processes and Applications (2013).

[BDM12] E. Bacry, K. Dayri, , and J.-F. Muzy, *Non-parametric kernel estimation for symmetric Hawkes processes. Application to high frequency financial data*, Eur. Phys. J. B (2012).

[BFL09] J.-P. Bouchaud, D.J. Farmer, and F. Lillo, *How markets slowly digest changes in supply and demand*, Handbook of Financial Markets: Dynamics and Evolution (2009).

[BGPW04] J.-P. Bouchaud, Y. Gefen, M. Potters, and M. Wyart, *Fluctuations and response in financial markets: The subtle nature of random price changes*, Quantitative Finance (2004).

[Bil99] P. Billingsley, *Convergence of Probability Measures*, Wiley, 2nd ed., 1999.

[BM96] P. Brémaud and L. Massoulié, *Stability of nonlinear Hawkes processes*, The Annals of Probability (1996).

[BM13] E. Bacry and J.F. Muzy, *Hawkes model for price and trades high-frequency dynamics*, Preprint (2013).

[BMP02] J.-P. Bouchaud, M. Mézard, and M. Potters, *Statistical properties of stock order books: empirical results and models*, Quantitative Finance (2002).

[Bow07] C.G. Bowsher, *Modelling security market events in continuous time: intensity based, multivariate point process models*, Journal of Econometrics (2007).

[CdL12] R. Cont and A. de Larrard, *Order book dynamics in liquid markets: limit theorems and diffusion approximations*, Preprint (2012).

[CdL13] _____, *Price dynamics in a Markovian limit order book market*, SIAM Journal for Financial Mathematics (2013).

[CI80] D.R. Cox and V. Isham, *Point Processes*, Chapman & Hall/CRC, 1980.

[CS01] D. Challet and R. Stinchcombe, *Analyzing and modeling* $1 + 1d$ *markets*, Physica A (2001).

[CS03] _____, *Non-constant rates and over-diffusive prices in a simple model of limit order markets*, Quantitative Finance (2003).

[CST10] R. Cont, S. Stoikov, and R. Talreja, *A stochastic model for order book dynamics*, Operations Research (2010).

[DFG+03] M.G. Daniels, J.D. Farmer, L. Gillemot, G. Iori, and E. Smith, *Quantitative model of price diffusion and market friction based on trading as a mechanistic random process*, Physical Review Letters (2003).

[DM85] L.M. Delves and J. Mohamed, *Computational Methods for Integral Equations*, Cambridge University Press, 1985.

[DVJ03] D.J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes*, Springer, 2nd ed., 2003.

[Eco11] *Computing power and stockmarkets: Moore and more*, The Economist (2011).

[EGG10] E. Errais, K. Giesecke, and L. Goldberg, *Affine point processes and portfolio credit risk*, SIAM Journal on Financial Mathematics (2010).

[Fer08] A. Ferraris, *Equity market impact models*, Presentation, 2008.

[FGLM06] J.D. Farmer, A. Gerig, F. Lillo, and S. Mike, *Market efficiency and the long-memory of supply and demand: Is price impact variable and permanent or fixed and temporary?*, Quantitative Finance (2006).

[FPZ05] J.D. Farmer, P. Patelli, and I. Zovko, *The predictive power of zero-intelligence in financial markets*, PNAS (2005).

[GO10] Gatheral and R. Oomen, *Zero-intelligence realized variance estimation*, Finance and Stochastics (2010).

[Har02] L. Harris, *Trading and Exchanges: Market Microstructure for Practitioners*, Oxford University Press, 2002.

[Haw71a] A. Hawkes, *Point spectra of some mutually exciting point processes*, Journal of the Royal Statistical Society (1971).

[Haw71b] _____, *Spectra of some self-exciting and mutually exciting point processes*, Biometrika (1971).

[HO74] Hawkes and D. Oakes, *A cluster process representation of a self-exciting process*, Journal of Applied Probability (1974).

[HP13] U. Horst and M. Paulsen, *A law of large numbers for limit order books*, Preprint (2013).

[Jac75] J. Jacod, *Multivariate point processes: predictable projections, Radon-Nikodym derivatives, representation of martingales*, Z. Wahrsch. Verw. Gebiete (1975).

[KL13] A.A. Kirilenko and A.W. Lo, *Moore's law vs. Murphy's law: algorithmic trading and its discontents*, Journal of Economic Perspectives (2013).

[Lar07] J. Large, *Measuring the resiliency of an electronic limit order book*, Journal of Financial Markets (2007).

[LF04] F. Lillo and J.D. Farmer, *The long memory of the efficient market*, Studies in Nonlinear Dynamics & Econometrics (2004).

[LMZ02] A.W. Lo, A.C. MacKinlay, and J. Zhang, *Econometric models of limit-order executions*, Journal of Financial Economics (2002).

[LS79] P.A.W. Lewis and G.S. Shedler, *Simulation of nonhomogeneous Poisson processes by thinning*, Naval Res. Logistics Quart. (1979).

[Mas00] S. Maslov, *Simple model of a limit order-driven market*, Physica A (2000).

[MT92] S. Meyn and R.L. Tweedie, *Stability of Markovian processes I: criteria for discrete-time chains*, Advances in Applied Probability (1992).

[MT93a] _____, *Stability of Markovian processes II: continuous-time processes and sampled chains*, Advances in Applied Probability (1993).

[MT93b] _____, *Stability of Markovian processes III: Foster-Lyapunov criteria for continuous-time processes*, Advances in Applied Probability (1993).

[MT09] _____, *Markov Chains and Stochastic Stability*, Cambridge University Press, 2nd ed., 2009.

[MT11a] I. Muni-Toke, *An introduction to Hawkes processes with applications to finance*, Working paper (2011).

[MT11b] _____, *Market making behavior in an order book model and its impact on the spread*, Econophysics of Order-driven Markets (2011).

[MTP12] I. Muni-Toke and F. Pomponio, *Modelling trades-through in a limit order book using Hawkes processes*, Economics: The Open-Access Open-Assessment E-Journal (2012).

[NKPS05] Y. Nevmyvaka, M. Kearns, A. Papandreou, and K. Sycara, *Electronic trading in order-driven markets: efficient execution*, IEEE Conference on Electronic Commerce (CEC) (2005).

[Oga78] Y. Ogata, *The asymptotic behavior of maximum likelihood estimators for stationary point processes*, Ann. Inst. Statist. Math. (1978).

[Oga81] _____, *On Lewis' simulation method for point processes*, IEEE Transactions on Information Theory (1981).

[Oza79] T. Ozaki, *Maximum likelihood estimation of Hawkes self-exciting point processes*, Ann. Inst. Statist. Math. (1979).

[PGPS06] T. Preis, S. Golke, W. Paul, and J.J. Schneider, *Multi-agent-based order book model for financial markets*, Europhysics Letters (2006).

[PS08] C.A. Parlour and D. Seppi, *Limit order markets: a survey*, Handbook of Financial Intermediation & Banking (2008).

[PTVF07] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes, The Art of Scientific Computing*, Cambridge University Press, 3nd ed., 2007.

[Rub72] I. Rubin, *Regular point processes and their detection*, IEEE Transactions on Information Theory (1972).

[SFGK03] E. Smith, J.D. Farmer, L. Guillemot, and S. Krishnamurthy, *Statistical theory of the continuous double auction*, Quantitative Finance (2003).

[SM91] D. Snyder and M. Miller, *Random point processes in time and space*, Springer-Verlag, 1991.

[TLD+11] B. Tóth, Y. Lempérière, C. Deremble, J. de Lataillade, J. Kockelkoren, and J.-P. Bouchaud, *Anomalous price impact and the critical nature of liquidity in financial markets*, Physical Review X (2011).

[Whi02] W. Whitt, *Stochastic-Process Limits*, Springer, 2002.

[ZRA13] B. Zheng, F. Roueff, and F. Abergel, *Ergodicity and scaling limit of a constrained multivariate Hawkes process*, Preprint (2013).