High-frequency financial data modelling with hybrid marked point processes

A thesis presented for the degree of

Doctor of Philosophy of Imperial College London

and the

Diploma of Imperial College

by

Maxime Morariu-Patrichi

Department of Mathematics

Imperial College

180 Queen's Gate, London SW7 2BZ

October 2018

I certify that this thesis, and the research to which it refers, are the product of my own work, and that any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline.

Signed: Maxime Morariu-Patrichi

Copyright

Thesis advisor: Mikko S. Pakkanen Author: Maxime Morariu-Patrichi

High-frequency financial data modelling with hybrid marked point processes

Abstract

The rise of electronic order-driven financial markets has brought a profusion of new high-frequency data to study, with an opportunity to understand the price formation mechanism at the smallest timescales. The original motivation of this thesis is to find a stochastic process that provides an accurate statistical dynamic description of this new data.

A critical analysis of the literature reveals a dichotomy between two main sorts of model, Hawkes processes and continuous-time Markov chains, each having qualities that the other lacks. In particular, models of the former sort are successful at capturing excitation effects between different event types but fail to incorporate the state of the market. We resolve this dichotomy by introducing state-dependent Hawkes processes, an extension of Hawkes processes where events can now interact with an auxiliary state process. These new stochastic processes provide us with the first model that features both excitation effects and an explicit feedback loop between events and the state of the market. The application of this new model to high-quality data demonstrates that the excitation effects are indeed strongly state-dependent.

State-dependent Hawkes processes come however with theoretical challenges: under which conditions do they exist, are they unique and do not explode? To answer these questions, we view state-dependent Hawkes processes as ordinary point processes of higher dimension, which we then generalise to the class of hybrid marked point processes. This class provides a framework that unifies and extends the existing high-frequency models. Since hybrid marked point processes are defined implicitly via their intensity, one can address the above questions by studying instead a Poisson-driven stochastic differential equation (SDE). We are able to solve this SDE under general assumptions that dispense with the Lipchitz condition usually required in the literature, which yields, as a corollary, the existence and uniqueness of non-explosive state-dependent Hawkes processes.

To Anisia.

Acknowledgments

First and foremost, I would like to thank Mikko for being my supervisor during four years. Growing under his guidance and working with him has been a real pleasure. In addition to his patience, support and availability, I would like to praise him for the research approach that we followed together. Thank you Mikko.

I would also like to thank several people who provided me with some helpful, stimulating and encouraging feedback throughout these years: Rama Cont, Fabrizio Lillo, Charles-Albert Lehalle, Jean-Philippe Bouchaud, Olav Kallenberg, Daryl Daley and Nicholas Bingham.

Last but not least, I would like to thank my family for their support and their understanding of my decision to pursue a PhD.

Contents

List of figures

List of tables

0

Introduction

0.1 Area of study

The last twenty years have seen the rise of electronic order-driven financial markets where agents submit anonymised buy and sell orders to a virtual exchange via their computers. In these new markets, buyers and sellers trade directly with one another at the prices they specify. This is the antonym of quote-driven markets where traders can only exchange securities with dedicated dealers who control the buy and sell prices.

In order-driven markets, the collection of outstanding orders is called the limit order book (LOB) and is made visible to market participants. The adoption of this order-driven trading mechanism by major stock exchanges has brought a profusion of new intraday high-frequency data to study, with an opportunity to understand the price formation mechanism at the smallest timescales. A new area of research emerged: modelling the dynamics of LOBs.

This field has turned out to be multidisciplinary, with contributions from (econo)physicists, economists, mathematicians, statisticians, computer scientists and regulators. It lends itself to different, yet complementing, approaches, from purely empirical studies to theoretical and abstract considerations. For mathematicians and statisticians, LOB modelling is an open problem that can be tackled with many different kinds of tools, which include: stochastic processes, statistical inference, machine learning, game theory, mean-field games and partial differential equations. In fact, LOBs can even be a source of inspiration and a catalyst for new breakthroughs in these areas.

0.2 Main contributions

This thesis tackles the LOB modelling problem from the angle of stochastic processes and statistical inference. The original motivation is to find a stochastic process that provides an accurate statistical dynamic description of the observed high-frequency data, that is, the sequence of orders and the corresponding states of the LOB. This contrasts with another approach, known as agent-based modelling, where one tries to explain how the key (statistical) features of the data derive from the behaviour of market participants.

A critical analysis of the literature reveals a dichotomy between two main classes of stochastic processes: Hawkes processes and continuous-time Markov chains. On the one hand, Hawkes-process models account for self- and cross-excitation effects between multiple event types (i.e, orders with different effects on the market), but ignore the state of the underlying system they influence. On the other hand, Markov-process models focus on the state of the LOB, but their dynamics are exclusively driven by the LOB's current shape, meaning that past-dependent excitations as in Hawkes processes are not captured. In effect, Hawkes processes and continuous-time Markov chains can have either an *event* viewpoint or a *state* viewpoint, respectively.

To resolve this dichotomy, we introduce the class of state-dependent Hawkes processes, an extension of Hawkes processes where the events can now interact with an auxiliary state process. This class gives rise to the first non-Markovian LOB models in continuous-time that introduce a feedback loop between the order flow and the LOB's shape. We show how statistical inference for these new processes can be achieved via the maximum likelihood principle and create a public Python package, called `mpoints`, that implements simulation and estimation algorithms. An application to high-quality high-frequency financial data reveals that the excitation effects in the order flow are indeed sensitive to the LOB's shape and that the model is able to capture the *event-state* structure of LOBs. Moreover, we observe a striking state-dependent criticality, that is, intermittent high levels of endogeneity, with a market activity that increases when the LOB is in a state of disequilibrium.

This new class of state-dependent Hawkes processes comes with theoretical challenges. Because of the hybrid nature of their definition and the full coupling between the events and the state process, it is not clear how existence, non-explosion and uniqueness can be proved. To solve this problem, we lift the events and the auxiliary state process to a single point process of higher dimension. We explain how the dynamics of this larger point process are linked to the original state-dependent Hawkes process. This shift in viewpoint transposes the above theoretical challenges to the setting of point process theory. Under this alternative viewpoint, state-dependent Hawkes processes are naturally generalised to the class of hybrid marked point processes, which encompasses and extends continuous- time Markov chains and (state-dependent) Hawkes processes. This larger class provides a unifying and flexible framework for the joint modelling of events and the state of a system. While still maintaining the *event-state* structure of LOBs, the feedback loop between the events and the system can now take virtually any form.

To address the existence and uniqueness of non-explosive marked point processes that are specified implicitly via their intensity, we study a well-known Poisson-driven stochastic differential equation. The strong existence and uniqueness results that are available in the literature rely on a Lipschitz condition that hybrid marked point process may fail to satisfy. Consequently, we propose an intuitive pathwise construction that instead requires only a sublinear behaviour. Using a domination argument, we are able to verify that this construction indeed yields a solution. As an important corollary, we obtain the strong existence and uniqueness of hybrid marked point processes and, in particular, state-dependent Hawkes processes.

Finally, our theoretical framework for marked point processes relies on some fundamental properties of the so-called weak-hash metric, a distance function on a space of integer-valued measures. However, the original proofs of these properties assume that a certain term is monotonic, which is not always the case as we show by a counterexample. We manage to clarify these original proofs by addressing the parts that rely on this assumption and finding alternative arguments.

Because we test, investigate, develop and solidify the idea of state-dependent Hawkes processes at different levels of abstraction, from the purely empirical to the foundations of point process theory, this thesis is an illustration of how the field of LOB modelling stimulates the interplay between theory and practice.

0.3 Outline and chronology

In Chapter 1, we present the LOB mechanism in detail and survey the literature, with a focus on the application of stochastic processes. In Chapter 2, we conduct a preliminary empirical analysis that motivates the theoretical developments in the remainder of the thesis. In Chapter 3, we introduce the class of state-dependent Hawkes processes, explain how statistical inference can be achieved via maximum likelihood and present our Python package `mpoints`. In Chapter 4, state-dependent Hawkes processes are applied to high-frequency financial data. In Chapter 5, we set a theoretical framework for marked point processes, allowing us, in Chapter 6, to generalise state-dependent Hawkes processes to hybrid marked point processes. In Chapter 7, we address the existence and uniqueness of marked point processes that are defined implicitly via their intensity. In Chapter 8, we prove two fundamental properties of the weak-hash metric for boundedly finite integer-valued measures.

We chose to order the contents of this thesis from the concrete to the abstract, from the applied to the theoretical. However, the actual research was a back and forth between these two poles. The identification of the gap between Hawkes processes and continuous-time Markov chains (Chapter 1) and the formulation of state-dependent Hawkes processes (Chapter 3) were two intertwined actions. The preliminary data analysis of Chapter 2 was then carried to test the relevance of these new ideas. Our focus then shifted to the theoretical foundations of state-dependent Hawkes processes, generalising them to hybrid marked point processes and working on the existence and uniqueness problem (Chapters 5, 6 and 7). While doing so, we studied in detail the theory of point processes, which lead us to find and solve the problem of Chapter 8. Only then, we started to work on estimation and simulation (Chapter 3). The application of state-dependent Hawkes processes to high-frequency financial data (Chapter 4) was in fact the closing project of the PhD experience.

0.4 Three papers

The research behind this thesis resulted in the following three papers.

1. Morariu-Patrichi, M. and Pakkanen, M. S. (2017). Hybrid marked point processes: characterisation, existence and uniqueness. Preprint, available at: `http://arxiv.org/abs/1707.06970`.

2. Morariu-Patrichi, M. (2018). On the weak-hash metric for boundedly finite integer-valued measures. *Bulletin of the Australian Mathematical Society*, 98(2):265–276.

3. Morariu-Patrichi, M. and Pakkanen, M. S. (2018). State-dependent Hawkes processes and their application to limit order book modelling. Preprint, available at: `http://arxiv.org/abs/1809.08060`.

The first paper was split into Chapters 5, 6 and 7, while its appendix constitutes the last section of Chapter 8, the core of which consists of the second paper. The third paper was divided into Chapters 3 and 4.

1

Background on limit order book modelling

In this chapter, we present in detail the trading mechanism of LOBs and define key quantities that characterise them. We explain the motivations of LOB modelling and review the main empirical findings available in the literature. We then survey the use of stochastic processes and, in particular, identify a dichotomy between two of the prominent modelling approaches, that is, continuous-time Markov chains and Hawkes processes. In a sense, resolving this dichotomy will become the main purpose of this thesis, motivating the empirical analysis of Chapter 2 and leading to the introduction of state-dependent Hawkes processes in Chapter 3.

1.1 LOB mechanism

Our main reference for this section and the following is the survey on LOBs written by Gould et al. [55]. We adopt most of their notations.

Historically, a stock exchange was a physical location where participants traded via open outcry, that is, by shouting and using hand signals in the trading pit area [104]. With the arrival of telephone and electronic trading at the end of the 20th century, stock exchanges virtually became networks of agents interacting at distance. This transformation came with new trading mechanisms and a distinction between *quote-driven* and *order-driven* markets emerged [88, 110, 57]. In the former, the buy and sell prices are quoted by a group of competing dealers and market participants can only trade via this group of intermediaries. In the latter, market participants submit orders at the price of their choice and are matched directly with one another via a LOB mechanism that nowadays is used on more than half of markets [70]. In practice, the distinction between the two is not always clear with some major stock exchanges such as the New York Stock Exchange and the Nasdaq operating a hybrid system. For instance, the Nasdaq was originally a quote-driven exchange. While it now employs a LOB mechanism, there is a group of dedicated market making firms who are by contract obliged to constantly post buy and sell offers.

1.1.1 Order-matching algorithm

To explain the LOB mechanism, we first have to define what is meant by *an order*. In the following, it remains implicit that an order targets a specific financial security on a given exchange.

Definition 1.1.1 (Order)**.** An order $x = (q_x, p_x, t_x)$ of size $q_x > 0$ (respectively $q_x < 0$) with price p_x submitted at time t_x is a commitment to sell (respectively buy) up to $|q_x|$ units of the traded asset at a price no less (respectively no greater) than p_x.

Note that we adopt the convention that positive (respectively negative) quantities correspond to sell (respectively buy) orders. The price and size of an order cannot take any values but must be multiples of the resolution parameters, which can vary across assets and exchanges.

Definition 1.1.2 (Resolution parameters: lot size and tick size)**.** The resolution parameters are the lot size υ and the tick size ρ. For any order x,

q_x must be a multiple of υ, that is, $q_x \in \{k\upsilon : k \in \mathbb{Z}^*\}$, and p_x must be a multiple of ρ, that is, $p_x \in \{k\rho : k \in \mathbb{N}\}$.

We now explain how buy and sell orders are matched. When, say, a buy order x is submitted, if the set A of outstanding sell orders x' with prices $p_{x'} < p_x$ is not empty, x is matched with the orders $x' \in A$ that provide the smallest price(s) p' until the quantity q_x is traded or all the orders in A are matched. If A is empty or x was only partially matched (i.e., all the orders in A were exhausted), the order or its remaining part becomes *active*, meaning that it enters the list of outstanding orders. Besides, agents are allowed to cancel their active orders at any time. Note that a cancellation can be partial, that is, the order remains active but its size is decreased. Thus, an order is removed from the list of active orders either when it is totally matched or cancelled and the size of an active order is updated when it is partially matched or cancelled. The collection of active orders is called the limit order book (LOB).

Definition 1.1.3 (Active order and limit order book)**.** An active order is an order that has neither been matched nor cancelled yet. The limit order book (LOB) is the set of active orders and evolves in time as new orders are submitted and previous orders are cancelled. The LOB at time t is denoted by $\mathcal{L}_t$.

Notation 1.1.4 (State at and after submission)**.** For any order or cancellation x, we denote by $\mathcal{L}_{x-}$ (respectively $\mathcal{L}_x$) the state of the LOB just before (respectively right after) its submission. We carry over this notation to all time-varying quantities.

Remark 1.1.5. Strictly speaking, the above formalism does not allow to update the size q_x of an order $x = (q_x, p_x, t_x)$ since q_x is not time-dependent. However, this can be handled by noticing that updating the size of an order x to q'_x is equivalent to removing x and adding a new order $x' = (q'_x, p_x, t_x)$ as if it were submitted at time t_x.

To illustrate the matching algorithm and the previous definitions, we present an example of LOB and clarify how it evolves with the submission of new orders. To this end, we introduce a common way to represent the state of LOBs, that is, the depth profile.

Definition 1.1.6 (Depth profile)**.** The depth profile $D_t(\cdot)$ of the LOB $\mathcal{L}_t$ at time t is defined by

$$D_t(p) := \sum_{\{x \in \mathcal{L}_t \,:\, p_x = p\}} q_x, \quad p \in \mathbb{R}.$$

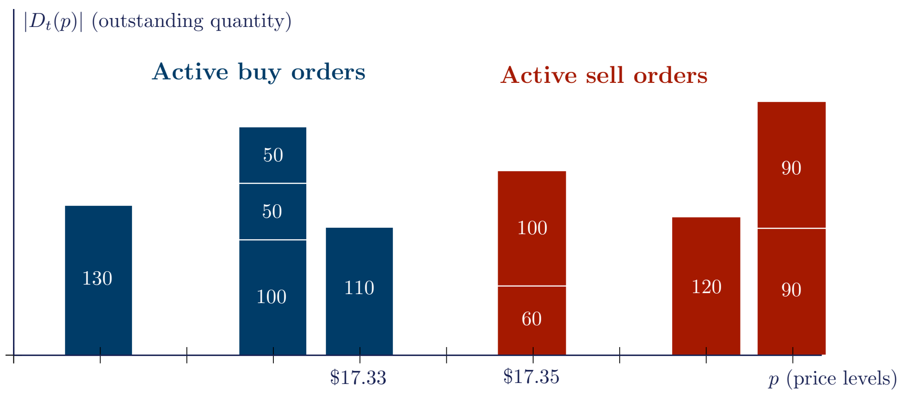


Figure 1.1: Fictional example of LOB represented by its depth profile. The sign of $D_t(p)$ can be read from the blue (negative) and red (positive) colours. The tick size is set to $\rho = \$0.01$. Each block corresponds to an active order. The number inside each block indicates the size of the order.

A fictional example of depth profile is shown in Figure 1.1, from which one can read the cumulative size of orders $|D_t(p)|$ at each price level p. In fact, this figure shows more than the depth profile since the size of each individual order is also indicated. Still, it does not allow to reconstruct $\mathcal{L}_t$ as it does not inform us on the submission times of orders. The depth profile informs us on the LOB's shape but does not fully characterise it. The blue (commitments to buy) and red (commitments to sell) regions are called the *bid* and *ask sides*, respectively. These two regions cannot overlap, otherwise it would mean that orders that should be matched were not matched.

Definition 1.1.7 (Ask and bid sides)**.** The ask side $\mathcal{A}_t$ (respectively bid side $\mathcal{B}_t$) of the LOB $\mathcal{L}_t$ is the set of active sell (respectively buy) orders at time t:

$$\mathcal{A}_t := \{x \in \mathcal{L}_t \,:\, q_x > 0\},$$

$$\mathcal{B}_t := \{x \in \mathcal{L}_t : q_x < 0\}.$$

Remark 1.1.8. The above definition obviously implies that $\mathcal{L}_t = \mathcal{A}_t \cup \mathcal{B}_t$.

Example 1.1.9. Assuming that the LOB $\mathcal{L}_t$ at time t has the depth profile of Figure 1.1, if the next order is $x_1 = (150, \$17.32, t + \Delta)$, $\Delta > 0$, and no cancellations occurs in the meantime, x_1 is immediately matched: 110 units are sold at \$17.33 and 40 at \$17.32. The updated depth profile satisfies $D_{t+\Delta}(17.33) = 0$ and $D_{t+\Delta}(17.32) = -160$. If instead the next order is $x_2 = (150, \$17.33, t+\Delta)$, it is partially matched: 110 units are sold at \$17.33 and the rest of the order becomes active. Then, the updated depth profile satisfies $D_{t+\Delta}(17.33) = 40$ and $D_{t+\Delta}(17.32) = -200$.

1.1.2 Priority rules

In Example 1.1.9, when order x_1 is submitted, it is matched with 40 of the 200 units available at the price level \$17.32. However, there are 3 buy active orders, denoted by $(y_i)_{i=1,2,3}$, at this price. Which of these is matched with x_1?

The answer to this question depends on the priority rule of the exchange. The most common one is *time* priority: x_1 is matched with the oldest order, that is, y_{i^*} where $i^* = \arg\min t_{y_i}$. This priority rule encourages the early submission of orders [103].

Futures markets however commonly operate a *pro-rata* rule: x_1 is matched with the three buy orders, proportionally to their sizes $(q_{y_i})_{i=1,2,3}$. More precisely, $40 \times 50/200 = 10$ units are matched with each of the two orders of size 50, while $40 \times 100/200 = 20$ units are matched with the larger order of size 100. This rule incentivises market participants to submit orders that are larger than the amount they actually intend to trade.

In general, different priority rules may result in different strategic behaviours.

1.1.3 Limit and market orders

Two essential order types can now be defined: *limit* and *market orders*. A limit order is an order that enters the list of active orders without resulting in any trade, while a market order is an order that is entirely matched at its submission time.

Definition 1.1.10 (Limit order)**.** An order $x = (q_x, p_x, t_x)$ is called a limit order if

$$\{y \in \mathcal{A}_{x-} : p_y \leq p_x\} = \emptyset, \text{ when } q_x < 0;$$
$$\{y \in \mathcal{B}_{x-} : p_y \geq p_x\} = \emptyset, \text{ when } q_x > 0.$$

Definition 1.1.11 (Market order)**.** An order $x = (q_x, p_x, t_x)$ is called a market order if

$$\begin{aligned} \sum_{\{y \in \mathcal{A}_{x-} \,:\, p_y \leq p_x\}} q_y &\geq -q_x, && \text{when } q_x < 0; \\ -\sum_{\{y \in \mathcal{B}_{x-} \,:\, p_y \geq p_x\}} q_y &\geq q_x, && \text{when } q_x > 0. \end{aligned}$$

Remark 1.1.12. Note that any order can be seen as a combination of market orders and limit orders. In Example 1.1.9, the order $x_2 = (150, \$17.33, t+\Delta)$ is equivalent to a market order $(110, \$17.33, t + \Delta)$ followed by a limit order $(40, \$17.33, t+\Delta)$. Thereby, the LOB can be viewed as the set of active limit orders, whence its name.

1.1.4 Real examples

Using the Nasdaq data introduced in Chapter 2, we created animations of real depth profiles. These were uploaded to a Youtube channel*. By watching these videos, one quickly notices the high-frequency nature of today's markets: in a blink of an eye, the LOB's shape changes multiple times. Quantifying and studying these small-timescale dynamics will be the object

*https://www.youtube.com/user/maximemorariu/featured of Chapter 4.

We would also like to stress that the Youtube channel contains an animation of the flash crash of May 2010. In particular, one can observe that the number of active limit orders evaporates as the price plunges, perhaps reflecting the uncertainty of the situation and people's confusion.

1.1.5 Market fragmentation

With the purpose of increasing competition and improving the fairness of prices, new regulations were introduced: the Markets in Financial Instruments Directive (MiFID) in 2004 in the European Union and Regulation National Market System (Reg NMS) in 2005 in the United-States of America. This changed drastically the trading landscape with, in particular, the apparition of new trading venues that would compete with the historical exchanges, a phenomenon known as market fragmentation [100]. Concretely, this implies that shares of the same company can be bought simultaneously on multiple exchanges. Hence, to trade a single stock, one has to monitor and interact with multiple LOBs. Benefiting from the increase in competition between trading venues and market makers came at the cost of an increase in complexity.

1.2 Characterising quantities

Now that the LOB mechanism has been exposed in detail, we define a list of key characterising quantities that are studied in the literature.

1.2.1 Prices, levels and returns

We begin by defining the bid and ask prices, the spread, the mid price and the returns. These are without a doubt the most famous quantities and the first numbers that a trader would look at. We also define was is meant by the level-n LOB and order flow.

Definition 1.2.1 (Best bid and ask prices)**.** The best bid and ask prices at time t, denoted by B_t and A_t respectively, are defined as

$$B_t := \max_{\{x \in \mathcal{B}_t\}} p_x, \quad A_t := \min_{\{x \in \mathcal{A}_t\}} p_x.$$

The best ask (respectively bid) price is the lowest (respectively highest) price at which the ask (respectively bid) side is willing to sell (respectively buy). Thus, buy (respectively sell) market orders are first matched with limit orders sitting at the best ask (respectively bid) price.

Definition 1.2.2 (Spread)**.** The spread S_t at time t is given by

$$S_t := A_t - B_t.$$

The spread measures the distance between the bid and ask sides or by how much they disagree on the fair price. It is often taken as a proxy for the level of liquidity (how much and how quickly one can trade the security) and execution costs.

Definition 1.2.3 (Mid price)**.** The mid price M_t at time t is defined as

$$M_t := \frac{A_t + B_t}{2}.$$

The mid price can be understood as the average price at which the security currently trades. Note however that the mid price is a fictional price, in the sense that the first order after time t can never result in a transaction at price M_t.

More generally, one can define the nth ask (bid) price as the nth price, counting from the best ask (bid) price, at which sell (buy) limit orders are currently posted.

Definition 1.2.4 (nth bid and ask prices)**.** Let $n \in \mathbb{N}$. The nth bid and ask prices at time t, denoted by $B_t^{(n)}$ and $A_t^{(n)}$ respectively, are defined as

$$B_t^{(n)} := \sup \left\{ p \leq M_t \;:\; \sum_{p \leq p' \leq M_t} \mathbb{1}(D_t(p) < 0) = n \right\},$$
$$A_t^{(n)} := \inf \left\{ p \geq M_t \;:\; \sum_{M_t \leq p' \leq p} \mathbb{1}(D_t(p) > 0) = n \right\}.$$

The level-n LOB and order flow are then naturally defined as follows. Perhaps because of the difficulty in having access to more comprehensive data sets, many studies focus on the level-I LOB exclusively.

Definition 1.2.5 (Level-n LOB and order flow)**.** Let $n \in \mathbb{N}$. The level-n LOB at time t, denoted by $\mathcal{L}_t^{(n)}$, is given by

$$\mathcal{L}_t^{(n)} := \left\{ x \in \mathcal{L}_t \;:\; p_x \in [B_t^{(n)}, A_t^{(n)}] \right\}.$$

The level-n order flow is the sequence of orders and cancellations y such that $\mathcal{L}_{y-}^{(n)} \neq \mathcal{L}_y^{(n)}$.

Finally, a quantity for which there are many empirical studies, as we will later see, is the mid price returns.

Definition 1.2.6 (Mid price returns)**.** The time series of mid price returns with constant time step $\tau > 0$ and initial time t_0 is denoted by $(R_{t_k}^{(\tau)})_{k \in \mathbb{N}^*}$ and given by

$$R_{t_k}^{(\tau)} := \ln \left(\frac{M_{t_{k+1}}}{M_{t_k}} \right)$$

where $t_k := t_0 + k\tau$.

1.2.2 Relative price and profile

To study the depth profile and the behaviour of market participants on each side of the LOB, it can be more appropriate to think in terms of *relative* price.

Definition 1.2.7 (Relative price)**.** The relative prices of a sell order x and buy order x' at time t are defined by

$$\delta_t^a(p_x) := p_x - A_t \quad \text{and} \quad \delta_t^b(p_{x'}) := B_t - p_{x'}, \quad \text{respectively.}$$

Active orders with positive relative prices do not belong to the level-I LOB whereas incoming orders with negative relative prices can reduce the spread and/or be market orders.

If one wants to study how orders are distributed on each side of the LOB, taking the time average of the depth profile would not be adequate because a given price level can pass through both bid and ask regions. This motivates the introduction of the relative depth profile.

Definition 1.2.8 (Relative depth profile)**.** The relative depth profile $D_t^a(\cdot)$ of the ask side of the LOB $\mathcal{L}_t$ at time t is defined as

$$D_t^a(p) := D_t(A_t + p) = \sum_{\{x \in \mathcal{A}_t \,:\, \delta_t^a(p_x) = p\}} q_x, \quad p \geq 0.$$

The relative depth profile $D_t^b(\cdot)$ of the bid side of the LOB $\mathcal{L}_t$ at time t is similarly given by

$$D_t^b(p) := D_t(B_t - p) = \sum_{\{x \in \mathcal{B}_t \,:\, \delta_t^b(p_x) = p\}} q_x, \quad p \geq 0.$$

Definition 1.2.9 (Mean relative depth profile)**.** The mean relative depth profile $\overline{D}_{t_1,t_2}^a(\cdot)$ of the ask side on the time interval $[t_1, t_2]$ is defined as

$$\overline{D}_{t_1,t_2}^a(p) := \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} D_t^a(p) dt, \quad p \geq 0.$$

The mean relative depth profile $\overline{D}_{t_1,t_2}^b(\cdot)$ of the bid side on the time interval $[t_1, t_2]$ is defined analogously.

For example, $\overline{D}_{t_1,t_2}^b(0)$ is the average number of units sitting at the best bid price.

1.2.3 Imbalances

We end this section with two measures of imbalance between buyers and sellers. These two quantities play an important role in Chapters 2 and 4. In the next section, we will review empirical studies that investigate their influence on the price dynamics.

The queue imbalance is an indicator of the shape of the level-I LOB.

Definition 1.2.10 (Queue imbalance)**.** The queue imbalance (QI) of the LOB $\mathcal{L}_t$ at time t is defined as

$$QI_t := \frac{|D_t(B_t)| - |D_t(A_t)|}{|D_t(B_t)| + |D_t(A_t)|}. \tag{1.1}$$

Thanks to the normalisation, $QI_t \in [-1, 1]$. Some studies however work with the unnormalised QI (i.e., the denominator in (1.1) is replaced by one). A positive QI, for example, means that the cumulative size of orders sitting at the best bid price is larger than at the best ask price, which can be interpreted as a buy pressure.

The order flow imbalance, introduced by Cont et al. [36], measures the differential between the inflows of level-I buy and sell orders.

Definition 1.2.11 (Order flow imbalance, ask events, bid events)**.** The order flow imbalance (OFI) between times t_1 and t_2 is defined as

$$OFI_{t_1,t_2} := Bid_{t_1,t_2} - Ask_{t_1,t_2},$$

where Bid_{t_1,t_2} (respectively Ask_{t_1,t_2}) is the cumulative size of bid (respectively ask) events in the time interval $(t_1, t_2]$. We count as bid events: level-I buy orders and level-I sell cancellations (analogous statement for ask events).

1.3 A variety of motivations

Quantifying and understanding the dynamics of LOBs is crucial in many interconnected aspects.

1.3.1 Price formation mechanism

LOB models can shed light on the forces that drive the market and the way they combine to generate price moves and volatility. One problem is to understand how the main (statistical) features of a LOB (e.g., the price patterns and mean relative depth profile) derive from the strategic behaviour of agents [108]. Another problem is to reconcile the dynamics of the price process at large timescales (typically modelled as an Itô diffusion) with its behaviour at small timescales (typically modelled as a pure jump process) [34, 66, 71, 71].

1.3.2 Regulation

The resolution parameters and the priority rules can lead to different market dynamics and statistical regularities, because of the different order placement strategies they incentivise. For instance, one issue for exchanges and regulators is to predict the effect of a change in the tick size on the cost of trading [64]. The analysis of LOBs can also provide new perspectives on market stability. Indeed, studying the market dynamics from the viewpoint of LOBs allows one to take into account the *liquidity factor*, that is, how quickly one can trade large quantities without significantly impacting the price [44]. More generally, the field of LOB modelling can assist regulators in the design of their policies.

1.3.3 Quantitative trading

LOB modelling is obviously important for practitioners who seek optimal decisions when trading. One major question, known as the optimal liquidation problem, is to find the best possible way to sell (or buy) a large number of shares in a given time window (from one hour to a few trading days) [5]. As one usually searches for an optimal sequence of smaller trades, understanding the impact of a trade on the LOB dynamics, and in particular the price, seems crucial to tackle this question effectively. Optimal order placement is another topic that requires a deep comprehension of LOBs [35]. In this topic, one aims to determine the optimal way of combining limit and market orders, potentially on multiple trading venues, to quickly trade a small quantity of a considered asset.

The analysis of LOB data can also reveal robust statistical patterns that can be exploited to increase the profit of market makers or generate alpha, that is, returns that beat the overall market performance [23, 82]. In particular, statistically accurate LOB models can inspire or enhance high-frequency trading algorithms that operate at very short timescales and react to the order flow in sometimes less than a millisecond. Besides, the development of realistic models can result in reliable market simulation tools [97], which are essential for the backtesting of trading strategies.

1.4 Empirical studies

In this section, we finally turn to the empirical studies on LOB data. We shall here focus on model-free analyses, that is, no specific stochastic process is postulated as the data generating process. The application of Hawkes processes and continuous-time Markov chains to high-frequency data will be covered in Sections 1.6 and 1.7, respectively. We put up a list of *statistical facts* that seem robust across all studies. We will later refer to these by using their number in the list (e.g., Fact (xii)).

Before reviewing studies on aspects that are more specific to LOBs (the mid price is not a concept that is specific to order-driven markets), we must mention the *stylised facts* of (mid) price returns, which are now well established, at least for stock prices [33, 25]. LOB models that aim for a realistic behaviour (e.g., LOB simulators) should generate a price process that exhibits such statistical features.

(i) **Heavy tails of mid price returns and aggregational Gaussianity.** The unconditional distribution of mid price returns $(R_{t_n}^{(\tau)})_{n\in\mathbb{N}}$ (cf. Definition 1.2.6) displays heavy tails. As the sampling time step τ increases, the shape of the distribution changes and resembles more and more a normal distribution. Cont [33] calls this phenomenon *aggregational Gaussianity.*

(ii) **Uncorrelated returns.** The mid price returns have a negligible autocorrelation except over short time scales, for which it is reported to be negative.

(iii) **Volatility clustering.** The autocorrelation function $\gamma_{vol}(\cdot)$ of the squared mid price returns $((R_{t_n}^{(\tau)})^2)_{n\in\mathbb{N}}$ is positive and decays slowly. Some people argue that it exhibits *long memory*, with $\gamma_{vol}(n) \sim n^{-\alpha}$ for some $\alpha \in (0,1)$ (it decays as a non-integrable power law), but this is subject to debate.

1.4.1 Unconditional statistical regularities

Compared to the mid price returns, establishing a consensus on the stylised facts of LOB-specific aspects seems more delicate. Not only the number of variables to investigate is significantly larger, but different studies can sometimes reach contradictory conclusions. For instance, the distribution of the size of limit orders on the Nasdaq and the Island ECN has radically different shapes [27, 90]. Special rules such as bilateral trade agreements (on foreign exchange markets) can also alter the statistical features of the data [54]. Moreover, since the statistical characteristics of LOBs must somehow result from the traders' behaviour, there is no apparent reason to expect stationary data. In addition to the daily to monthly potential changes in behaviour, the regulatory and technological transformation of markets has probably impacted the decision-making process of agents over the last two decades. This possible non-stationarity at multiple timescales renders any statistical analysis more fragile and makes it difficult to compare empirical studies over different time periods. Arguing that there are not enough papers using present-day data, Gould et al. [55, p. 1734] insist that "[s]tudies of recent, high-quality LOB data that are conducted with stringent awareness of potential statistical pitfalls are needed to understand better the LOBs of today".

With the above caveats in mind, we will now proceed with presenting what seems to be statistical regularities of LOBs, beginning with their unconditional behaviour. Our main references here are [55] and [25].

(iv) **Non-Poissonian submission times.** The sequence of times at which orders are submitted cannot be described by a stationary Poisson process, which strongly suggests that orders are not sent independently of one another. Modelling the dependence structure between the submission times of different order types is one of the main contributions of this thesis. Our empirical findings in Chapter 4 do confirm this statistical regularity.

(v) **U-shaped intraday seasonality.** On average, the market activity, measured by the number of orders submitted, is higher at the beginning and end of the day, but lower around midday, exhibiting a so-called U-shape. We do find this pattern in our empirical analysis (Chapter 4).

(vi) **High autocorrelations in the order flow.** Define the sign of the nth trade X_n as $X_n = -1$ if the trade is triggered by a buy market order and $X_n = +1$ otherwise. The autocorrelation function of $(X_n)_{n \in \mathbb{N}}$ decays slowly, with many studies arguing for long memory [53], although this could be an artefact caused by structural breaks in the data [7]. This remains true if one considers instead limit orders or cancellations [16, 84]. As suggested by Tóth et al. [120], this apparent long memory in the order flow could be due to traders splitting large quantities to execute, called *parent* or *meta orders*, into a sequence of smaller actual orders, their goal being to reduce their market impact.

(vii) **Power law for the relative price of orders.** The distribution of the relative price of incoming orders, $\delta^a_{x-}(p_x)$ and $\delta^b_{x-}(p_x)$ (cf. Definition 1.2.7), is close to a power law [89] with an exponent systematically above -2, meaning that the distribution is not integrable. Bouchaud et al. [17, p.252] found some limit orders on the Paris Bourse that would be filled only if the mid price moved by $\pm 50\%$.

(viii) **High cancellation rate.** Most of active orders are removed from the

LOB because they are cancelled, not because they are matched [55, Section IV.C]. A a group of studies on the Island ECN, the S&P 500 futures contracts, foreign exchange markets and an exchange-traded fund tracking the Nasdaq 100 showed that approximately 70-80% of orders end up being cancelled. Gould et al. [55] refer to one of their papers in preparation that shows that more than 99.99% of active orders are cancelled on today's foreign exchange markets. They suggest that this might be due to the intensification of algorithmic trading.

(ix) **Hump-shaped mean relative depth profile.** The mean relative depth profiles $(\overline{N}^a_{t_1,t_2}(p))_{p\geq 0}$ and $(\overline{N}^b_{t_1,t_2}(p))_{p\geq 0}$ (cf. Definition 1.2.9) display a hump, meaning that, on average, the longest queue of orders is not situated at the best bid and ask prices, but deeper in the LOB. Gould et al. [55, Section IV.D] refer to studies on the Paris Bourse, the Nasdaq, the Stockholm Stock Exchange and the Shenzhen Stock Exchange. For example, one can examine the results in [17, Paris Bourse, data from February 2001]. The location of the hump depends on the stock or the market however.

1.4.2 Conditional behaviour in LOBs

Given the large number of variables that define the order flow and the state of the LOB, there are potentially infinitely many conditional relationships to investigate. However, we will focus on three conditional behaviours that are perhaps the most robust and popular.

(x) **Concave price impact.** The term *price impact* refers to the expected mid price shift following the submission of a single market order of size q_x. Even though different definitions are considered in the literature, the absolute price impact is always found to be a concave function of $|q_x|$, usually a power law [60, 85, 105, 15]. The average price slippage during the execution of a large quantity Q (*meta* or *parent order*) that is split into a sequence of smaller orders is also reported to be a concave function of Q [120].

(xi) **Price-predictive power of the QI.** The QI (Definition 1.2.10) carries information on both the direction of the next price move (up or down) [52] and the mid price variation over a fixed (short) time window [90, 23, 82]. A positive (respectively negative) QI, close to 1 (respectively -1), tends to be followed by an increase (respectively decrease) in the mid price.

(xii) **Prices driven by the OFI.** After introducing the OFI (Definition 1.2.11), Cont et al. [36] showed that a linear regression of the mid price increments $M_{t+\Delta} - M_t$ on $OFI_{t,t+\Delta}$ results in a very good fit with an R^2 around 65%.

Facts (xi) and (xii) will be confirmed, commented and studied in Chapter 2. For now, let us just stress that the QI and the OFI are different in nature. The QI is a measure of the *current state* of the level-I LOB and informs us on the *future* price move. The OFI is a measure of the level-I *order flow* during a fixed time window and explains the price variation during the *same* time window.

1.4.3 Large- and small-tick stocks

To end this section, we would like to explain the distinction between large- and small-tick stocks. The ratio ρ/M_t (tick size over current price) can be interpreted as the relative cost of bypassing a queue of pre-existing limit orders to gain priority (i.e., submitting a limit order one tick closer to the mid price or a market order). Moreover, the smaller ρ/M_t, the more there are price levels relatively close to the mid price. For these reasons, one can expect ρ/M_t to influence the traders' strategic behaviour and, thus, the shape and dynamics of the LOB. Indeed, for large-tick stocks (high ρ/M_t), it is observed and generally accepted that all price levels around the mid price are occupied by large quantities of orders while $S_t = \rho$ holds most of the time (locked spread) [6, 34, 14]. For small-tick stocks however, many price levels are unoccupied, orders tend to spread across small queues and $S_t > \rho$ holds most of the time.

It is important to note that Facts (xi) and (xii) are only shown to hold for large-tick stocks and seem to fade as ρ/M_t decreases [52]. The last part of Chapter 2 will propose a concept of *deep imbalance* that (partially) restores these statistical regularities when small-tick stocks are considered.

1.5 Different modelling approaches

The purpose of this section is to acknowledge that the general modelling approach considered in the remainder of the thesis (stochastic modelling) cannot answer to all the motivations mentioned in Section 1.3 and that other complementing philosophies exist. Here, we will simply make the distinction between agent-based and stochastic models, but a more subtle classification and additional references can be found in [55].

On the one hand, agent-based models explicitly represent the market as an aggregation of agents. Each agent follows a predefined objective (e.g., expected utility maximisation) or acts according to a predetermined strategy. The goal with such models is to derive the LOB dynamics that result from the interaction between agents. To this end, one usually needs to make some unrealistic assumptions (e.g., each agent trades only one unit of the asset and then leaves the market) and invoke parameters that cannot be directly measured (e.g., a risk-aversion coefficient in a utility function). Alternatively, when models are more realistic but intractable, one can resort to simulation to study the generated market dynamics. The downside of this however is that it becomes difficult, or at least costly, to understand the impact of each parameter and each model component [26]. Perhaps the main value of agent-based models resides in their ability to explain how some statistical regularities of LOBs stem from the strategic behaviour of traders. Therefore, agent-based models are key at shedding light on the price formation mechanism and can also help regulators understand how certain rules shape the market dynamics.

Just to mention an example, Roşu [108] proposes a model where the utility function of impatient traders is penalised by the time it takes to match their (limit) order. In other words, this model explores the idea that traders can be driven by their impatience to access liquidity. Using game theory to derive the implied market dynamics, Roşu [108] finds that, in the model, the LOB displays a hump-shaped profile (cf. Fact (ix)). Hence, this model explains Fact (ix) by the existence of patient traders, who are less penalised by long waiting times, and who anticipate large market orders and thus post limit orders beyond the level-I LOB.

On the other hand, stochastic models describe directly the dynamics of observable market quantities (e.g, mid, ask and bid prices, order sizes, order submission times) without giving any causal explanation of the phenomenon. There are no explicit agents that drive the LOB. Instead, one directly models the probability distribution and dependence structure of the order flow and/or LOB variables, the goal being to provide an accurate and/or tractable statistical dynamic description. As we shall see, the estimation of such models from market data can provide a framework for studying the conditional behaviour of LOBs and reveal patterns that might not be visible or measurable with a model-free approach. Compared to agent-based models, stochastic models usually rely on more reasonable assumptions and their parameter values can be directly inferred from the available data. That's why they are perhaps more suited to answer to the motivations discussed in Subsection 1.3.3 (quantitative trading).

The next two sections will review two main classes of stochastic models: Hawkes processes and continuous-time Markov chains.

1.6 Hawkes processes

Hawkes processes, named after the person who introduced them first [62], are a class of self- and cross-exciting processes in which events of different types can precipitate each other. Since they break the memorylessness property of Poisson processes, they are a potential candidate for modelling the dependence structure of the order flow (cf. Fact (iv)). A general literature review on Hawkes processes was proposed by Laub et al. [81] while a survey focusing on their applications to finance was provided by Bacry et al. [10].

1.6.1 Basics of point process theory

To introduce Hawkes processes, we briefly recall some central concepts of point process theory [19, 40]. A d_e-dimensional multivariate point process consists of an increasing sequence of positive random times $(T_n)_{n\in\mathbb{N}}$ and a matching sequence of random marks $(E_n)_{n\in\mathbb{N}}$ in $\mathcal{E} := \{1, \ldots, d_e\}$. We interpret, for any $n \in \mathbb{N}$, the pair (T_n, E_n) as an event of type E_n occurring at time T_n. A point process can be identified by its counting process representation $\boldsymbol{N} := (N_1, \ldots, N_{d_e})$, where

$$N_e(t) := \sum_{n\in\mathbb{N}} \mathbb{1}_{\{T_n \le t, E_n = e\}}, \quad t \ge 0, \quad e \in \mathcal{E},$$

counts how many events of type e have occurred by time t. The counting process $\boldsymbol{N}$ is said to be *non-explosive* if $\lim_{n\to\infty} T_n = \infty$ with probability one. Given a filtration $\mathbb{F} = (\mathcal{F}_t)_{t\ge 0}$ to which $\boldsymbol{N}$ is adapted, we say that a non-negative $\mathbb{F}$-adapted process $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_{d_e})$ is the $\mathbb{F}$-*intensity* of $\boldsymbol{N}$ if

$$\lim_{h\downarrow 0} \frac{\mathbb{E}\left[N_e(t+h) - N_e(t) \,|\, \mathcal{F}_t\right]}{h} = \lambda_e(t), \quad t \ge 0, \quad e \in \mathcal{E}.$$

Intuitively, $\lambda_e(t)$ is the infinitesimal rate of new events of type e at time t. Note that this heuristic definition will be completed by a rigorous one in Chapter 5, where we present a theoretical framework for marked point processes.

1.6.2 Definition and interpretation

Definition 1.6.1 (Hawkes process)**.** Let $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_{d_e}) \in \mathbb{R}^{d_e}_{>0}$ and $\boldsymbol{k} = (k_{e'e})_{e',e\in\mathcal{E}}$, where $k_{e'e} : \mathbb{R}_{>0} \to \mathbb{R}_{\ge 0}$. A point process $\boldsymbol{N}$ is a *Hawkes process* with base rate vector $\boldsymbol{\nu}$ and excitation kernel $\boldsymbol{k}$, abbreviated as Hawkes$(\boldsymbol{\nu}, \boldsymbol{k})$, if $\boldsymbol{N}$ has a $\mathbb{F}^{\boldsymbol{N}}$-intensity $\boldsymbol{\lambda}$ that satisfies

$$\lambda_e(t) = \nu_e + \sum_{e'\in\mathcal{E}} \int_{[0,t)} k_{e'e}(t-s)\, dN_{e'}(s), \quad t \ge 0, \quad e \in \mathcal{E},$$

where $\mathbb{F}^{\boldsymbol{N}}$ is the natural filtration of $\boldsymbol{N}$.

To illustrate how past events may influence the arrival rate of future events, the sample path of a univariate Hawkes process with an exponential kernel is sketched in Figure 1.2.

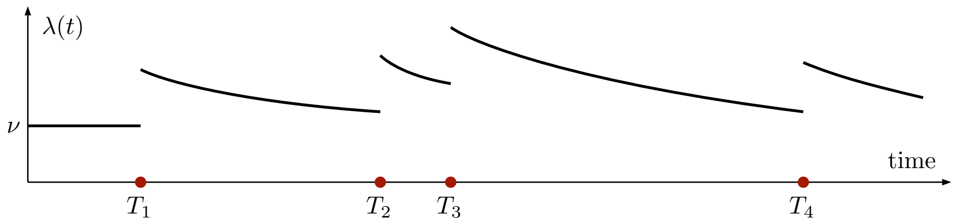


Figure 1.2: Sample path of a univariate Hawkes process ($d_e = 1$) with an exponential kernel of the form $k(t) = \alpha \exp(-\beta t)$, where $\alpha, \beta \in \mathbb{R}_{>0}$.

The component $k_{e'e}$ of $\boldsymbol{k}$ determines the magnitude and the timescale of the excitation effect that events of type e' have on events of type e. More precisely, an event of type $e' \in \mathcal{E}$ at time $T = T_n$ for some $n \in \mathbb{N}$ increases the infinitesimal rate of new events of type $e \in \mathcal{E}$ at time $t > T$ by $k_{e'e}(t - T)$. Notice that when $\boldsymbol{k} \equiv 0$, the Hawkes process collapses to a multivariate Poisson process with rate $\boldsymbol{\nu}$.

1.6.3 Cluster representation, endogeneity and stability

Additional intuition on Hawkes processes can be gained by understanding their cluster representation [63, 73], which we here summarise for the univariate case ($d_e = 1$). First, generate a Poisson process with rate ν and denote by G_0 (generation 0) the corresponding set of time points. Then, apply the following recursion. For each point t' of the nth generation G_n, generate a Poisson process starting at t' with inhomogeneous rate $k(t - t')$, $t > t'$, and denote by $G_{n+1}^{t'}$ the corresponding set of time points. Generation $n + 1$ is defined by $G_{n+1} := \cup_{t \in G_n} G_{n+1}^{t}$. Sorting the superposition of all generations $\cup_{n \in \mathbb{N}} G_n$ into an increasing sequence of random times $(T_n)_{n \in \mathbb{N}}$ defines a Hawkes(ν, k).

Every event time t from generation G_n, $n \geq 1$, is the children of an event time $t' < t$ from the previous generation G_{n-1}. Points from the initial generation G_0 are exogenous whereas points from all the other generations are endogenous. In spite of this very transparent endogeneity and even though it has been shown that Hawkes processes capture the *Granger causality* between different event types [45], one should remain careful and avoid saying that "t' causes t". Large [78] prefers instead the expression "t' precipitates t".

The cluster representation of Hawkes processes allows us to study their dynamics using the theory of branching processes [59], in particular multi-type Galton-Watson processes. Remaining in the univariate case ($d_e = 1$), each event generates on average $\rho := ||k||_{1,\infty} := \int_0^\infty k(t)dt$ children. Consequently, each element in G_0 generates on average

$$\underbrace{\rho}_{\text{mean number of children}} + \underbrace{\rho \times \rho}_{\text{mean number of grandchildren}} + \dots = \sum_{n \geq 1} \rho^n$$

events in total, which is finite if and only if $\rho < 1$. Hence, the process is non-explosive under the sufficient condition that $\rho < 1$. This condition also implies the existence of a stationary version [40, p. 184], but, again, is not necessary.

In the multivariate case, an analogous result holds, where ρ becomes the spectral radius of the matrix of kernel norms. Furthermore, still under the condition $\rho < 1$, convergence to a unique stationary version can be proved. We recall that the spectral radius of a square matrix is the largest modulus of its eigenvalues.

Theorem 1.6.2 (Stability condition)**.** *Define the matrix* $\boldsymbol{M} = (m_{ij})$ *where* $m_{ij} := ||k_{ji}||_{1,\infty}$, $i, j \in \mathcal{E}$, *and denote its spectral radius by* ρ. *If* $\rho < 1$, Hawkes$(\boldsymbol{\nu}, \boldsymbol{k})$ *converges to a unique non-explosive stationary version.*

Proof. See Brémaud and Massoulié [21, Theorem 7]. □

1.6.4 Estimation techniques

The problem of estimating the base rate vector $\boldsymbol{\nu}$ and kernel $\boldsymbol{k}$ from the observation a sample path has been the focus of many publications.

Thanks to the general theoretical result that expresses the likelihood of a sample path in terms of the intensity process [40], parametric kernels can be estimated by a numerical maximisation of the likelihood function [98, 18]. However, computing the likelihood is very costly as it requires $\mathcal{O}(N^2)$ operations, where N is the number of events in the sample path, which makes the optimisation problem practically insolvable. Happily, in the case of (a superposition of) exponential kernels, that is, $k_{e'e}(t) = \alpha_{e'e}\exp(-\beta_{e'e}t)$, $\alpha_{e'e} \in \mathbb{R}_{\geq 0}$, $\beta_{e'e} \in \mathbb{R}_{>0}$, the computational complexity reduces to $\mathcal{O}(N)$ [102]. For this reason, parametric estimation via maximum likelihood is always performed for kernels of this special form in the literature, even though Ogata [99] reminds us that other more general parameterisations can also reduce the complexity to $\mathcal{O}(N)$. More details will be given in Chapter 3 where we extend this parametric framework to state-dependent Hawkes processes. To close this paragraph on parametric estimation, we note the availability of the generalised method of moments [38, 39].

In terms of non-parametric techniques, perhaps the most popular and robust one was introduced by Bacry and Muzy [12]. It consists in solving a Wiener-Hopf equation that links the kernel $\boldsymbol{k}$ to the second-order statistics of the process and can be adapted to the case of slowly decreasing kernels (e.g., power laws). Another interesting approach was proposed by Kirchner [77] who approximates a Hawkes process by an integer-valued vector autoregressive time series. Additional references and non-parametric techniques are covered in [10].

Finally, we would like to mention the existence of an alternative technique based on an Expectation-Maximisation (EM) algorithm that takes advantage of the cluster representation and can be used in both parametric and non-parametric frameworks (see, again, [10] and references therein).

1.6.5 Applications to LOB modelling

The survey [10] demonstrates that Hawkes processes have found numerous applications in finance. In particular, they have been used as models of market risk [46], market endogeneity [49], market impact [11], the effect of news announcements [107] and the microstructure price [8], even in the context of optimal execution [4]. More generally, the Hawkes-process approach fits into the longer trend of using point processes to model sequences of irregularly-spaced market events in continuous time [47, 61, 13]. Besides, the study of scaling limits of Hawkes processes has provided a new way of understanding how the large-timescale dynamics of a stock price derive from the short-timescale LOB behaviour [71], with Jaisson and Rosenbaum [72] offering the first *microstructural* explanation of the *roughness of volatility*, recently discovered by Gatheral et al. [50]

Coming back to the focus of this thesis, Hawkes processes have been used as a model of the (level-I) order flow. While different model specifications and estimation techniques have been used, the general approach remains the same. First, one specifies a list of order categories that is used to classify or discard the orders in the data. For example, one might only retain the submission times of market orders and classify them according to their direction (buy or sell). Then, one estimates a Hawkes process from the timestamps of the selected event types, enabling one to gain new insights on the dependence structure of the order flow.

One can (and we will) try to explain the shape of the estimated kernel components in terms of the traders' strategic behaviour. However, the kernel arguably captures the *aggregate* market reaction, which is very probably the product of numerous *heterogeneous* agents. Inferring the traders' behaviour from the estimated kernel remains thus a delicate exercise. The kernel estimate can also be understood as a measure of *market impact*, that is, how orders influence the LOB dynamics, complementing the studies on *price impact* (Fact (x)). It is also interesting to note that estimating the kernel $\boldsymbol{k}$ somehow implies disentangling the impact of orders to infer the contribution of each event type in isolation, which model-free approaches like in [14] cannot achieve. Hawkes processes essentially provide us with a new lens to read the market reaction to different order types and their interaction.

Large [78] was the first to use a Hawkes process as a model of the order flow in a LOB market. Using a parametric framework with exponential kernels, his focus is on formalising and measuring the resiliency of the LOB, that is, its capacity to replenish following a market order that moves the mid price. He considers $d_e = 10$ event types: buy/sell market/limit orders that move the mid price, buy/sell market/limit orders that do not move the mid price and cancellations on the bid/ask side. Focusing on the level-I order flow, Bacry et al. [9] apply their non-parametric estimation technique to $d_e = 8$ event types: buy/sell market/limit/cancel orders that do not change the mid price and orders that push the mid price up/down. Rambaldi et al. [106] perform a similar study but take into account the size of orders, increasing the number of event types to $d_e = 24$. The type of an event is determined by three components: the order type (market, limit, cancellation), the direction (buy, sell) and the bin in which the order size falls (4 bins are defined). From these three studies, we stress the following main findings.

- The self-excitation effects dominate the cross-excitation effects, except for price moves, as upwards (respectively downwards) moves mainly precipitate downwards (respectively upwards) moves.
- The dominating excitation effects are persistent, that is, the corresponding kernel components decrease slowly. Note that this first two items are reminiscent of and consistent with the high autocorrelations in the order flow (Fact (vi)).
- There is a buy-sell symmetry, meaning that self-excitation on the bid side and the cross-excitation from the bid side to the ask side mirror the self-excitation on the ask side and the cross-excitation from the ask side to the bid side, respectively.
- A high degree of endogeneity is observed, although no actual measures of the spectral radius are given.

- In [9] and [106], the timescale of the self- and cross-excitation effects can be very short, in the order of 0.1–1 millisecond. The timescales reported by Large [78] are much larger (in the order of 10 seconds), most probably because of the poor timestamp precision in his data, which is of one second.

Except for the second item, our estimation results in Chapter 4, where we apply state-dependent Hawkes processes to high-quality LOB data, are consistent with these findings.

1.7 Continuous-time Markov chains

We now turn to another major LOB modelling approach which is based on continuous-time Markov chains.

1.7.1 Zero-intelligence models

The first stochastic models of the LOB viewed the order flow as a homogenous Poisson process [112], neglecting the dependence structure empirically observed (Fact (iv)). This assumes that agents are submitting orders independently of the LOB history, manifesting no strategic behaviour, whence the name of *zero-intelligence models*.

In spite of this unrealistic assumption, this pioneering approach has led to interesting results. Not only the probability of different events conditional on the current state of the LOB can be computed in (semi-) closed form [37, 34], but the volatility of the price process can be explicitly linked to the parameters of the order flow [34, 1].

As an illustrative example, let us review in more detail the model of Cont et al. [37]. Orders, assumed to be of unit size, are placed on a finite grid $\{1, ..., N\}$ of price levels, where N is taken *large enough*. The queue lengths on this price grid are represented by the process $\boldsymbol{X}_t = (X_t^1, ..., X_t^N)$, which in terms of the depth profile D_t (cf. Definition 1.1.6) is simply given by $X_t^n = D_t(n\rho)$, where remember that ρ is the tick size. The positions of the best bid and ask queues on the finite grid are denoted by $n_t^b := B_t/\rho$ and $n_t^a := A_t/\rho$, where remember that B_t and A_t are the bid and ask prices. The arrival rate of limit orders at a given price level is just a function of the distance to the opposite best quote. For example, sell limit orders with price level $n_t^b + i$, $i \geq 0$, are submitted according to a Poisson process with rate $\lambda(i) \in \mathbb{R}_{\geq 0}$. Similarly, cancellations at a distance $i \in \mathbb{N}$ from the opposite best quote are given by a Poisson process with rate $\theta(i) X_t^{n_t^{a/b} \pm i}$: the rate is proportional to the number of outstanding orders at the corresponding price level. The submission of market orders is also described by independent exponential durations: trades are described by two Poisson processes with rate μ, one for buys and one for sells. All Poisson processes are assumed mutually independent. The process $\boldsymbol{X}_t$ is thus a continuous-time Markov chain with state space $\mathbb{Z}^N$, meaning that the probability distribution of future states depends only on the current state, as opposed to the full history.

Employing Laplace transform methods, Cont et al. [37] are able to compute several conditional probabilities, in particular the probability that the next mid price move is an increase. The model is strikingly able to replicate Fact (xi) (price-predictive power of QI): the probability of an upward move increases as the best bid queue lengthens and the best ask queue remains fixed. We note that a simplified version of this model in [34], which accounts only for the level-I LOB, yields a similar relationship. This finding suggests that Fact (xi) is mainly a *mechanical* effect that can be explained without appealing to the anticipation and strategic behaviour of agents. Our empirical results in Chapter 2 will support this hypothesis as well. Finally, let's also mention that the model reproduces Fact (ix) (hump-shaped profile) and generates a price process with a volatility that has realistic magnitude.

1.7.2 Queue-reactive models

Huang et al. [65] propose a model where the arrival rates of orders can depend on the relative depth profile, going beyond zero-intelligence models but remaining in the setting of continuous-time Markov chains. Their approach introduces a concept of reference price P_t, different from the mid price, that can be driven by both endogenous and exogenous factors and moves on the shifted grid $\{(n + 0.5)\rho : n \in \mathbb{N}\}$. The other component of the model, denoted by $\boldsymbol{Q}_t = (Q_t^{-N}, ..., Q_t^{-1}, Q_t^{1}, ..., Q_t^{N})$, is the depth profile *relative to the reference price*. For example, when $i > 0$, Q_t^i is the total size of outstanding sell limit orders with price $P_t + (i - 0.5)\rho$. Most importantly, the point process that describes the order flow has an intensity that is a function of the relative depth profile, that is $\boldsymbol{\lambda}(t) = f(\boldsymbol{Q}_t)$, and, thus, is not necessarily a Poisson process anymore.

The estimation of this queue-reactive model for $N = 3$ and different forms of f reveals that the arrival rates are indeed state-dependent. For example, buy market orders are submitted faster when the best ask queue is relatively short. This could be explained by market participants expecting a mid price increase (Fact (xi)), making it more interesting to consume the remaining orders in the best ask queue.

Huang and Rosenbaum [66] extend this queue-reactive model to a more general class of processes, although we note that the behaviour of the reference price and its impact on the LOB dynamics are slightly different in this follow-up. In particular, the effect of LOB events on the reference price are not instantaneous anymore. This comprehensive class of continuous-time Markov chains nests for example the reduced-form model in [34] (a slight variation of the zero-intelligence model in [37]) and an analogue of the original queue-reactive model in [65]. Under general assumptions, Huang and Rosenbaum [66] prove that the system is ergodic and the rescaled reference price process converges to a Brownian motion.

Even though this framework allows for a general dependence of the order flow on the relative depth profile, it is important to clarify that it does not consider all possible state variables. For instance, the actual number of orders in a queue as well as their individual sizes and submission times might influence the behaviour of traders. These state variables of the LOB are unfortunately not contained in the relative depth profile.

The model of Cartea et al. [23] which describes the joint evolution of the mid price, spread and QI as a continuous-time Markov chain with state-dependent arrival rates could also be classified as a *queue-reactive* model, although it does no fall exactly within the framework of Huang and Rosenbaum [66].

To end this section, let us mention that Swishchuk and Vadori [117] extend the zero-intelligence model in [34] in a different direction. They allow the type of the next event and the probability distribution of its arrival time to depend on the type of the previous event. Since the times between events are not necessarily driven by exponential random variables anymore, this extension leaves the setting of continuous-time Markov chain. The new model must be studied from the viewpoint of Markov renewal processes.

1.8 A dichotomy to resolve

In light of the previous two sections, a dichotomy between Hawkes processes and continuous-time Markov chains emerges.

Shorthand 1.8.1. Throughout the thesis, we will refer to the dichotomy that is explained in this section as the *HM dichotomy*. This shorthand is simply based on the last-name initials of Alan Hawkes and Andrey Markov.

While LOB models based on Hawkes processes have been rather successful in describing the dynamics of the order flow, they do not to incorporate any endogenous state variables describing the LOB, such as prices, volumes or the bid–ask spread, nor their influence on the arrival rate of orders. The Hawkes-process models that we reviewed in Section 1.6 account only for the dynamics of the order flow and ignore the state of the underlying LOB. We note that Hawkes processes have been used as building blocks to full LOB models [97, 2], but the arrival rate of orders in these models is still not influenced by the state of the LOB. In effect, the existing Hawkes-process models neglect the feedback loop between the order flow and the LOB.

By contrast, models that are based on continuous-time Markov chains focus on the state of the LOB, represented by its depth profile. In particular, the queue-reactive models assume that the arrival rate of orders is exclusively driven by the shape of the LOB. This has the merit of introducing a feedback loop between the order flow and the state of the LOB but it omits the self- and cross-excitation effects evidenced by the empirical work based on Hawkes processes.

In short, each of these two approaches to LOB modelling has desirable qualities that the other lacks. Hawkes processes and continuous-time Markov chains can have either an *event* viewpoint or a *state* viewpoint, respectively. The former approach models the aggregate market reaction to different event types while the latter captures the adjustment of the market dynamics to different shapes of the LOB. In practice, the agents' strategic behaviour probably takes into account both the (recent) order flow history and the current state of the LOB, as both carry different, yet complementing, information.

The dichotomy between Hawkes processes and Markovian models, and the need for something bridging the two was also noticed by Bacry et al. [9, p. 1190–1191] , Taranto et al. [118] and Gonzalez and Schervish [51]. In fact, Gonzalez and Schervish [51] show empirically that the type of the next order depends on both the type of the previous order and the state of the LOB, which highlights the need for a more general modelling framework that can somehow combine the two approaches.

Resolving the HM dichotomy is the high-level motivation and contribution of this thesis.

2

Preliminary data analysis

In this chapter, we present empirical results that were obtained relatively early in the PhD process. On top of providing additional motivation for addressing the HM dichotomy, we investigate further some of the statistical regularities discussed in the previous chapter, namely Facts (xi) (price-predictive power of the QI) and (xii) (prices driven by the OFI) as well as the distinction between small- and large-tick stocks. We find that the QI (cf. Definition 1.2.10) is a significant state variable of the LOB in two ways (Section 2.2). First, even if the OFI is the main *pathwise* driver of price moves, the QI carries complementing information that dominates *on average*. Second, the aggregate market reaction to order flow events varies with the QI. Furthermore, we propose a concept of *deep imbalance* which is able to restore the price-predictive power of the QI for small-tick stocks (Section 2.3). This chapter is also an opportunity to introduce the data that is used throughout this thesis (Section 2.1).

2.1 LOBSTER data

The LOB data that we use is supplied by LOBSTER* (Limit Order Book System – The Efficient Reconstructor), a tool that offers high-quality historical records of the LOB, for all the stocks that are traded on the Nasdaq. LOBSTER is based on Nasdaq's *Historical TotalView-ITCH data*†, which is a record of the "standard Nasdaq data feed for serious traders"‡. However, this record consists only of the sequence of orders and, thus, LOBSTER has its own algorithm to reconstruct the evolution of the LOB throughout each trading day.

When requesting data for a given stock, one must specify a number of levels N. For each trading day, LOBSTER supplies the level-N order flow and level-N LOB (cf. Definition 1.2.5) in the form of two CSV files: the *message file* and the *order book file*. Each row in the message file corresponds to an order flow event and intersects six columns:

- the timestamp of the event which, for the time periods we study, is always stored with nanosecond precision;
- the event type, stored as an integer between 1 and 7 (see Table 2.1);
- the order ID, a unique reference number that identifies the limit order associated to the event;
- the size, that is, the number of shares traded, cancelled or added to the LOB;
- the price of the limit order associated to the event;
- the direction of the limit order associated to the event (buy or sell).

Regarding the order book file, it stores the first N ask and bid prices as well as the outstanding quantities at those prices. More precisely, the ith row of

*https://lobsterdata.com
†http://nasdaqtrader.com/Trader.aspx?id=ITCH
‡http://www.nasdaqtrader.com/Trader.aspx?id=Totalview2

Event type number	Description
1	Submission of a new limit order
2	Partial cancellation of a limit order
3	Total cancellation of a limit order
4	Execution of a visible limit order
5	Execution of a hidden limit order
6	Aggregate of trades that happen during an auction
7	Trading halt indicator

Table 2.1: Event types in LOBSTER.

the order book file stores the level-N depth profile following the ith event in the message file as follows (cf. Definition 1.2.4):

$$A_i^{(1)},\ |D_i(A_i^{(1)})|, B_i^{(1)},\ |D_i(B_i^{(1)})|,\ \ldots,\ A_i^{(N)},\ |D_i(A_i^{(N)})|, B_i^{(N)},\ |D_i(B_i^{(N)})|.$$

Note that the order book file does not provide the detailed list of limit orders at each price level. The data does not inform us on the number of outstanding orders in each queue as well as their respective sizes and submission times. For this reason, a more accurate name for this file would have been the *depth profile file*.

It is also important to stress that each row in the message file does not correspond to the submission or cancellation of an order, but to one of the event types in Table 2.1. In particular, instead of recording the submission of buy (respectively sell) market orders, LOBSTER saves the execution of sell (respectively buy) limit orders. As a market order can be matched with more than one limit order, the submission of a market order can generate multiple consecutive rows with tied timestamps in the message file. Even though it is in theory possible that two separate market orders are sent at exactly the same time (with nanosecond precision), we will always count a string of consecutive events with type numbers in $\{4, 5\}$, tied timestamps and identical directions as a single market order.

As shown in Table 2.1, the Nasdaq allows market participants to hide their limit orders. A hidden limit order is simply not displayed in the data feed and thus not visible by other market participants. At each price level, hidden limit orders must however always give priority to visible limit orders, that is, a hidden limit order can be matched with an incoming market order only after all visible limit orders with the same price are matched.

In the US equity markets, the tick size is fixed to \$0.01 by Rule 612 of Regulation National Market System (Reg NMS), with the exception of stocks priced below \$1.00 per share. For all the stocks we will consider, the tick size ρ is equal to \$0.01. Let us also clarify that the Nasdaq operates a time-priority rule (cf. Subsection 1.1.2).

2.2 The queue imbalance: a significant state variable

2.2.1 The relationship between returns and imbalances

As already discussed, Cont et al. [36] defined a new variable called the OFI in order to explain the relationship between price moves and the level-I order flow. Remember that a positive (respectively negative) OFI essentially indicates an excess of buy (respectively sell) level-I orders. For the purpose of this section, we also need to recall the definition of bid and ask events (cf. Definition 1.2.11).

Inspired by a stylised view of the LOB where all price levels are occupied by the same amount of outstanding quantities, Cont et al. [36] propose the following linear regression model, where (X_t) is the log-mid-price process, that is, $X_t = \ln M_t$.

Linear Regression Model 1.

$$X_{t+\Delta} - X_t = \beta_{N(t)} OFI_{t,t+\Delta} + \varepsilon^{(1)}_{t,t+\Delta}$$

Here, the function $N : \mathbb{R} \to \mathbb{N}$ maps time to the number of the corresponding time bin. Over each time bin, the coefficient β is assumed to be constant. For time bins of 30 minutes and $\Delta = 10$ seconds, Cont et al. [36] find an average coefficient of determination R^2 of 65% when performing an ordinary least square regression over 50 stocks and one calendar month of data. Note that in the original paper, Model 1 is specified in terms of mid price increments instead of mid price returns. We prefer to work with (X_t) since it guarantees a positive price.

This result catches our attention as it gives an explicit *pathwise* relation simplifying the functional that maps the initial state of the LOB and the subsequent order flow to the corresponding mid price trajectory. Given the initial state $\mathcal{L}_t$ of the LOB at time t and the detailed order flow $OF_{t,t+\Delta}$ in the time interval $(t, t+\Delta]$, there is a *deterministic* functional $\varphi(\mathcal{L}_t, OF_{t,t+\Delta})$ that returns the log-price increment $X_{t+\Delta} - X_t$. The quantity $\varphi(\mathcal{L}_t, OF_{t,t+\Delta})$ is computed simply by applying the LOB mechanism described in Section 1.1. However, even if this functional is straightforward to compute on a case by case basis, it cannot be written in an explicit form that would inform us on its general behaviour. The finding of Cont et al. [36] reveals that this functional can in fact be approximated by

$$X_{t+\Delta} - X_t = \varphi(\mathcal{L}_t, OF_{t,t+\Delta}) \approx \beta_{N(t)} OFI_{t,t+\Delta}. \tag{2.1}$$

As a side remark, this approximation could be an effective shortcut for understanding how the large-timescale dynamics of the mid price derive from the short-timescale behaviour of the LOB. Given any model of the level-I order flow, Equation (2.1) can virtually be used as a first order approximation of the implied price process.

Putting aside the fact that $\beta_{N(t)}$ is a proxy for the average relative depth profile in the considered time bin, it is striking that the approximation in (2.1) does not use the information contained in the initial state $\mathcal{L}_t$. Moreover, how can the price-predictive power of the QI (Fact (xi)) be explained from Model 1?

To investigate this question, we first use LOBSTER data on the stock of Apple (AAPL) from the 8th of September 2015. As demonstated in Figure 2.1, we observe that Fact (xi) holds indeed: a higher QI implies an increased expected mid price. However, as shown in Figure 2.2, it happens that, on the

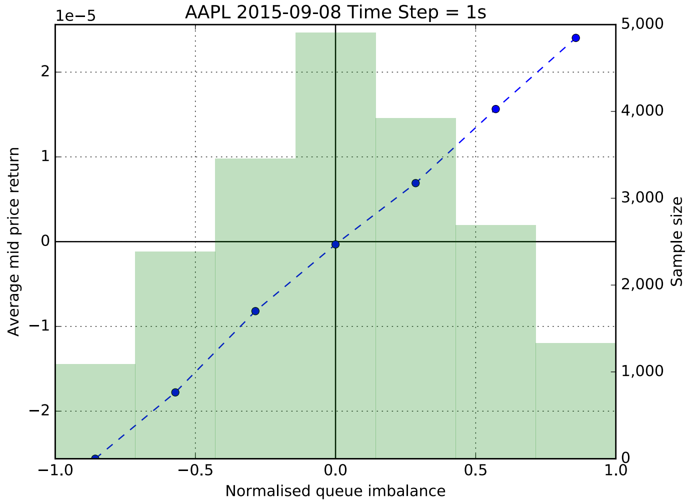


Figure 2.1: Average mid price returns conditioned on the QI for the stock of Apple (AAPL) on the 8th of September 2015 between 10.30am and 4pm. The mid price returns are computed over a time window of one second. We condition on the value of the QI at the beginning of the time window. The values of the QI are split into 7 bins.

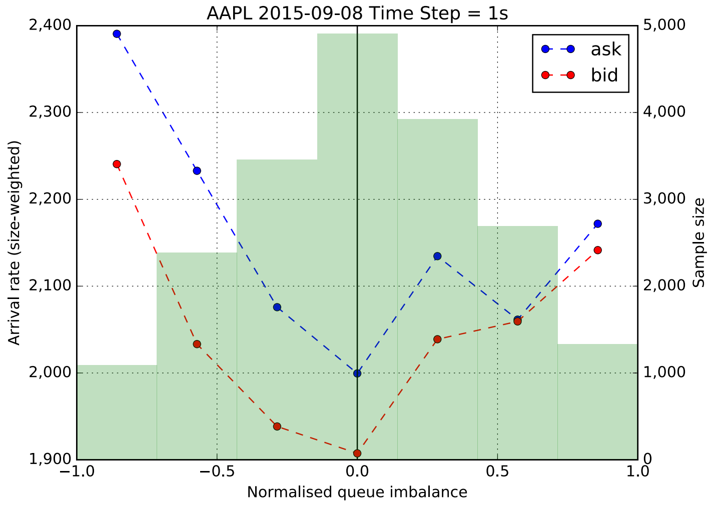


Figure 2.2: Average ask and bid arrival rates conditioned on the QI for the stock of Apple (AAPL) on the 8th of September 2015 between 10.30am and 4pm. The arrival rates are computed over a time window of one second. We condition on the value of the QI at the beginning of the time window.

same day, the average arrival rate of ask events is bigger than the average arrival rate of bid events, even when we condition on the QI. This implies that the average OFI conditioned on the QI is systematically negative. Hence, if one applies Model 1 and treats the residuals $\varepsilon_{t,t+\Delta}$ as *pure* noise, one should conclude that the expected mid price returns conditioned on the QI are always negative, which is of course inconsistent with Figure 2.1. This suggests that the residuals $\varepsilon_{t,t+\Delta}$ must depend on the LOB's shape and cannot be treated as *purely exogenous* variables.

To check this claim, we extend Model 1 by introducing the QI as an additional explanatory variable.

Linear Regression Model 2.

$$X_{t+\Delta} - X_t = \beta_{N(t)} OFI_{t,t+\Delta} + \gamma_{N(t)} QI_t + \varepsilon^{(2)}_{t,t+\Delta}$$

We compare Models 1 and 2 using 4 months of data (5 Jan. 2016 to 29 Apr. 2016) on the stocks of Apple (AAPL) and Twitter (TWTR). In the following, we only discuss the results for Twitter but similar plots are obtained for Apple. As in [36], we split trading days (9.30am to 4pm) into 13 bins of 30 minutes but choose $\Delta = 1$s. For every trading day, we estimate the parameters β and γ by performing an ordinary least square regression over each 30-minute bin separately.

Figure 2.3 shows that most of the variance is explained by the OFI with an average R^2 around 70% for Model 1 (if one disregards the first 30 minutes of trading). Model 2 allows to slightly increase the R^2 but the contribution of the QI remains small. Nevertheless, for almost all days across all time bins, t-tests with a significance level of 1% reject the null hypothesis that $\gamma_n = 0$ (cf. Figure 2.4). Moreover, Figure 2.5 shows that the residuals $\varepsilon^{(1)}_{t,t+\Delta}$ in Model 1 depend on the LOB's shape as claimed previously. This figure estimates the functions $q \mapsto \mathbb{E}[\varepsilon^{(i)}_{t,t+\Delta} | QI_t = q]$, $i = \{1, 2\}$, by splitting the possible values of QI_t into 11 bins. One can observe that the dependence of the residuals on the QI is significantly reduced in Model 2.

This finding shows that the price-predictive power of the QI cannot be ex-

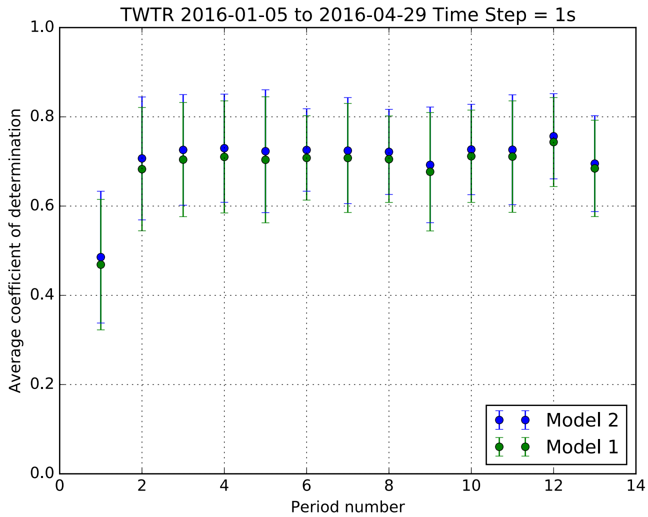


Figure 2.3: Average and standard deviation of R^2 for each 30-minute bin of the day in Linear Regression Models 1 and 2 ($\Delta = 1$s). We use 4 months of data (Jan-Apr 2016) for the stock of Twitter.

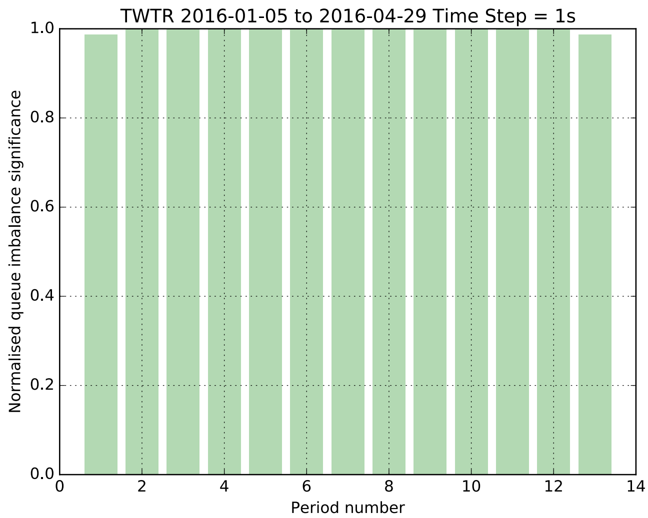


Figure 2.4: Significance of the QI in Linear Regression Model 2 ($\Delta = 1$s). For each day and for each time bin, we test the null hypothesis that $\gamma_n = 0$ using a t-test. The plot shows, for each time bin, the fraction of days for which the null hypothesis was rejected at a significance level of 1%.

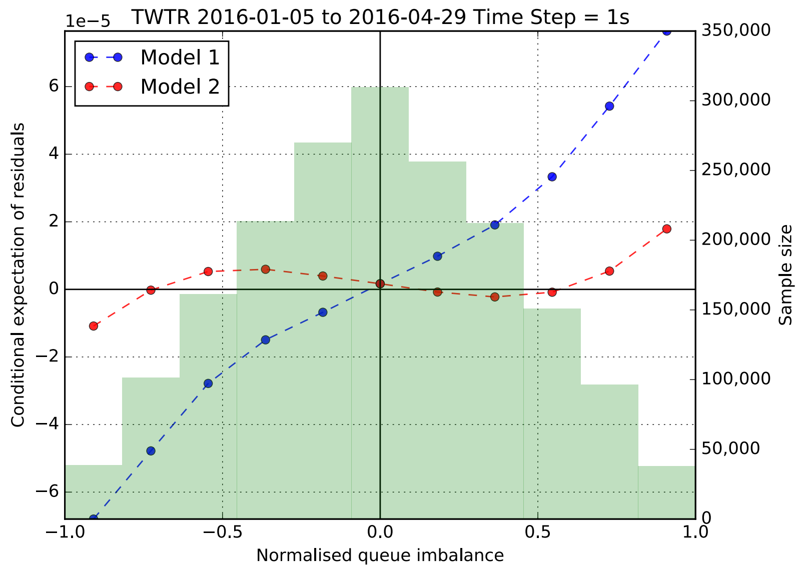


Figure 2.5: Average of the residuals in Models 1 and 2 conditioned on the QI.

plained from Model 1. The reason a positive QI is followed on average by a positive mid price return is not an adjustment of the order flow to the LOB's shape such that bid events outweigh ask events. Instead, these results suggest that the influence of the QI on the mid price moves is mainly *mechanical*. A positive QI means that, the amount of bid events required to increase the mid price is smaller than the amount of ask events required to decrease it. The QI acts virtually as an initial offset to the OFI. This can be summarised by rewriting the approximation in Equation (2.1) as

$$X_{t+\Delta} - X_t = \varphi(LOB_t, OF_{t,t+\Delta}) \approx \gamma_{N(t)} QI_t + \beta_{N(t)} OFI_{t,t+\Delta}. \qquad (2.2)$$

While from a pathwise perspective the OFI is the main driver of price moves, the arrival rates of bid and ask events compensate each other such that, on average, the mechanical effect of the QI dominates.

2.2.2 Influence on the market impact of LOB events

The application of Hawkes processes to LOB data shows that the order flow exhibits significant self- and cross-excitation effects (cf. Subsection 1.6). This

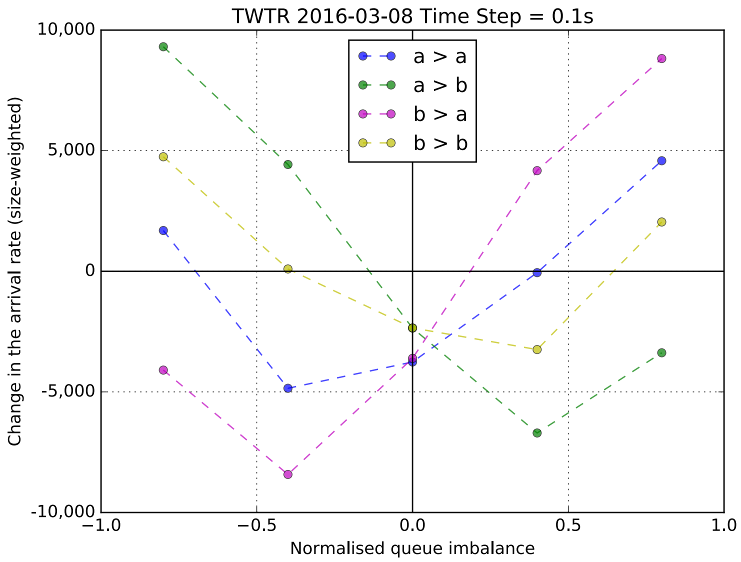


Figure 2.6: Influence of the queue imbalance on the excitation effects in the level-I order flow. For each possible couple of event types $(e, e') \in \{\text{ask}, \text{bid}\}^2$, we plot $I(e \to e' \,|\, \cdot)$ as a function of the QI (the bin's mid point). We use the data from 10.30am to 4pm on the 8th of March 2016 for the stock of Twitter. For example, the curve *b>a* corresponds to the couple $(e, e') = (\text{bid}, \text{ask})$.

subsection presents evidence that the aggregate market reaction to order flow events depends in fact on the QI and is thus state-dependent.

We split the possible values of the QI into $n_{bins} = 5$ bins. For each bin $i \in \{1, ..., n_{bins}\}$ and for each couple of event types $(e, e') \in \{\text{ask}, \text{bid}\}^2$, we do the following: for each event of type e such that the value of the QI just after the event falls in the i^{th} bin, we count the number $N(e' \leftarrow e \,|\, i)$ of events of type e' within the past Δ seconds (weighted by the size of orders) and the number $N(e \to e' \,|\, i)$ of events of type e' within the next Δ seconds (again, weighted by the size of orders). We then compute the empirical average of $N(e \to e' \,|\, i) - N(e' \leftarrow e \,|\, i)$, denoted by $I(e \to e' \,|\, i)$, which we interpret as a proxy of the excitation effect of events of type e on events of type e' when the QI is in the i^{th} bin.

Figure 2.6 plots $I(e \to e' \,|\, i)$ as a function of the bin's mid point for the four possible couples $(e, e') \in \{\text{ask}, \text{bid}\}^2$. Here, we use the data on the stock of

Twitter from the 8th of March 2016 and we choose $\Delta = 1\text{s}$. Consider for example the curve $b{>}a$ which corresponds to the function $I(\text{bid} \to \text{ask} \,|\, \cdot)$. When the QI is close to 1 (buy pressure), the occurrence of a bid event tends to increase the arrival rate of ask events. However, when the QI is close to -1 (sell pressure), the impact of bid events is different as they tend to decrease the arrival rate of ask events. Similar observations can be made for the other curves. Furthermore, this pattern is robust across the 4 months of data (5 Jan. 2016 to 29 Apr. 2016).

If the order flow were described by a Hawkes process as in Section 1.6, and thus not influenced by the LOB, all the curves in Figure 2.6 should be flat. Therefore, this empirical result can be taken as evidence that the impact of order flow events varies with the state of the LOB. Actually, this suggests that the excitation kernel of the Hawkes process could depend on the LOB's shape, an idea that is explored in detail in Chapters 3 and 4. More generally, this finding complements the existing empirical evidence on the state-dependence of the order flow (Section 1.7) and provides additional motivation for addressing the HM dichotomy.

2.3 From level-I to deep imbalances

As mentioned in Subsection 1.4.3, the relationship between returns and imbalances, that is, Facts (xi) and (xii), seems to fade as one goes from large-tick stocks to small-tick stocks [52]. In this section, we try to restore these statistical regularities for small-tick stocks by introducing a concept of *deep* imbalance.

Facts (xi) and (xii) suggest that mainly orders with prices that are *close enough* to the mid price influence price returns. However, for large-tick stocks, the *relative* distance between the mid price and the bid and ask prices is usually much larger than for small-tick stocks. This implies that, for small-tick stocks, level-n orders with $n > 1$ can still be *close enough* to the mid price to influence its trajectory. The following definition of the *deep* QI applies this idea of relative distance and takes into account the contribution of orders that sit deeper in the LOB.

Definition 1 (Deep queue imbalance)**.** *Let $\delta > 0$. The deep queue imbalance (deep QI) at time t is defined as*

$$QI_t^\delta := \frac{\sum_{M_t e^{-\delta} < p < M_t} |D_t(p)| - \sum_{M_t < p < M_t e^\delta} |D_t(p)|}{\sum_{M_t e^{-\delta} < p < M_t} |D_t(p)| + \sum_{M_t < p < M_t e^\delta} |D_t(p)|}.$$

For example, to compute the contribution of buy limit orders, one adds up the outstanding quantities at all the prices levels in the range $(M_t \exp(-\delta), M_t)$, that is, within a distance δ of the mid price on a logarithmic scale. Depending on the value of δ and the current state of the LOB, this computation can involve more than the level-I LOB. In a similar fashion, we define the *deep* OFI, denoted by OFI_{t_1,t_2}^δ, where one considers the order flow within a log-distance δ of the log-mid-price, instead of the level-I order flow (cf. Definition 1.2.11).

We use data on the stocks of Amazon (AMZN, \$550), Google (GOOG, \$650), iShares NASDAQ Biotechnology Index (IBB, \$300), Intuitive Surgical (ISRG, \$475) and Priceline Group (PCLN, \$1,300) from the 1st of October 2015 to the 23rd of October 2015 (17 days of trading), where we indicated in parenthesis the symbol and average price. Remember that the tick size is fixed to $\rho = \$0.01$, hence these assets can be considered as small-tick stocks compared to companies like Twitter (TWTR, \$30) [14].

For each stock, we set the value of δ using the following procedure. We sample the log-distance $\ln(A_t/M_t)$ between the ask price and the mid price and set δ to the 95th percentile of the empirical distribution. This means that 95% of the time, the deep queue imbalance QI_t^δ takes into account at least the best bid and ask queues. The other 5% of the time, we set QI_t^δ to zero. In practice, with this choice of δ, the number of levels used to compute the deep QI varies from 1 to 15. The reason for choosing δ in this manner is that we postulate that the best ask and bid prices should most of the time be *close enough* to the mid price. However, one could of course investigate further what is the optimal value of δ and how it varies across stocks.

In Figure 2.8, for AMZN, we plot the average of mid price returns conditioned on the deep QI. Compared to Figure 2.7 (same scale on the vertical axis)

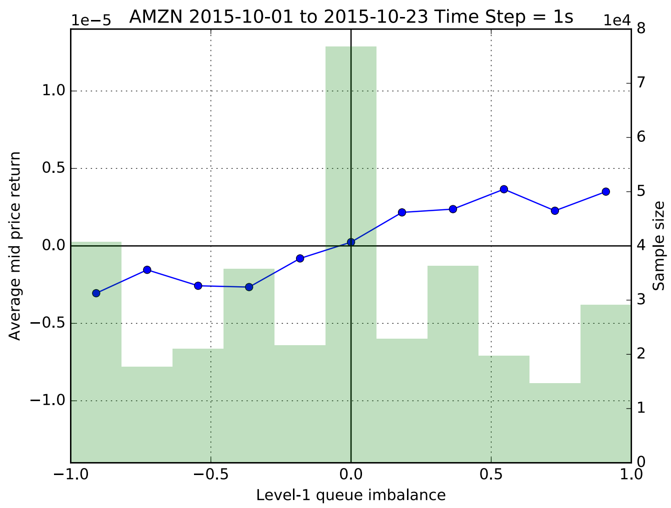


Figure 2.7: Average mid price returns conditioned on the normalised queue imbalance for the stock of Amazon (AMZN). The mid price returns are computed over a time window of one second. We condition on the value of the queue imbalance at the beginning of the time window. The values of the normalised queue imbalance are split into 11 bins.

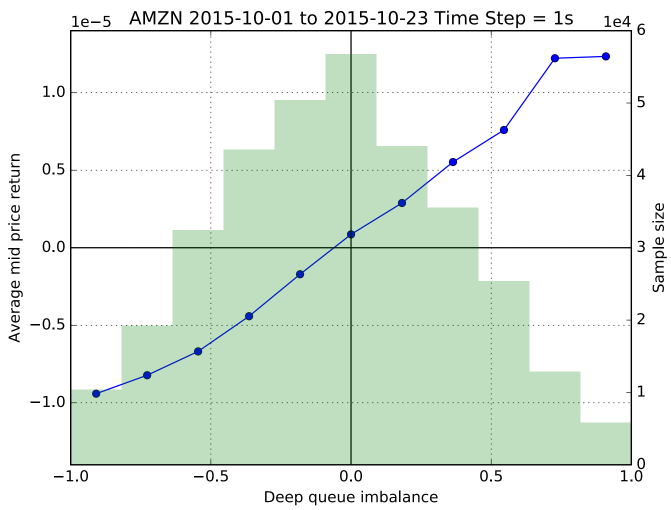


Figure 2.8: Average mid price returns conditioned on the normalised deep queue imbalance for the stock of Amazon (AMZN). The mid price returns are computed over a time window of one second. We condition on the value of the deep queue imbalance at the beginning of the time window. The values of the normalised queue imbalance are split into 11 bins.

where we condition instead on the QI, we retrieve a relationship that is stronger, less noisy and similar to the case of large-tick stocks (cf. Figure 2.1). Notice also that the distribution of the deep QI is more *regular* than that of the QI, resembling more the distribution of the QI for large-tick stocks. We obtain similar results for the other stocks mentioned above.

We also fitted Linear Regression Model 1 using the deep order flow imbalance OFI^δ instead of OFI and found that this allows to increase R^2 from 40% to 50%, excluding the beginning and the end of the trading day (cf. Figure 2.9).

As a final remark, notice that for large-tick stocks, with the way we choose δ, the deep imbalances QI^δ and OFI^δ almost coincide with the level-1 imbalances QI and OFI, since the spread is most of the time equal to one tick. In this sense, the deep imbalances QI^δ and OFI^δ generalise the level-I imbalances QI and OFI to small-tick stocks.

To summarise, the present results suggest that the mid price is driven by orders that are *close enough* to it but that the concept of *close enough* should be formalised in terms of *relative* price instead of number of ticks or number of levels.

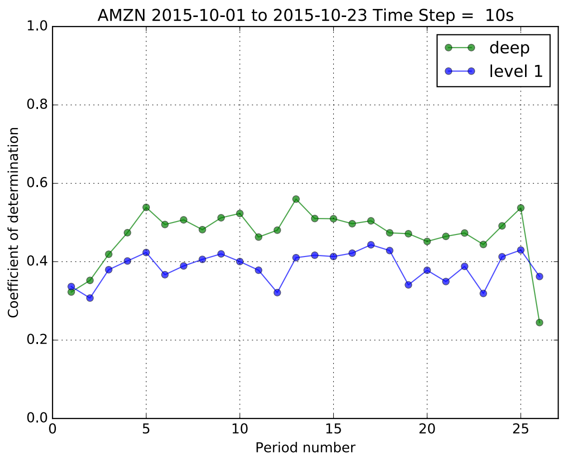


Figure 2.9: Average of R^2 for each 15-minute bin of the day for Linear Regression Model 1 ($\Delta = 10$s). For the green curve, we use the deep order flow imbalance OFI^δ as the explanatory variable instead of OFI. We use 17 days of data on the stock of Amazon (1st of October 2015 to the 23rd of October 2015).

3

State-dependent Hawkes processes

This chapter addresses the HM dichotomy by introducing the class of state-dependent Hawkes processes, an extension of Hawkes processes where the point process $\boldsymbol{N}$ is endowed with a piecewise-constant state process X. The novelty here is that $\boldsymbol{N}$ and X are fully coupled, just like the order flow and the LOB. The kernel that determines the excitation effects between the different event types can now depend on the history of the state process X. Moreover, X can jump only at the event times of $\boldsymbol{N}$, with transition probabilities that depend on the event type.

A key idea is to lift $(\boldsymbol{N}, X)$ to a single ordinary point process of higher dimension. This allows us to use the framework of hybrid marked point processes (Chapters 6 and 7) to obtain the existence and uniqueness of non-explosive state-dependent Hawkes processes and derive a simulation algorithm.

Furthermore, we explain how the maximum likelihood estimation principle applies to parametric specifications of this new class. In particular, we find a convenient separability property of the likelihood which implies that the transition probabilities and the kernel can be estimated independently of one another, in spite of the full coupling between $\boldsymbol{N}$ and X.

The chapter provides also some details on the package `mpoints`, a Python library that we created for the simulation and estimation of state-dependent Hawkes processes with exponential kernels.

3.1 Definition

To begin with, recall the basic definitions of point process theory from Subsection 1.6.1 and the notations therein. Then, to define state-dependent Hawkes processes, we endow the point process $\boldsymbol{N}$ with a càdlàg state process $(X(t))_{t\geq 0}$ that takes values in a finite state space $\mathcal{X} := \{1, \ldots, d_x\}$ and denote by $\mathbb{F}^{\boldsymbol{N},X}$ the natural filtration of the pair $(\boldsymbol{N}, X)$.

Definition 3.1.1. Let $\boldsymbol{\phi} = (\boldsymbol{\phi_e})_{e\in\mathcal{E}}$ be a collection of $d_x \times d_x$ transition probability matrices, $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_{d_e}) \in \mathbb{R}^{d_e}_{>0}$ and $\boldsymbol{k} = (k_{e'e})_{e',e\in\mathcal{E}}$, where $k_{e'e} : \mathbb{R}_{>0} \times \mathcal{X} \to \mathbb{R}_{\geq 0}$. The pair $(\boldsymbol{N}, X)$ is a *state-dependent Hawkes process* with transition distribution $\boldsymbol{\phi}$, base rate vector $\boldsymbol{\nu}$ and excitation kernel $\boldsymbol{k}$, abbreviated as sdHawkes$(\boldsymbol{\phi}, \boldsymbol{\nu}, \boldsymbol{k})$, if

(i) $\boldsymbol{N}$ has $\mathbb{F}^{\boldsymbol{N},X}$-intensity $\boldsymbol{\lambda}$ that satisfies

$$\lambda_e(t) = \nu_e + \sum_{e'\in\mathcal{E}} \int_{[0,t)} k_{e'e}(t-s, X(s))dN_{e'}(s), \quad t \geq 0, \quad e \in \mathcal{E}; \tag{3.1}$$

(ii) X is piecewise constant and jumps only at the event times $(T_n)_{n\in\mathbb{N}}$, so that

$$\mathbb{P}\left(X(T_n) = x \,|\, E_n, \mathcal{F}^{\boldsymbol{N},X}_{T_n-}\right) = \phi_{E_n}(X(T_n-), x), \quad n \in \mathbb{N}, \quad x \in \mathcal{X}, \tag{3.2}$$

where $\mathcal{F}^{\boldsymbol{N},X}_{T_n-} := \bigvee_{\varepsilon>0} \mathcal{F}^{\boldsymbol{N},X}_{T_n-\varepsilon}$ and $X(T_n-) := \lim_{t\uparrow T_n} X(t)$.

In applications to LOB modelling, the counting process $\boldsymbol{N}$ and state process X will represent the order flow and the state of the LOB, respectively. An event of type $e' \in \mathcal{E}$ at time $T = T_n$, for some $n \in \mathbb{N}$, increases the infinitesimal rate of new events of type $e \in \mathcal{E}$ at time $t > T$ by $k_{e'e}(t - T, X(T))$. In particular, the self- and cross-excitation effects now depend on the state process X. Reciprocally, at each event in $\boldsymbol{N}$, the state process X may switch to a new state according to the probability (3.2) that depends on the event type. Thereby, mimicking the mechanics of the LOB, the processes $\boldsymbol{N}$ and X are fully coupled. This feature differentiates state-dependent Hawkes processes from Markov-modulated Hawkes processes, studied in Wang et al. [122], Cohen and Elliott [32], Vinkovskaya [121] and Swishchuk [116], where the state follows an exogeneous Markov process that lacks the relationship (3.2). In the case $d_x = 1$, Definition 3.1.1 reduces to that of an ordinary linear Hawkes process. A simulated sample path of a state-dependent Hawkes process with $d_e = 1$, $d_x = 2$ and exponential kernel is shown in Figure 3.1. In this example, the process exhibits self-excitation only in the second state.

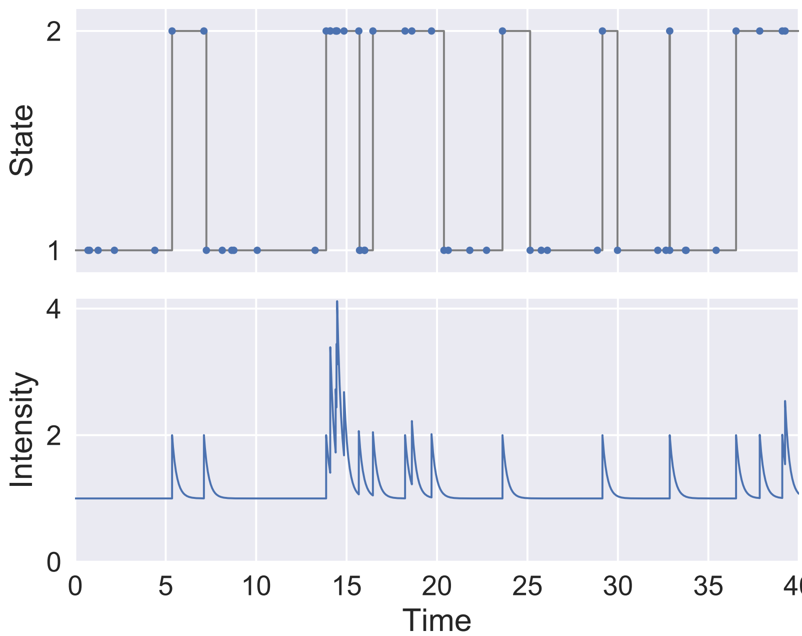


Figure 3.1: Simulation of a state-dependent Hawkes process with $d_e = 1$, $d_x = 2$. The upper plot shows the evolution of the state process. The blue dots indicate the event times and the lower plot represents the intensity. The process is specified so that $\nu_1 = 1$ and $k_{11}(t, x) = \exp(-4t)\mathbb{1}_{\{x=2\}}$, that is, in state 2 the process exhibits exponential self-excitation whereas no self-excitation occurs in state 1.

Since Definition 3.1.1 is implicit in the sense that the counting process $\boldsymbol{N}$ is defined by intensity $\boldsymbol{\lambda}$ that depends on the past of $\boldsymbol{N}$ and X, care is needed to establish the existence and uniqueness of a pair $(\boldsymbol{N}, X)$ that solves (3.1) and (3.2) so that $\boldsymbol{N}$ is non-explosive. To this end, we lift $(\boldsymbol{N}, X)$ to a $d_e d_x$-dimensional multivariate point process $\tilde{\boldsymbol{N}} = (\tilde{N}_{ex})_{e\in\mathcal{E},x\in\mathcal{X}}$, where $\tilde{N}_{ex}$ counts the number of events of type e after which the state is x, formally,

$$\tilde{N}_{ex}(t) = \sum_{n\in\mathbb{N}} \mathbb{1}_{\{T_n\le t, E_n=e, X(T_n)=x\}}, \quad t \ge 0, \quad e \in \mathcal{E}, \quad x \in \mathcal{X}.$$

The marks corresponding to $\tilde{\boldsymbol{N}}$ are given by (E_n, X_n), $n \in \mathbb{N}$, where E_n is as above and $X_n := X(T_n)$ is the value of the state process following the nth event and X_0 is the initial state. Thus, given $\tilde{\boldsymbol{N}}$, the state $X(t)$ can be recovered from the most recent mark at time t, which makes the relationship between $\tilde{\boldsymbol{N}}$ and $(\boldsymbol{N}, X)$ bijective.

By applying the general characterisation result in Chapter 6, the dynamics of $\tilde{\boldsymbol{N}}$ can be expressed in terms of the dynamics of $(\boldsymbol{N}, X)$, and vice versa. The natural filtration of $\tilde{\boldsymbol{N}}$ is denoted by $\mathbb{F}^{\tilde{\boldsymbol{N}}}$.

Theorem 3.2.1. *The pair $(\boldsymbol{N}, X)$ is a non-explosive* sdHawkes$(\boldsymbol{\phi}, \boldsymbol{\nu}, \boldsymbol{k})$ *process if and only if $\tilde{\boldsymbol{N}}$ is non-explosive, admitting $\mathbb{F}^{\tilde{\boldsymbol{N}}}$-intensity $\tilde{\boldsymbol{\lambda}}$ that satisfies*

$$\tilde{\lambda}_{ex}(t) = \phi_e(X(t), x) \left(\nu_e + \sum_{e'\in\mathcal{E}, x'\in\mathcal{X}} \int_{[0,t)} k_{e'e}(t-s, x') d\tilde{N}_{e'x'}(s) \right) \tag{3.3}$$

for all $t \ge 0$, $e \in \mathcal{E}$, $x \in \mathcal{X}$.

The above theorem in fact shows that state-dependent Hawkes processes belong to the class of *hybrid marked point processes* studied in Chapters 6 and 7. The general existence and uniqueness results therein translate to the present framework as follows.

Theorem 3.2.2. *A unique, non-explosive* sdHawkes$(\boldsymbol{\phi}, \boldsymbol{\nu}, \boldsymbol{k})$ *process exists if one of the following two conditions is satisfied:*

(i) *the components of* $\boldsymbol{k}$ *are bounded functions;*

(ii) $\sum_{e' \in \mathcal{E}, x' \in \mathcal{X}} \int_0^\infty k_{e'e}(t, x')dt < (\max_{x' \in \mathcal{X}} \phi_e(x', x))^{-1}$ *for all* $e \in \mathcal{E}$ *and* $x \in \mathcal{X}$.

In Chapter 4, since we will only be using bounded exponentially decaying kernels, we will rely on the first condition exclusively. Note that when $d_x = 1$, the second condition is the classical stability condition in Massoulié [91].

3.3 Simulation

Another implication of Theorem 3.2.1 is that the simulation of a state-dependent Hawkes process can be reduced to the simulation of a multivariate point process with an intensity given by (3.3) and, thus, many simulation techniques from point process theory can be reused [83, 40].

In fact, the sample path in Figure 3.1 has been generated using *Ogata's thinning algorithm* [99], which is an exact simulation algorithm and is adaptable for state-dependent Hawkes processes as follows. We write $R(t) := \sum_{e \in \mathcal{E}, x \in \mathcal{X}} \tilde{\lambda}_{ex}(t) = \sum_{e \in \mathcal{E}} \lambda_e(t)$, which is a function of all (T_n, E_n, X_n) such that $T_n < t$.

Algorithm 3.3.1 Iterative step in Ogata's thinning algorithm for state-dependent Hawkes processes

Require: $(T_i, E_i, X_i)_{i=1,\ldots,n-1}$

set $T := T_{n-1}$
set $\xi := 0$
while $\xi = 0$ **do**
    draw $U \sim \mathrm{Exp}(R(T))$
    set $\xi := 1$ with probability $\frac{R(T+U)}{R(T)}$
    set $T := T + U$
end while
set $T_n := T$
draw $E_n \in \mathcal{E}$ with probabilities proportional to $(\lambda_e(T_n))_{e \in \mathcal{E}}$
draw $X_n \in \mathcal{X}$ with probabilities $(\phi_{E_n}(X_{n-1}, x))_{x \in \mathcal{X}}$
return (T_n, E_n, X_n)

Remark 3.3.1. (i) In Algorithm 3.3.1, we are implicitly assuming that the components of the kernel $\boldsymbol{k}$ are non-increasing, which guarantees that $R(t') \leq R(t)$ for all $t' \in [t, T_n]$. In the general case, one needs to define $R(t)$ so that it bounds the total intensity $\sum_{e \in \mathcal{E}} \lambda_e(t')$ for all $t' \in [t, T_n]$.

(ii) Lines 9–10 of Algorithm 3.3.1 use the product form (3.3) to simulate the marks, which avoids the computation of $d_e d_x$ products at the cost of generating an additional random number.

3.4 Parametric estimation via maximum likelihood

Estimating the base rate vector $\boldsymbol{\nu}$ and kernel $\boldsymbol{k}$ of an ordinary Hawkes process has become a vibrant research topic in statistics and statistical finance literature. Whilst the recent focus has mostly been on non-parametric methodology [12, 77, 45, 109, 3], our aim in this chapter is to extend the classical parametric framework [102], comprehensively summarised in Bowsher [18], to state-dependent Hawkes processes. From now on, we work with a kernel $\boldsymbol{k} = \boldsymbol{k}^{\boldsymbol{\theta}}$ parametrised by a vector $\boldsymbol{\theta} \in \mathbb{R}^p$.

3.4.1 Likelihood function

We know from Subsections 3.2 and 3.3 that a state-dependent Hawkes process $(\boldsymbol{N}, X)$ can be lifted to a $d_e d_x$-dimensional point process $\tilde{\boldsymbol{N}}$. Given a realisation $(t_n, e_n, x_n)_{n=1,\ldots,N}$ of $\tilde{\boldsymbol{N}}$ over a time horizon $[0, T]$, the likelihood function $\mathcal{L}(\boldsymbol{\phi}, \boldsymbol{\nu}, \boldsymbol{\theta})$ can be informally understood as the probability that $\sum_{e \in \mathcal{E}, x \in \mathcal{X}} \tilde{N}_{ex}(T) = N$ and $(T_n, E_n, X_n)_{n=1,\ldots,N}$ lies in a small neighbourhood of $(t_n, e_n, x_n)_{n=1,\ldots,N}$, under the assumption that $(T_n, E_n, X_n)_{n \in \mathbb{N}}$ is generated by a state-dependent Hawkes processes with parameters $(\boldsymbol{\phi}, \boldsymbol{\nu}, \boldsymbol{\theta})$. More rigorously, the likelihood function is the density of the *Janossy measure* with respect to the Lebesgue measure on $\mathbb{R}^N$ [40, p. 125, 213].

For ordinary Hawkes processes, the likelihood function $\mathcal{L}(\boldsymbol{\nu}, \boldsymbol{\theta})$ can be expressed directly in terms of $\boldsymbol{\lambda}$ [40] and the maximum likelihood (ML) estimator $(\hat{\boldsymbol{\nu}}, \hat{\boldsymbol{\theta}})$ is obtained by maximising $\mathcal{L}(\hat{\boldsymbol{\nu}}, \hat{\boldsymbol{\theta}})$, in practice numerically. For state-dependent Hawkes processes, we are able to express $\mathcal{L}(\boldsymbol{\phi}, \boldsymbol{\nu}, \boldsymbol{\theta})$ in terms of $\boldsymbol{\phi}$ and $\boldsymbol{\lambda}$, and find that maximising the likelihood is conveniently achieved by solving two independent optimisation problems.

Theorem 3.4.1. *The log likelihood function of an* sdHawkes$(\boldsymbol{\phi}, \boldsymbol{\nu}, \boldsymbol{k}^{\boldsymbol{\theta}})$ *process is given by*

$$\ln \mathcal{L}(\boldsymbol{\phi}, \boldsymbol{\nu}, \boldsymbol{\theta}) = \sum_{n=1}^{N} \ln \phi_{e_n}(x_{n-1}, x_n) + \sum_{n=1}^{n} \ln \lambda_{e_n}(t_n) - \int_0^T \sum_{e \in \mathcal{E}} \lambda_e(t) dt. \quad (3.4)$$

Furthermore, $(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\nu}}, \hat{\boldsymbol{\theta}}) \in \arg\max_{\boldsymbol{\phi}, \boldsymbol{\nu}, \boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\phi}, \boldsymbol{\nu}, \boldsymbol{\theta})$ *if and only if*

$$\begin{cases} \hat{\phi}_e(x, x') = \dfrac{\sum_{n=1}^{N} \mathbb{1}(x_{n-1} = x, e_n = e, x_n = x')}{\sum_{n=1}^{N} \mathbb{1}(x_{n-1} = x, e_n = e)}, \quad e \in \mathcal{E}, \quad x, x' \in \mathcal{X}, \\ (\hat{\boldsymbol{\nu}}, \hat{\boldsymbol{\theta}}) \in \arg\max\limits_{\boldsymbol{\nu}, \boldsymbol{\theta}} \sum_{n=1}^{N} \ln \lambda_{e_n}(t_n) - \int_0^T \sum_{e \in \mathcal{E}} \lambda_e(t) dt. \end{cases} \quad (3.5)$$

The upshot of Theorem 3.4.1 is that the ML estimation of state-dependent Hawkes processes is no harder than that of ordinary Hawkes processes. Namely, $\boldsymbol{\phi}$ is estimated in a straightforward manner by the empirical transition probabilities, whilst $(\boldsymbol{\nu}, \boldsymbol{\theta})$ is estimated by maximising the log quasi-likelihood of $\boldsymbol{N}$, that is, the Radon-Nikodym derivative of a change of measure that transforms a standard Poisson process into $\boldsymbol{N}$ [19, Theorem 10, p. 241], which is similar to the log likelihood of a multivariate ordinary Hawkes process. It is remarkable that, in spite of the strong coupling between the events and the state process, the estimation of $\boldsymbol{\phi}$ and $(\boldsymbol{\nu}, \boldsymbol{\theta})$ is decoupled due to the separable form 3.4 of the log likelihood function.

Remark 3.4.2. It is not clear how to derive the likelihood function $\mathcal{L}$ directly from Definition 3.1.1. It is once again the lift $\tilde{\boldsymbol{N}}$ that allow us to transpose the problem to the setting of point process theory, where classical results can be reused (see the proof of Theorem 3.4.1 in Section 3.6).

We note that, in the case of ordinary Hawkes processes ($d_x = 1$), consistency and asymptotic normality results for the ML estimator $(\hat{\boldsymbol{\nu}}, \hat{\boldsymbol{\theta}})$ are available in the literature, see for example [98, 31]. However, these results rely on the stationarity and ergodicity of the underlying process and, unfortunately, these properties need not extend to state-dependent Hawkes processes in general. Therefore, the analysis of the asymptotic properties of the ML estimator is not straightforward and is left for future work. However, in Subsection 3.5.1, we present Monte Carlo results that exemplify the finite-sample performance of the ML estimator. In particular, the results provide evidence that $(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\nu}}, \hat{\boldsymbol{\theta}})$ is consistent as the length of the estimation window increases.

3.4.2 Goodness-of-fit diagnostics using residuals

Denote by $(t_n^e)_{n=1,\dots,n_e}$ the sequence of times at which an event of type e occurred and by $(t_n^{ex})_{n=1,\dots,n_{ex}}$ the sequence of times at which an event type e occurred and after which the state was x, and set $t_0^e = t_0^{ex} = 0$, $e \in \mathcal{E}$, $x \in \mathcal{X}$. To assess the goodness-of-fit of a state-dependent Hawkes process with given parameters $(\boldsymbol{\phi}, \boldsymbol{\nu}, \boldsymbol{\theta})$, introduce the event residuals r_n^e and total residuals $\tilde{r}_n^{ex}$

$$
\begin{aligned}
r_n^e &:= \int_{t_{n-1}^e}^{t_n^e} \lambda_e(t)dt, \quad n = 1, \dots, n_e, e \in \mathcal{E}, \\
\tilde{r}_n^{ex} &:= \int_{t_{n-1}^{ex}}^{t_n^{ex}} \tilde{\lambda}_{ex}(t)dt, \quad n = 1, \dots, n_{ex}, e \in \mathcal{E}, x \in \mathcal{X},
\end{aligned}
$$

where $\boldsymbol{\lambda}$ and $\tilde{\boldsymbol{\lambda}}$ are computed by plugging $(\boldsymbol{\phi}, \boldsymbol{\nu}, \boldsymbol{\theta})$ into (3.1) and (3.3), respectively.

It is a classical result that under the right time-change, $\boldsymbol{N}$ and $\tilde{\boldsymbol{N}}$ become standard unit-rate Poisson processes [93], with the event residuals and total residuals as their respective time increments.

Theorem 3.4.3. (i) *Suppose that $(t_n^e)_{n=1,\dots,n_e,e\in\mathcal{E}}$ was generated by a d_e-dimensional multivariate point process $\boldsymbol{N}$ with an $\mathbb{F}^{\boldsymbol{N},X}$-intensity $\boldsymbol{\lambda}$ satisfying* (3.1)*, where X is a given state process. Then the event residuals $(r_n^e)_{n=1,\dots,n_e}$ for any $e \in \mathcal{E}$ are i.i.d. and follow the* Exp(1) *distribution.*

(ii) *Suppose that $(t_n^{ex})_{n=1,\dots,n_{ex},e\in\mathcal{E},x\in\mathcal{X}}$ was generated by an* *sdHawkes*$(\boldsymbol{\phi}, \boldsymbol{\nu}, \boldsymbol{k}^{\boldsymbol{\theta}})$ *process* $(\boldsymbol{N}, X)$ *such that* $\tilde{N}_{ex}(t) \to \infty$ *as* $t \to \infty$ *with probability one for any* $e \in \mathcal{E}$ *and* $x \in \mathcal{X}$*. Then the total residuals* $(\tilde{r}_n^{ex})_{n=1,\dots,n_{ex}}$ *for any* $e \in \mathcal{E}$ *and* $x \in \mathcal{X}$ *are i.i.d. and follow the* Exp(1) *distribution.*

Consequently, given an estimate $(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\nu}}, \hat{\boldsymbol{\theta}})$, the goodness-of-fit of the process can be assessed by comparing the empirical distribution of each sequence among $(r_n^e)_{n=1,\dots,n_e}$, $e \in \mathcal{E}$ and $(\tilde{r}_n^{ex})_{n=1,\dots,n_{ex}}$, $e \in \mathcal{E}$, $x \in \mathcal{X}$ to the Exp(1) distribution,, where $\boldsymbol{\lambda}$ and $\tilde{\boldsymbol{\lambda}}$ are now computed by plugging $(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\nu}}, \hat{\boldsymbol{\theta}})$ into (3.1) and (3.3), respectively. This can by achieved by inspecting a Q–Q plot (see Subsection 4.6). Goodness-of-fit diagnostics can also check whether the residuals are mutually independent, which can be assessed by examining their correlogram. One can of course go beyond these visual assessments by applying formal statistical tests [18].

3.4.3 The special case of exponential kernels

The direct computation of the term $\sum_{n=1}^{N} \ln \lambda_{e_n}(t_n)$ in the log likelihood function (3.4) involves a double sum requiring $\mathcal{O}(N^2)$ operations to evaluate, which may render ML estimation numerically infeasible when the sample is large. However, for ordinary Hawkes processes ($d_x = 1$) with exponential kernels, it is known that this computation can be achieved in $\mathcal{O}(N d_e^2)$ operations [102, 99]. Fortunately, this property carries over to state-dependent Hawkes processes with kernel $\boldsymbol{k}$ having exponential form

$$k_{e'e}(t, x) = \alpha_{e'xe} \exp(-\beta_{e'xe} t), \quad t > 0, \quad e', e \in \mathcal{E}, \quad x \in \mathcal{X}, \tag{3.6}$$

where the *impact coefficients* $\boldsymbol{\alpha} := (\alpha_{e'xe})$ and *decay coefficients* $\boldsymbol{\beta} := (\beta_{e'xe})$ are non-negative. The reduction in computational cost becomes apparent from the derivation in Subsection 3.5.2. In particular, it is due to the recursive relationship between the sums $S_{e'x'e}$ therein, reducing the computational cost from $\mathcal{O}(N^2)$ to $\mathcal{O}(N)$ operations. A similar remark holds for the computation of the gradient and the residuals.

3.5 The mpoints package

The algorithms that we have developed to simulate and estimate state-dependent Hawkes processes with exponential kernels were released publicly as Python package called `mpoints` that is available at https://github.com/maximemorariu/mpoints. We also worked to provide a full documentation of the package that can be found at https://mpoints.readthedocs.io. In particular, `mpoints` can be easily installed via a one-line command in the terminal using the package management system `pip` and comes with a tutorial that demonstrates its main features. We believe that sharing one's code in an accessible manner is essential for research reproducibility and transparency. Besides, it is hopefully an alternative channel to disseminate state-dependent Hawkes processes beyond the mathematical finance community. Note also that the `mpoints` package comes with specialised plotting services that facilitate the visualisation of parameters and sample paths.

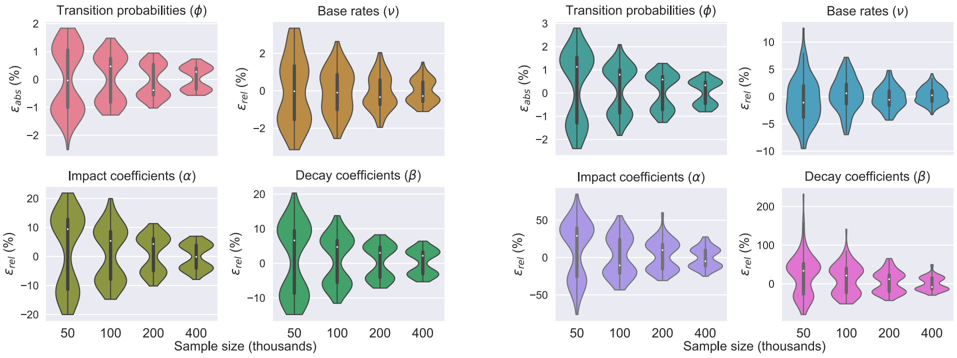
In this section, we shall test the consistency of the ML estimator and give the formulae that we use for the computation of the log likelihood, its gradient and the residuals. We will provide some details on our numerical optimisation methodology in the next chapter, in Subsection 4.10.1.

3.5.1 Test on simulated data

We assess the finite-sample performance of the ML estimator in a small Monte Carlo experiment using a state-dependent Hawkes process with exponential kernel of the form (3.6) when $d_e = 2$ and $d_x = 5$. The parameters are naturally split into four groups: the transition probabilities in $\boldsymbol{\phi}$, the base rates in $\boldsymbol{\nu}$, the impact coefficients in $\boldsymbol{\alpha}$ and the decay coefficients in $\boldsymbol{\beta}$. For each group of parameters $(\theta_{i_j})_{j=1,\dots,p_i}$ and their estimator $(\hat{\theta}_{i_j})_{j=1,\dots,p_i}$, we define the *worst relative error* as

$$\varepsilon_{rel} := \frac{\hat{\theta}_{i_{j^\star}} - \theta_{i_{j^\star}}}{\theta_{i_{j^\star}}}, \quad \text{where} \quad j^\star = \arg\max_j \frac{|\hat{\theta}_{i_j} - \theta_{i_j}|}{\theta_{i_j}}. \tag{3.7}$$

For two different sets of parameter values and four different sample sizes,



(a) Specification 1. **(b)** Specification 2.

Figure 3.2: Violin plots of the worst estimation errors (3.7) under two different sets of parameter values (specifications 1 and 2). For every specification and sample size N, we simulate 100 paths with sample size N and perform ML estimation for each of them. The true parameters are used as the initial guess in the optimisation procedure to speed up estimation and reduce the computational cost.

we estimate the distribution of ε_{rel} for each of the four groups of parameters using Monte Carlo via Algorithm 3.3.1. However, for the transition probabilities, we measure instead the worst *absolute* error, defined by replacing the denominator in (3.7) by one. The first set of parameter values (Specification 1, Table 3.1) is constructed simply by averaging the daily estimates of Model_{QI} in Chapter 4 using INTC data over the 19 trading days of February 2018. The second set of parameter values (Specification 2, Table 3.2) is artificial, chosen to produce more drastic changes in behaviour from one state to another. The results are displayed in Figure 3.2 and indeed support the conjectured consistency of the ML estimators of the state-dependent Hawkes process. Note that the observed bimodality is due to the fact that the worst relative error inherently alternates between positive and negative values.

3.5.2 Formulae for state-dependent Hawkes processes with exponential kernels

In the case of an exponential kernel $\boldsymbol{k}$ given by (3.6), the following formulae can be derived for the second and third terms of the log likelihood function

α_{i1j}		α_{i2j}		α_{i3j}		α_{i4j}		α_{i5j}	
11,314	2,311	8,098	3,556	17,016	270	3,198	535	34,747	220
139	33,460	724	4,497	178	15,497	3,778	7,943	2,353	12,452

(a) Impact coefficients.

β_{i1j}		β_{i2j}		β_{i3j}		β_{i4j}		β_{i5j}	
17,480	9,563	15,187	19,637	29,890	1,456	6,493	3,071	45,546	1,097
613	44,994	4,614	8,384	1,026	27,522	20,537	14,272	10,443	18,988

(b) Decay coefficients.

$\phi_1(i,j)$					$\phi_2(i,j)$				
.89	.02	.05	.03	.01	.78	.19	.02	<.01	<.01
.13	.83	.01	<.01	.02	<.01	.75	.2	.02	<.01
.02	.17	.79	<.01	.02	.02	<.01	.79	.17	.02
<.01	.02	.21	.75	.01	.02	<.01	<.01	.82	.14
<.01	<.01	.02	.19	.78	<.01	.03	.06	.02	.89

(c) Transition distribution.

ν_i
2.2
2.2

(d) Base rate vector.

Table 3.1: Parameter values for Specification 1.

α_{i1j}		α_{i2j}		α_{i3j}		α_{i4j}		α_{i5j}	
2	10	1	3	1,000	30	2,000	40	100	1,000
10	2	3	1	20	3,000	2,000	50	60	2,000

(a) Impact coefficients.

β_{i1j}		β_{i2j}		β_{i3j}		β_{i4j}		β_{i5j}	
10	15	8	4	3,000	500	6,000	160	500	8,000
15	10	4	8	1000	5,000	10,000	300	120	5,000

(b) Decay coefficients.

$\phi_1(i,j)$					$\phi_2(i,j)$				
.7	.3	.0	.0	.0	.0	.1	.2	.3	.4
.1	.8	.1	.0	.0	.2	.1	.4	.2	.1
.0	.1	.6	.2	.1	.1	.3	.1	.3	.2
.2	.2	.3	.1	.2	.0	.0	.1	.8	.1
.1	.3	.3	.1	.2	.1	.0	.1	.1	.7

(c) Transition distribution.

ν_i
5
1

(d) Base rate vector.

Table 3.2: Parameter values for Specification 2.

in (3.4), denoted by l_+ and l_-, respectively. Here, we consider a general time horizon $(t_0, T]$, meaning that the origin of time is not necessarily $t_0 = 0$ and that the times $t_n^{ex} \leq t_0$ are treated like an initial condition, and we have

$$
\begin{aligned}
l_+ &= \sum_e \sum_{t_n^e > t_0} \ln \lambda_e(t_n^e) = \sum_e \sum_{t_n^e > t_0} \ln \left(\nu_e + \sum_{e'x'} \alpha_{e'x'e} S_{e'x'e}(t_n^e) \right), \\
S_{e'x'e}(s,t) &:= \sum_{s < t_i^{e'x'} < t} \exp(-\beta_{e'x'e}(t - t_i^{e'x'})), \\
S_{e'x'e}(t) &:= S_{e'x'e}(-\infty, t), \\
S_{e'x'e}(t) &= \exp(-\beta_{e'x'e}(t-s)) S_{e'x'e}(-\infty, s] + S_{e'x'e}(s,t), \\
l_- &= \int_{t_0}^T \sum_e \lambda_e(t) dt \\
&= \int_{t_0}^T \sum_e \left(\nu_e + \sum_{e'x'} \sum_{t_i^{e'x'} < t} \alpha_{e'x'e} \exp(-\beta_{e'x'e}(t - t_i^{e'x'}) \right) dt \\
&= \sum_e \nu_e (T - t_0) \\
&\quad + \sum_{e'x'} \sum_{t_i^{e'x'} \leq t_0} \frac{\alpha_{e'x'e}}{\beta_{e'x'e}} (\exp(-\beta_{e'x'e}(t_0 - t_i^{e'x'})) - \exp(-\beta_{e'x'e}(T - t_i^{e'x'})) \\
&\quad + \sum_{t_i^{e'x'} > t_0} \frac{\alpha_{e'x'e}}{\beta_{e'x'e}} (1 - \exp(-\beta_{e'x'e}(T - t_i^{e'x'})).
\end{aligned}
$$

The gradients can then also be computed via

$$
\begin{aligned}
\frac{\partial l_+}{\partial \nu_k} &= \sum_{t_n^k > t_0} \left(\nu_k + \sum_{e'x'} \alpha_{e'x'k} S_{e'x'k}(t_n^k) \right)^{-1}, \\
\frac{\partial l_+}{\partial \alpha_{ijk}} &= \sum_{t_n^k > t_0} \frac{S_{ijk}(t_n^e)}{\nu_k + \sum_{e'x'} \alpha_{e'x'k} S_{e'x'k}(t_n^k)}, \\
\frac{\partial l_+}{\partial \beta_{ijk}} &= - \sum_{t_n^k > t_0} \frac{\alpha_{ijk} S_{ijk}^{(1)}(t_n^e)}{\nu_k + \sum_{e'x'} \alpha_{e'x'k} S_{e'x'k}(t_n^k)}, \\
S_{e'x'e}^{(1)}(s,t) &:= \sum_{s < t_i^{e'x'} < t} (t - t_i^{e'x'}) \exp(-\beta_{e'x'e}(t - t_i^{e'x'})),
\end{aligned}
$$

$$
\begin{aligned}
S^{(1)}_{e'x'e}(t) &:= S^{(1)}_{e'x'e}(-\infty, t), \\
S^{(1)}_{e'x'e}(t) &= \exp(-\beta_{e'x'e}(t-s)) \left(S^{(1)}_{e'x'e}(-\infty, s] + (t-s) S_{e'x'e}(-\infty, s] \right) \\
&\quad + S^{(1)}_{e'x'e}(s, t), \\
\frac{\partial l_-}{\partial \nu_k} &= T - t_0, \\
\frac{\partial l_-}{\partial \alpha_{e'x'e}} &= \sum_{t_i^{e'x'} \leq t_0} \frac{1}{\beta_{e'x'e}} (\exp(-\beta_{e'x'e}(t_0 - t_i^{e'x'})) - \exp(-\beta_{e'x'e}(T - t_i^{e'x'})) \\
&\quad + \sum_{t_i^{e'x'} > t_0} \frac{1}{\beta_{e'x'e}} (1 - \exp(-\beta_{e'x'e}(T - t_i^{e'x'})), \\
\frac{\partial l_-}{\partial \beta_{e'x'e}} &= \sum_{t_i^{e'x'} \leq t_0} \frac{\alpha_{e'x'e}}{\beta_{e'x'e}} ((T - t_i^{e'x'}) \exp(-\beta_{e'x'e}(T - t_i^{e'x'})) \\
&\qquad - (t_0 - t_i^{e'x'}) \exp(-\beta_{e'x'e}(t_0 - t_i^{e'x'}))) \\
&\quad - \frac{\alpha_{e'x'e}}{\beta^2_{e'x'e}} (\exp(-\beta_{e'x'e}(t_0 - t_i^{e'x'})) - \exp(-\beta_{e'x'e}(T - t_i^{e'x'})) \\
&\quad + \sum_{t_i^{e'x'} > t_0} \frac{\alpha_{e'x'e}}{\beta_{e'x'e}} (T - t_i^{e'x'}) \exp(-\beta_{e'x'e}(T - t_i^{e'x'})) \\
&\qquad - \frac{\alpha_{e'x'e}}{\beta^2_{e'x'e}} (1 - \exp(-\beta_{e'x'e}(T - t_i^{e'x'}))).
\end{aligned}
$$

Besides, the efficient computation of the residuals is based on the identity

$$
\begin{aligned}
\int_s^t \lambda_e(u) du = (t-s)\nu_e + \sum_{e'x'} C_{e'x'e}(t) - C_{e'x'e}(s) \\
- \sum_{e'x'} \frac{\alpha_{e'x'e}}{\beta_{e'x'e}} (S_{e'x'e}(t) - S_{e'x'e}(s)), \\
C_{e'x'e}(t) := \sum_{t_i^{e'x'} < t} \frac{\alpha_{e'x'e}}{\beta_{e'x'e}} \exp(-\beta_{e'x'e}(t_0 \vee t_i^{e'x'} - t_i^{e'x'})).
\end{aligned}
$$

The accompanying Python library `mpoints` implements the above formulae in C via the Cython extension of Python, allowing us to drastically reduce the computation time (up to 300 times faster computations compared to a plain Python implementation using NumPy). This played a crucial role in the feasibility of the study in Chapter 4.

Proof of Theorem 3.2.1. The statement follows by applying Theorem 6.2.5 in Chapter 6. We only need to check that if $(\boldsymbol{N}, X)$ is a state-dependent Hawkes process, then $\tilde{\boldsymbol{N}}$ admits an $\mathbb{F}^{\tilde{\boldsymbol{N}}}$-intensity. By Proposition 3.1 in Jacod [68], $\tilde{\boldsymbol{N}}$ admits an $\mathbb{F}^{\tilde{\boldsymbol{N}}}$-compensator Λ given by

$$\Lambda(dt, de, dx) = \sum_{n \in \mathbb{N}} \frac{G_n(dt - T_n, de, dx)}{G_n([t - T_n, \infty], \mathcal{E}, \mathcal{X})} \mathbb{1}_{\{T_n < t \leq T_{n+1}\}},$$

and, consequently, this holds if the conditional distributions G_n of the triplets $(T_{n+1} - T_n, E_{n+1}, X_{n+1})$ with respect to $\mathcal{F}^{\tilde{\boldsymbol{N}}}_{T_n}$ are absolutely continuous with respect to $dt\mu_e(de)\mu_x(dx)$ on $(0, T_{n+1} - T_n] \times \mathcal{E} \times \mathcal{X}$, where μ_e and μ_x are the counting measures on $\mathcal{E}$ and $\mathcal{X}$, respectively. Moreover, by definition of $\tilde{\boldsymbol{N}}$, $\Lambda(dt, de, \mathcal{X})$ is an $\mathbb{F}^{\tilde{\boldsymbol{N}}}$-compensator of $\boldsymbol{N}$. But since $\boldsymbol{N}$ admits an $\mathbb{F}^{\tilde{\boldsymbol{N}}}$-intensity, by uniqueness of the compensator [75, Theorem 1.25, p. 39], we necessarily have with probability one that

$$\sum_{x \in \mathcal{X}} G_n(dt - T_n, \{e\}, \{x\}) \mathbb{1}_{\{T_n < t \leq T_{n+1}\}} = f(t, e)dt \mathbb{1}_{\{T_n < t \leq T_{n+1}\}}, \quad e \in \mathcal{E}, \quad n \in \mathbb{N},$$

for some $\mathcal{F}^{\tilde{\boldsymbol{N}}}_{T_n} \otimes \mathcal{B}(\mathbb{R}) \otimes \mathcal{P}(\mathcal{E})$-measurable function f, which concludes the proof. □

Proof of Theorem 3.2.2. By Theorem 3.2.1, it is sufficient to show that each of the two conditions ensures the existence and uniqueness of a non-explosive point process $\tilde{\boldsymbol{N}}$ with an $\mathbb{F}^{\tilde{\boldsymbol{N}}}$-intensity given by (3.3), which is achieved by applying Theorems 7.4.1 and 7.5.2 in Chapter 7. □

Proof of Theorem 3.4.1. By Theorem 3.2.1, we know that $\tilde{\boldsymbol{N}}$ has intensity $\tilde{\boldsymbol{\lambda}}$, which is given by (3.3). Hence, by applying Proposition 7.3.III in Daley and Vere-Jones [40, p. 251], we can express the log likelihood function as

$$\ln \mathcal{L}(\boldsymbol{\phi}, \boldsymbol{\nu}, \boldsymbol{\theta}) = \sum_{n=1}^{N} \ln \tilde{\lambda}_{e_n x_n}(t_n) - \int_0^T \sum_{e \in \mathcal{E}, x \in \mathcal{X}} \tilde{\lambda}_{ex}(t) dt. \tag{3.8}$$

Plugging (3.3) in (3.8) and using that $\sum_{x' \in \mathcal{X}} \phi_e(x, x') = 1$, $e \in \mathcal{E}$, $x \in \mathcal{X}$, yields (3.4), from which it is immediate that $(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\nu}}, \hat{\boldsymbol{\theta}}) \in \arg\max_{\boldsymbol{\phi}, \boldsymbol{\nu}, \boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\phi}, \boldsymbol{\nu}, \boldsymbol{\theta})$ if and only if

$$\begin{cases} \hat{\boldsymbol{\phi}} & \in \arg\max\limits_{\boldsymbol{\phi}} \sum\limits_{n=1}^{N} \ln \phi_{e_n}(x_{n-1}, x_n), \\ (\hat{\boldsymbol{\nu}}, \hat{\boldsymbol{\theta}}) & \in \arg\max\limits_{\boldsymbol{\nu}, \boldsymbol{\theta}} \sum\limits_{n=1}^{N} \ln \lambda_{e_n}(t_n) - \int_0^T \sum\limits_{e \in \mathcal{E}} \lambda_e(t) dt, \end{cases}$$

where the first optimisation problem is performed under the constraint that $\boldsymbol{\phi_e}$ is a transition probability matrix, $e \in \mathcal{E}$. By solving this optimisation problem with the method of Lagrange multipliers, we obtain the claimed expression for $\hat{\phi}_e(x, x')$. □

Proof of Theorem 3.4.3. The statement follows directly from Theorem 1 in Brown and Nair [22]. Note that this theorem requires that, with probability one, $\int_0^t \lambda_e(s) ds \to \infty$ and $\int_0^t \tilde{\lambda}_{ex}(s) ds \to \infty$ as $t \to \infty$, $e \in \mathcal{E}$, $x \in \mathcal{X}$. The first condition is satisfied because, in Definition 3.1.1, we assume that all the base rates are strictly positive ($\boldsymbol{\nu} \in \mathbb{R}^{d_e}_{>0}$). By Lemma 17 in Brémaud [19, p. 41], the second condition is equivalent to $\tilde{N}_{ex}(t) \to \infty$, $t \to \infty$, with probability one, $e \in \mathcal{E}$, $x \in \mathcal{X}$, which is assumed here. □

4

Application to high-frequency financial data

We apply the new model and methodology of the previous chapter to high-frequency financial data, estimating for the first time a LOB model that accommodates an explicit feedback loop between the state of the LOB and the order flow with self- and cross-excitation effects. We estimate two specifications of a state-dependent Hawkes process on one year of level-I LOB data, using the bid–ask spread and QI, respectively, as the state variable. Our estimation results reveal that the magnitude and speed of excitation effects in the order flow depend significantly on both state variables, being effective at timescales ranging from 100 microseconds to 100 milliseconds. This is a very clear manifestation of the high-frequency automation of trading in modern electronic markets. Moreover, we find that the level of endogeneity in the order flow is state-dependent and, intriguingly, more pronounced in what can be regarded as *disequilibrium states* of the LOB.

4.1 Data and descriptive statistics

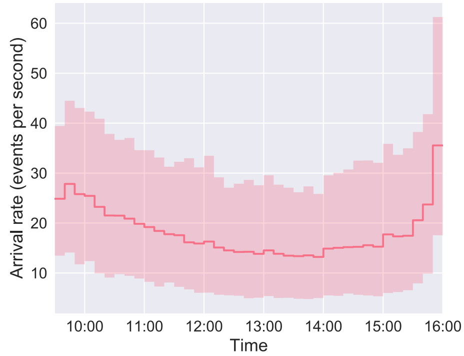
We use level-I data on the stock of Intel (INTC), traded on the Nasdaq, from the 1st of June 2017 to the 31st of May 2018. The 3rd of July and 24th of November 2017 are removed from the data set because of early market close at 13:00 (Independence Day and Black Friday), which leaves us with 250 full trading days. Besides INTC, we have also studied in the same manner the stocks of Advanced Micro Devices (AMD), Micron Technology (MU), Snap (SNAP) and Twitter (TWTR) from January to April 2018, but we only report our results on INTC as they are largely representative of the findings on the other stocks.

The first three stocks (INTC, AMD, MU) and the last two stocks (SNAP, TWTR) are primarily listed on the Nasdaq and the New York Stock Exchange (NYSE), respectively. As an illustration of the current market fragmentation (cf. Subsection 1.1.5), around 20% of the total volume is traded on the Nasdaq for the first group (AMD, INTC, MU), while nearly 10% for the second group (SNAP, TWTR). In particular, Nasdaq has the largest market share for INTC during the observation period. These estimates of Nasdaq's market share were given by Fidessa's Fragulator tool*.

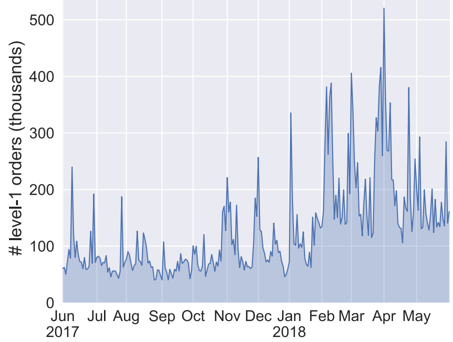
While there is no universally-accepted criterion to decide of the large-tick character of a stock, the prices of our five stocks were most the time below the threshold of $50 per share used by Bonart and Gould [14] and always below $60. A defining characteristic of large-tick stocks is that the spread is most often equal to one tick, which can be indeed checked in Figure 4.2a.

We will focus on the trading activity that begins at 12pm and ends at 2.30pm for the following reason. As depicted in Figure 4.1a, it is known that market activity is not constant throughout the day but tends to exhibit a U-shape (cf. Fact (v)). Because of that, enforcing a constant base rate vector $\boldsymbol{\nu}$ over the entire trading day can result in an overestimation of the self- and cross-excitation effects [107, 101]. The chosen time horizon exhibits a relatively small variation in market activity while still offering a large-enough sample

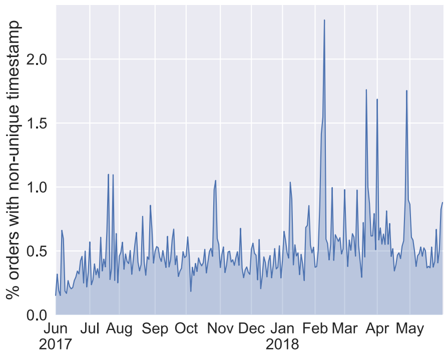
*http://fragmentation.fidessa.com/fragulator/



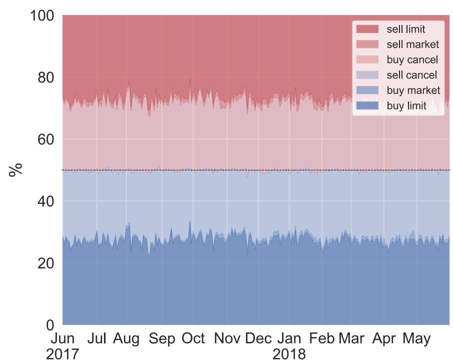
(a) Arrival rate of level-I orders.



(b) Number of level-I orders.



(c) Fraction of level-I orders with non-unique timestamp.



(d) Distribution of level-I order types.

Figure 4.1: Descriptive statistics of the level-I order flow of INTC. Except for Figure 4.1a, only the data between 12:00 and 14:30 are used. In Figure 4.1a, the sample mean of the arrival rate of level-I orders is computed over 10-minute bins. The translucent area represents the range of the arrival rate across the 250 trading days, excluding the bottom and top 5% values.

size (at least 50,000 level-I orders each day, see Figure 4.1b).

We stress again that timestamps are recorded with nanosecond precision (cf. Section 2.1). As a result, more than 99% of level-I orders have a unique timestamp (Figure 4.1c). This contrasts with older studies [18, 78] where lower time resolutions (e.g., one second) were used, leading to a significant amount of *tied* timestamps, shared by multiple events. As Hawkes processes capture the Granger causality between different event types [45], being able to accurately order events, even at the smallest timescales, is essential for the accurate estimation of Hawkes processes.

4.2 Model specification

We work with models, based on state-dependent Hawkes processes, that distinguish two event types, denoted ask and bid (that is, $\mathcal{E} = \{\mathsf{ask}, \mathsf{bid}\}$, $d_e = 2$). These event types are nothing else than the ask and bid events in the definition of the OFI (cf. Definition 1.2.11). Consequently, $N_{\mathsf{bid}} - N_{\mathsf{ask}}$ can be interpreted as a proxy of the OFI. From Figure 4.1d, we see that the distribution between these two event types is very balanced, with market orders accounting for less than 5% of the level-I activity. As mentioned above, only the level-I events are modelled and events occurring at deeper levels of the LOB are discarded. One could of course increase d_e to have a more granular classification of event types. Still, the focus of this chapter is on the modelling of state dependence and the choice $\mathcal{E} = \{\mathsf{ask}, \mathsf{bid}\}$ already leads to interesting results whilst keeping the dimensionality low, which makes the results easier to visualise.

As the state variable we consider the bid–ask spread and the QI (cf. Definition 1.2.10) , giving rise to two models named Model_S and Model_QI, respectively. In Model_S, we set $\mathcal{X} = \{\mathsf{1}, \mathsf{2+}\}$ ($d_x = 2$), where the states correspond to the spread being one tick ($X(t) = \mathsf{1}$) and two ticks or more ($X(t) = \mathsf{2+}$). Increasing the number of states beyond $d_x = 2$ in this setting would not be practically relevant since the spread is very rarely strictly wider than two ticks. In Model_QI, as in Cartea et al. [23], we split the interval $[-1, 1]$ into $d_x = 5$ bins of equal width which we label as follows:

$$\mathcal{X} = \{ \underbrace{\mathsf{sell{+}{+}}}_{[-1,-0.6)}, \underbrace{\mathsf{sell{+}}}_{[-0.6,-0.2)}, \underbrace{\mathsf{neutral}}_{[-0.2,0.2)}, \underbrace{\mathsf{buy{+}}}_{[0.2,0.6)}, \underbrace{\mathsf{buy{+}{+}}}_{[0.6,1]} \}, \tag{4.1}$$

whereby in Model_QI the state variable $X(t)$ indicates bin where $QI(t)$ is located.

Finally, given the large number of observations we are dealing with (see Figure 4.1b), we use the exponential specification (3.6) of the kernel $\boldsymbol{k}$ in both models, as it leads to a significant reduction of computational cost in estimation, as discussed in Subsection 3.4.3. The full specifications of the models are summarised in Table 4.1.

	$\mathbf{Model_S}$	$\mathbf{Model_{QI}}$
Event types	Ask and bid events, $d_e = 2$, $\mathcal{E} = \{\mathsf{ask}, \mathsf{bid}\}$	
State process	Spread, $d_x = 2$	Queue imbalance, $d_x = 5$
	$\mathcal{X} = \{\mathsf{1}, \mathsf{2+}\}$	$\mathcal{X} = \{\mathsf{sell++}, \mathsf{sell+}, \mathsf{neutral}, \mathsf{buy+}, \mathsf{buy++}\}$
Kernels	Exponential, of the form (3.6)	
# params.	26	92

Table 4.1: Summary of $\mathrm{Model_S}$ and $\mathrm{Model_{QI}}$.

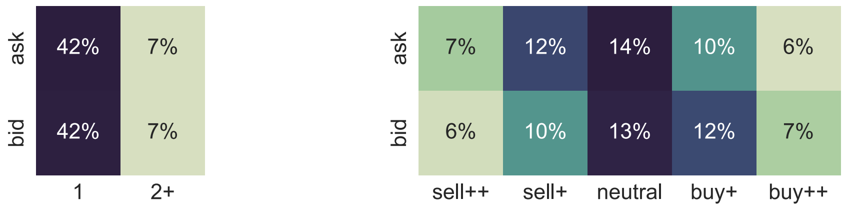
4.3 Visualising the estimated excitation effects

We use the following approach to present our estimation results on self- and cross-excitation effects. For each triple $(e, e', x) \in \mathcal{E}^2 \times \mathcal{X}$, ML estimation produces an estimated excitation profile $t \mapsto k_{e'e}(t, x)$, parameterised by the estimated impact coefficient $\hat{\alpha}_{e'xe}$ and decay coefficient $\hat{\beta}_{e'xe}$. However, instead of reporting $\hat{\alpha}_{e'xe}$ and $\hat{\beta}_{e'xe}$, we visualise the excitation profile by plotting the truncated L^1-norm

$$t \mapsto \|\hat{k}_{e'e}(\cdot, x)\|_{1,t} := \int_0^t \hat{k}_{e'e}(s, x)ds,$$

from which the magnitude and the effective timescale of the excitation effect is easier to gauge than from the numerical values of $\hat{\alpha}_{e'xe}$ and $\hat{\beta}_{e'xe}$. Extrapolating from the cluster representation of simple Hawkes processes (Subsection 1.6.3), we can interpret $\|\hat{k}_{e'e}(\cdot, x)\|_{1,t}$ as the average number of events of type e that have been directly triggered by an event of type e' in state x within t seconds of its occurrence. Besides, notice that the full L^1-norm is given by

$$\|\hat{k}_{e'e}(\cdot, x)\|_{1,\infty} = \lim_{t\to\infty} \|\hat{k}_{e'e}(\cdot, x)\|_{1,t} = \frac{\hat{\alpha}_{e'xe}}{\hat{\beta}_{e'xe}}.$$



(a) $\mathrm{Model}_{\mathrm{S}}$ (state process: bid–ask spread).

(b) $\mathrm{Model}_{\mathrm{QI}}$ (state process: queue imbalance).

Figure 4.2: Joint distribution of events and states for INTC, depicting the empirical distribution of the marks (E_n, X_n) for the two considered state processes.

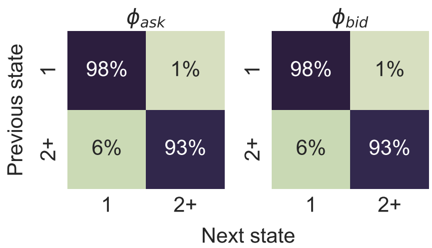
4.4 Estimation results for $\mathrm{Model}_{\mathrm{S}}$

We estimate the model parameters $(\hat{\boldsymbol{\phi}}^{(i)}, \hat{\boldsymbol{\nu}}^{(i)}, \hat{\boldsymbol{\alpha}}^{(i)}, \hat{\boldsymbol{\beta}}^{(i)})$ for each trading day i in the sample by ML estimation, as explained in Section 3.4. Practical details on the numerical solution of the underlying optimisation problem will be given in Subsection 4.10.1.

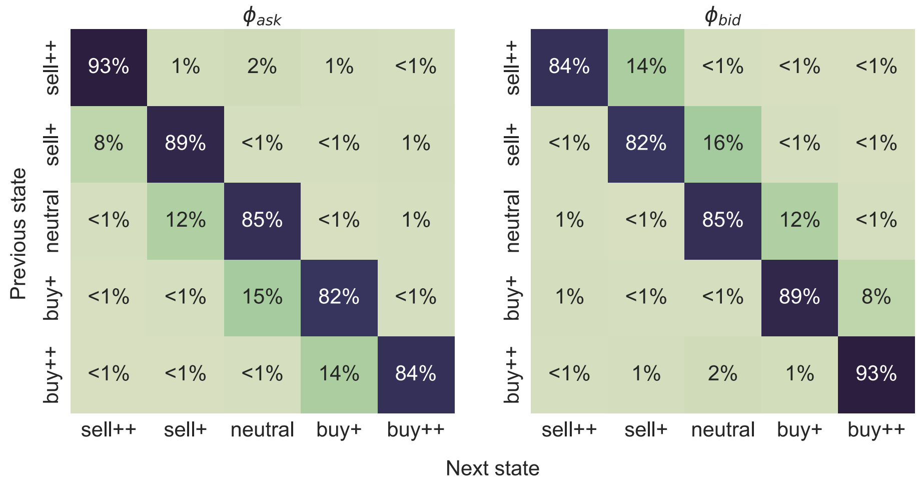
The estimated transition distribution $\hat{\boldsymbol{\phi}}$ of $\mathrm{Model}_{\mathrm{S}}$, obtained by averaging over the daily estimates $\hat{\boldsymbol{\phi}}^{(i)}$, is presented in Figure 4.3a. The state process describing the bid–ask spread exhibits persistent behaviour, in the sense that the probability of remaining in the current state is very high. We also observe higher likelihood of moving from state 2+ to 1 than vice versa, which is consistent with one-tick bid–ask spread being the equilibrium state for a large-tick stock like INTC. We also find that the transition probabilities are not sensitive to the event type, which is natural since the bid–ask spread not is expected to be influenced by the direction of orders, ceteris paribus.

The estimation results on the excitation kernel, given in Figure 4.4a, indicate that self-excitation effects surpass cross-excitation effects in both states. Their magnitude and timescales, however, vary between the two states. The effective timescales of these effects range from 0.1 to 100 milliseconds, which is in agreement with the predominantly algorithmic origin and multiscale nature of trading in modern electronic markets.

When the bid–ask spread increases to 2+, the magnitude of the self-excitation effects doubles whilst their timescale remains roughly the same. The timescale



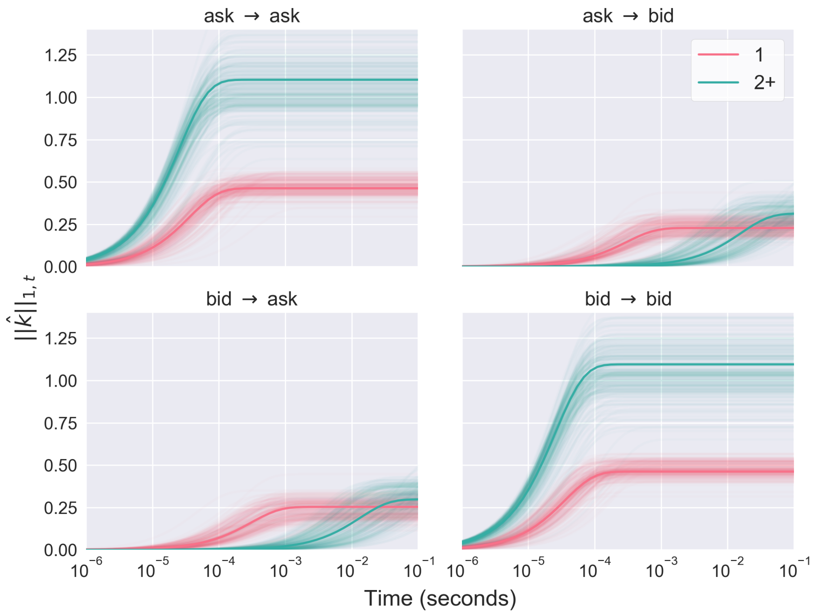
(a) Transition probabilities of the bid–ask spread ($\mathrm{Model_S}$).



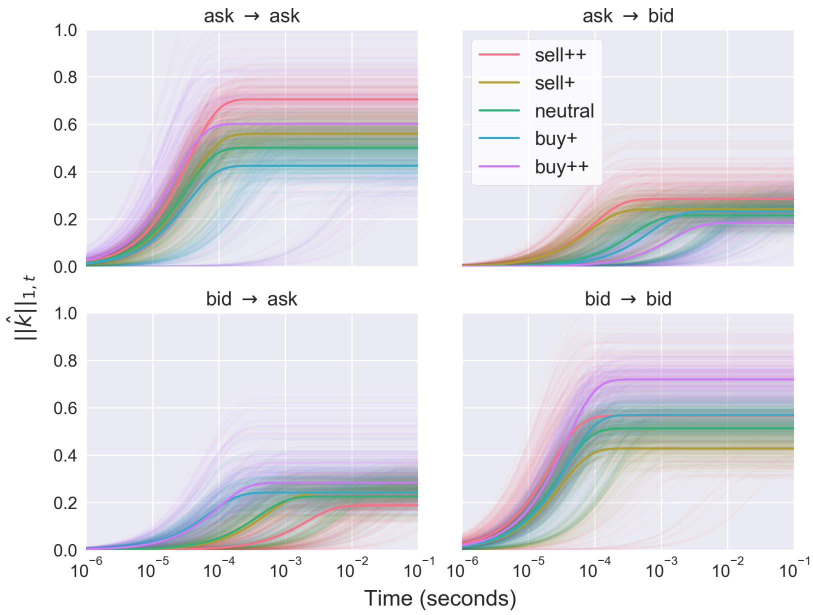
(b) Transition probabilities of the queue imbalance ($\mathrm{Model_{QI}}$).

Figure 4.3: Estimated transition distributions $\hat{\phi}$ of $\mathrm{Model_S}$ and $\mathrm{Model_{QI}}$. We report the average of $\hat{\phi}^{(i)}$ across the 250 trading days. (Daily estimates vary little from these averaged values.)

of cross-excitation, however, lengthens drastically, whilst their magnitude increases slightly. A plausible microstructural explanation for this pattern goes as follows. When the bid–ask spread is in state 2+, a trader can submit an aggressive limit order inside the spread, gaining queue priority at the cost of a less favourable price. A bid event can then be seen as a signal for an upwards price move, which may prompt limit orders from buyers seeking a favourable position in the new queue and cancellations of limit orders from sellers trying to avoid adverse selection. Traders who submit sell orders are, per contra, incentivised to wait and see if the expected price increase actually materialises.

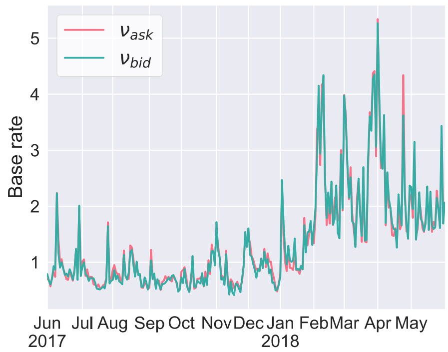


(a) $\mathrm{Model}_{\mathrm{S}}$ (state variable: bid–ask spread).

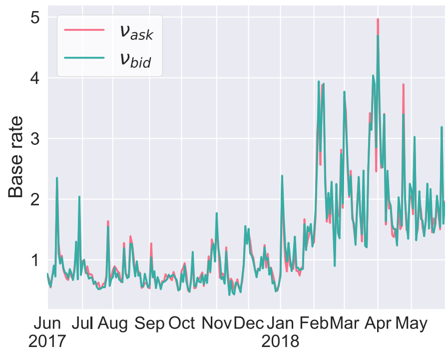


(b) $\mathrm{Model}_{\mathrm{QI}}$ (state variable: queue imbalance).

Figure 4.4: The estimated kernel $\hat{\boldsymbol{k}}$ under $\mathrm{Model}_{\mathrm{S}}$ and $\mathrm{Model}_{\mathrm{QI}}$. Each panel describes self- or cross-excitation as indicated by its title, whilst each colour corresponds to a different state. For example, in Figure 4.4a, the red curves in the second panel represent the estimates $\hat{k}^{(i)}_{e'e}(\cdot, x)$ where $e' = \text{ask}$, $e = \text{bid}$ and $x = 1$. All daily estimates are superposed with one translucent curve for each day. An "aggregate" kernel is represented by a solid line, computed using the median of $\hat{\boldsymbol{\alpha}}^{(i)}$ and $\hat{\boldsymbol{\beta}}^{(i)}$ across the 250 trading days.



(a) Model_S.



(b) Model_QI.

Figure 4.5: Estimated base rate vector $\hat{\boldsymbol{\nu}}^{(i)}$ for Model_S and Model_QI over time (in number of events per second).

The evolution of the base rate vector $\hat{\boldsymbol{\nu}}^{(i)}$ throughout the 250 days of data is displayed in Figure 4.5. We find a remarkable balance between buyers and sellers (i.e., $\nu_{\mathsf{bid}} \approx \nu_{\mathsf{ask}}$). We also notice that the evolution of $\hat{\boldsymbol{\nu}}^{(i)}$ mimics that of the total number of orders (Figure 4.1b), which suggests that the day-to-day variation in market activity is mainly due to exogenous factors (cf. the analysis of endogeneity in Section 4.7).

4.5 Estimation results for Model_QI

The estimated transition probabilities of Model_QI, presented in Figure 4.3b, convey a tendency to stick to the current state, similar to what is seen in Model_S. Here, however, this behaviour is more of an artefact — each ask and bid event, by definition, changes the queue imbalance but not necessarily the state variable that is confined to the bins (4.1).

In contrast to Model_S, the estimated transition probabilities now depend on the event type and we observe remarkable mirror symmetry, whereby $\hat{\phi}_{\mathsf{bid}}$ equals, up to 1 percentage point, $\hat{\phi}_{\mathsf{ask}}$ with the order of states reversed. This symmetry is natural, given the definition of the QI — a sell order always decreases the queue imbalance unless it is submitted inside the bid–ask spread or it depletes the current bid queue, whilst an analogous statement is true for buy orders. As in Model_S, we find again that the probability of a state

Figure 4.6: Estimated kernel norms $\|\hat{\boldsymbol{k}}\|_{1,\infty}$ for $\mathrm{Model}_{\mathrm{QI}}$ over time. The dashed line marks 5 February 2018 ("the return of volatility"), a day when the CBOE Volatility Index (VIX) jumped by 116% to 38 points, a level not seen since August 2015. This day seems to have introduced a systematic change in the magnitude of cross- and self-excitation. The spike in cross-excitation that occurs on 21 February 2018 is linked to a sudden change in market behaviour around 14:00 on that day, when Intel Corporation rolled out patches for its most recent generation of processors.

transition is higher when it is towards the equilibrium state, here neutral, consistent with ideas about the *resilience* of the LOB [78].

Looking at the estimation results for the excitation kernel in Figure 4.4b, we observe that, like in $\mathrm{Model}_{\mathrm{S}}$, self-excitation surpass cross-excitation, with the magnitude of the former and the timescale of the latter being manifestly sensitive to the current state. The mirror symmetry seen above in the context of the transition probabilities holds here as well, whereby it suffices to only speak about the results for ask events, whilst analogous conclusions can be drawn on bid events.

Even though the day-to-day variation of the estimates is more pronounced in this model compared to $\mathrm{Model}_{\mathrm{S}}$, some clear patterns emerge again. The self-excitation of ask events consistently increases under heavy sell pressure (state sell++). This can plausibly be explained by a combination of a flight to liquidity, through the submission of sell orders, and fear of adverse selection, leading to cancellations of buy orders, by traders expecting a downwards price move. Under mild buy pressure (state buy+), self-excitation decreases so that the corresponding kernel norm nearly halves.

In the state buy++ (heavy buy pressure), a closer look at the daily estimates of the self-excitation of ask events plotted in purple in Figure 4.4b reveals two distinct groups of curves above and below of the median curve. The daily estimates of the (full) kernel norms plotted in time in Figure 4.6 suggest that the self-excitation effect has undergone a structural break in early February 2018 — the moment that suddenly marked the end of a year-long period of unusually low volatility in the US equity markets, dubbed "the return of volatility" by some financial journalists. It is also worth mentioning that at the same time the price of INTC went above \$50, at which point the large-tick character of the stock starts to weaken. The behaviour characterised by the lower group of curves (pre-February 2018) can be interpreted as traders expecting a price increase and thus deferring the submission of sell orders, whereas the upper group of curves (post-February 2018) hints at a tendency of sellers to seek an advantageous position in the ask queue. Indeed, a queue imbalance close to one implies that that only a very small amount of liquidity is available at the ask price. Thus, placing a sell limit order following a succession of sell limit orders from other traders allows one to acquire a good position in the queue, should it be replenished (which brings the queue imbalance back to equilibrium). If, however, the ask queue progresses towards depletion, one has still time to cancel the order whilst the sell limit orders at the front of the queue are matched with incoming buy market orders.

Under buy pressure (states buy++ and buy+), the timescale of the cross-excitation from bid to ask events becomes almost as short as that of the self-excitation of bid events. This reflects the resilience of the LOB — in response to a bid event, ask events compete neck and neck with bid events precisely when they push the queue imbalance back towards the equilibrium state. Recall that the resilience of the LOB is also reinforced by the estimated

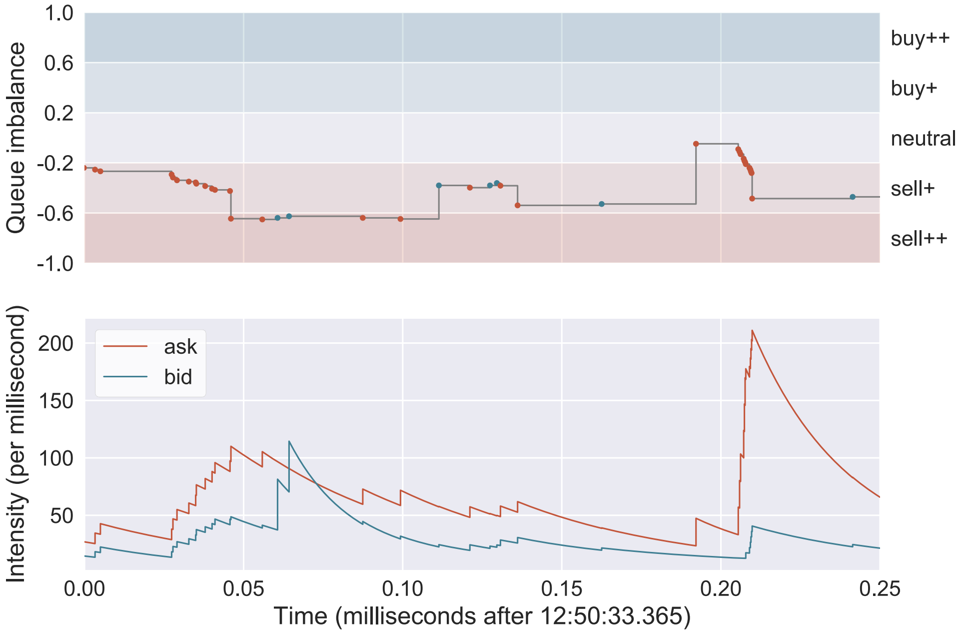


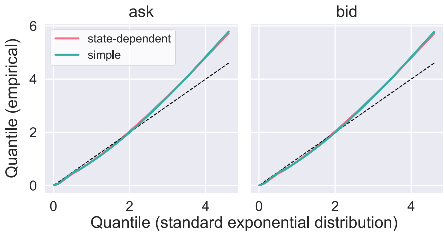
Figure 4.7: The upper panel depicts the evolution of the queue imbalance and level-I order flow of INTC on 13 February 2018 and the ask (red dots) and bid (blue dots) events. The lower panel displays the estimated intensities of $\mathrm{Model}_{\mathrm{QI}}$. (The self- and cross-excitation kernel norms of $\mathrm{Model}_{\mathrm{QI}}$ on 13 February 2018 are visualised in Figure 4.11.)

state transition probabilities, as discussed above.

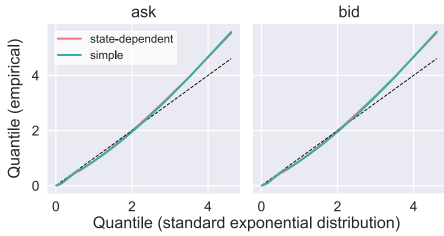
To exemplify the estimated intensity processes and their state dependence, a very brief extract from the estimated dynamics of $\mathrm{Model}_{\mathrm{QI}}$ is presented in Figure 4.7. In particular, we observe how pronounced the self-excitation of bid events becomes when the queue imbalance drops below -0.6, that is, to state sell++.

4.6 Goodness-of-fit diagnostics

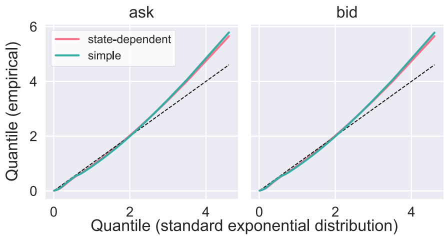
To assess the goodness of fit of the estimated models, we examine the event residuals r_n^e in sample (daily, 12:00–14:30) and out of sample (daily, 14:30–15:00) for both $\mathrm{Model}_{\mathrm{S}}$ and $\mathrm{Model}_{\mathrm{QI}}$. For comparison, we also estimated an ordinary Hawkes process ($d_x = 1$) with exponential kernel for the same event types $\mathcal{E} = \{$ask, bid$\}$ and computed its event residuals. Since the state-



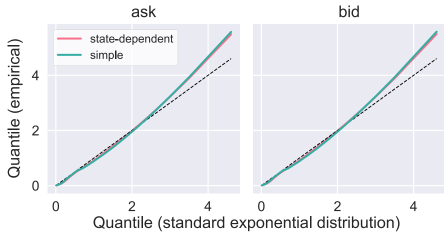
(a) $\mathrm{Model}_{\mathrm{S}}$: in sample.



(b) $\mathrm{Model}_{\mathrm{S}}$: out of sample.



(c) $\mathrm{Model}_{\mathrm{QI}}$: in sample.



(d) $\mathrm{Model}_{\mathrm{QI}}$: out of sample.

Figure 4.8: In-sample (12:00–14:30) and out-of-sample (14:30–15:00) Q–Q plots of event residuals under $\mathrm{Model}_{\mathrm{S}}$, $\mathrm{Model}_{\mathrm{QI}}$ (state-dependent) and an ordinary Hawkes process (simple). The residuals of the ith day are computed using the ML estimates $(\hat{\boldsymbol{\nu}}^{(i)}, \hat{\boldsymbol{\alpha}}^{(i)}, \hat{\boldsymbol{\beta}}^{(i)})$ obtained from the 12:00–14:30 period. The empirical quantiles are obtained by pooling the residuals of all 250 trading days. The two panels in each sub-figure correspond to the sequences of residuals (r_n^e) for $e \in \mathcal{E} = \{\mathsf{ask}, \mathsf{bid}\}$.

dependent Hawkes process nests the ordinary Hawkes process, the former will by construction provide a better fit in sample than the latter. The Q–Q plots in Figure 4.8 show that this improvement in the goodness of fit extends out of sample, albeit the improvement is smaller than in sample. This is a confirmation that that $\mathrm{Model}_{\mathrm{S}}$ and $\mathrm{Model}_{\mathrm{QI}}$, and their state-dependent features in particular, are not overfitted.

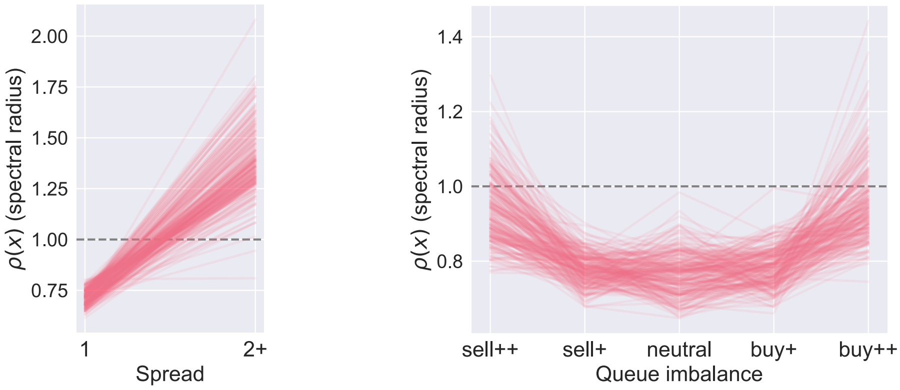
It should not come as a surprise that the improvement in goodness of fit provided by the state-dependent model looks meagre when one examines the Q–Q plots. Indeed, the behaviour of $\mathrm{Model}_{\mathrm{S}}$, $\mathrm{Model}_{\mathrm{QI}}$ and their ordinary Hawkes process alternative is quite similar when the bid–ask spread and queue imbalance are in their most likely states. It is only in the less likely states, 2+ in $\mathrm{Model}_{\mathrm{S}}$ and sell++ and buy++ in $\mathrm{Model}_{\mathrm{QI}}$, where the difference between the state-dependent and ordinary Hawkes process models becomes more pronounced. Thereby, the unconditional distribution of residuals does not vary much between these three models.

The recent popularity of Hawkes processes in the modelling of high-frequency financial data partly stems from their ability to capture the *endogeneity* of market activity. In particular, recall from Section 1.6.3 that the spectral radius of the matrix of kernel norms, denoted by ρ, can be interpreted as a measure of endogeneity and that the condition $\rho < 1$ ensures the stability of the process.

Univariate Hawkes processes fitted to high-frequency financial data often exhibit kernel norms slightly below one, whilst the base rates tend to be relatively low in comparison. This phenomenon has been interpreted by Filimonov and Sornette [49] and Hardiman et al. [58] as evidence that most market events are endogenous, mere responses to earlier events, dwarfing the flow of less frequent exogenous events that are driven by new information. They dub the phenomenon *critical reflexivity*, which is a nod to George Soros and his reflexivity theory on the endogeneity of financial markets [114].

In the context of state-dependent Hawkes processes, for any state $x \in \mathcal{X}$, the kernel $(k_{e'e}(\cdot, x))_{e',e \in \mathcal{E}}$ defines a multivariate ordinary Hawkes process. Setting $m_{ij} := \|k_{ji}(\cdot, x)\|_{1,\infty}$, we denote by $\rho(x)$ the spectral radius of the $d_e \times d_e$ matrix (m_{ij}), which measures the endogeneity in this ordinary Hawkes process. Figure 4.9 displays the daily estimates of $\rho(x)$ for both Model_{S} and Model_{QI} as a function of $x \in \mathcal{X}$. We observe a remarkably clear pattern of $\rho(x)$ being almost uniformly higher in the disequilibrium states (2+, sell++, buy++) than in the equilibrium states (1, sell+, neutral, buy+). In particular, in Model_{S}, the spectral radius $\rho(\text{2+})$ is systematically above the critical value 1, whilst in Model_{QI}, the values of $\rho(\text{sell++})$ and $\rho(\text{buy++})$ are above 0.9 half of the time, exceeding 1 occasionally. These results thus open a new perspective on critical reflexivity, showing that it is in fact a largely state-dependent phenomenon, observed only in particular circumstances. They also lend credence to Soros's remark that *"[e]ven in the financial markets demonstrably reflexive processes occur only intermittently"* [115, p. 29].

The increase in endogeneity in the disequilibrium states seems attributable to strategies employed by high-frequency traders (HFTs), who become active



(a) $\mathrm{Model_S}$

(b) $\mathrm{Model_{QI}}$

Figure 4.9: The estimated spectral radius $\hat{\rho}(x)$ as a function of $x \in \mathcal{X}$ under $\mathrm{Model_S}$ and $\mathrm{Model_{QI}}$. The daily profiles $x \mapsto \hat{\rho}(x)^{(i)}$, $i = 1, \ldots, 250$, are represented by the red translucent curves.

in these states, in anticipation of a price move. In a recent study, Lehalle and Neuman [82] analyse a unique data set from Nasdaq Stockholm, where the identities of the buyer and seller in each transaction were disclosed by the exchange until 2014. In particular, they show that when the QI increases, the trading activity of market participants they classify as proprietary HFTs is amplified, in the direction of the imbalance. The pronounced sub-millisecond self-excitation effects seen in Figure 4.4b, which are the key driver behind the high spectral radii $\rho(x)$, concur with the trading patterns observed by Lehalle and Neuman [82]. Besides its use as a trading signal, Lehalle and Neuman [82] find the QI to be mean-reverting, which is similarly compatible with our results. A higher spectral radius in disequilibrium states corresponds here to an immediate increase in market activity, which, reinforced by the structure of the estimated transition probabilities (Figure 4.3), is more likely to push the QI towards equilibrium than vice versa.

4.8 Event–state structure of LOBs

We could alternatively build a state-dependent variant of a Hawkes process in the following, conceptually simpler way. Using the representation $\tilde{\boldsymbol{N}}$ of the counting and state processes $(\boldsymbol{N}, X)$, one could specify an intensity of

Figure 4.10: In-sample (12:00–14:30) Q–Q plots of total residuals of $\mathrm{Model}_{\mathrm{QI}}$ (state-dependent) and the alternative model given by (4.2) (complex) on 11 May 2018. Each panel corresponds to a sequence of residuals $(\tilde{r}_n^{ex})$ for all $e \in \mathcal{E}$ and $x \in \mathcal{X}$.

the form

$$\tilde{\lambda}_{ex}(t) = \nu_{ex} + \sum_{e' \in \mathcal{E}, x' \in \mathcal{X}} \int_{[0,t)} k_{e'x'ex}(t-s) d\tilde{N}_{e'x'}(s), \quad t \geq 0, \quad e \in \mathcal{E}, \quad x \in \mathcal{X}, \tag{4.2}$$

instead of (3.3). This approach would in fact be tantamount to simply using a $d_e d_x$-dimensional ordinary Hawkes process.

The intensity (4.2) makes self- and cross-excitation state-dependent, but the simple structure of the state process X in Definition 3.1.1 is lost. Namely, under (4.2), the transition probabilities of the state process depend not only on the current state but on the entire history. However, LOBs enjoy a certain *event–state structure* — knowing the current state of the LOB and the characteristics of the next order suffices to (approximately) determine the next state. State-dependent Hawkes processes are by construction able to reproduce such an event-state structure and, therefore, compared to the alternative (4.2), we expect them to provide in general a better statistical description of the LOB. Moreover, the model given by (4.2) requires a kernel with $d_e^2 d_x^2$ components whereas a state-dependent Hawkes process can be specified more parsimoniously, using only $d_e^2 d_x$ components.

To compare this alternative model to state-dependent Hawkes processes, we estimate an ordinary Hawkes process with intensity (4.2), choosing an exponential form of the kernel $\boldsymbol{k}$ with event types in $\mathcal{E}$ and state space $\mathcal{X}$ as in Model_{QI}. Because of the higher dimensionality of this alternative model, estimation becomes computationally more expensive, which is why we use INTC data for May 2018 only. In Figure 4.10, this alternative model is compared to Model_{QI} via Q–Q plots of the total residuals. We choose to display the results for a day when the alternative model provides one of its best fits, although the goodness of fit does not vary markedly across the 22 trading days in May 2018. Even though it has fewer parameters (92 as opposed to 210), we observe that Model_{QI} provides a better in-sample fit. These empirical results thus underline the significance of the event–state structure of LOBs.

4.9 Discussion

State-dependent Hawkes processes enable us to model two-way interaction between a self- and cross-exciting point process, governing the temporal flow of *events*, and a state process, describing a *system*. In the context of LOB modelling, they provide a probabilistic foundation for a novel class of continuous-time models that encapsulate the feedback loop between the order flow and the shape of the LOB. Our estimation results, using these models and one year's worth of high-frequency LOB data on the stock of Intel Corporation, reveal that state dependence is indeed significant, as we uncover several robust patterns that persist throughout the daily estimation results. In particular, we find that market endogeneity, measured through the magnitude of self- and cross-excitation is state-dependent, being most pronounced in disequilibrium states of the LOB.

Our results also validate the event–state structure of LOBs that is embedded in the definition of a state-dependent Hawkes process. However, we do not claim that Model_{S} and Model_{QI} would be the best possible representations of the aforementioned feedback loop — we recognise that they could be refined as follows:

(i) The exponential excitation kernels could be replaced by power laws, motivated by the non-parametric estimation results on ordinary Hawkes processes [58, 9]. A numerically more convenient alternative would be to use a superposition of exponentials within the parametric framework to mimic the slow decay of a power law [107, 87].

(ii) The base rates could be made state-dependent by replacing ν_e with $\nu_e(X(t))$ in (3.1). Note that in this case, the model would contain both a continuous-time Markov chain ($\boldsymbol{k} \equiv 0$) and an ordinary Hawkes process as special cases.

(iii) One might argue for excitation kernels that allow for negative values, to capture inhibition effects that are known to exist in LOB data [87]. To ensure the non-negativity of the intensity, this would require transforming the right-hand side of (3.1) by a non-linear function.

(iv) More granular event types and states would provide more nuanced understanding of the LOB dynamics. Besides, notice that the present framework accommodates multiple state variables. For example, X could in fact jointly represent both the bid–ask spread and the QI using the state space

$$\mathcal{X} = \{(\mathsf{1},\ \mathsf{sell{+}{+}}), \dots, (\mathsf{1},\ \mathsf{buy{+}{+}}), (\mathsf{2{+}},\ \mathsf{sell{+}{+}}), \dots, (\{\mathsf{2{+}},\ \mathsf{buy{+}{+}})\}.$$

However, no matter how one modifies the intensity (3.1) by implementing any of (i)–(iv) to incorporate one's views on the feedback loop, the event–state structure of the model remains intact. The process still falls within the class of hybrid marked point processes that we will introduce in Chapter 6, whence the theoretical results of Chapter 3 (existence, uniqueness, separability of the likelihood function) still apply.

4.10 Details on maximum likelihood estimation

4.10.1 Numerical optimisation

Similarly to ordinary Hawkes processes [87], the maximum likelihood (ML) search (3.5) is broken down into d_e separate optimisation problems. For every $e \in \mathcal{E}$, we have an independent optimisation problem that involves only the parameters ν_e, $(\alpha_{e'xe})_{e' \in \mathcal{E}, x \in \mathcal{X}}$, $(\beta_{e'xe})_{e' \in \mathcal{E}, x \in \mathcal{X}}$, and which we solve using a conjugate gradient method.

More precisely, we call the `minimize` function in the `scipy.optimize` Python package and use the method `TNC`. Three random sets of parameters are generated and used as alternative initial guesses. An ordinary Hawkes process ($d_x = 1$) is estimated before the state-dependent one and its parameters are also used as an initial guess. Moreover, the estimate of the previous day is used as another initial guess and, thus, a total of five different initial guesses are employed. For Model_{QI}, around 50 iterations and 400 function evaluations are required for each day, event type $e \in \mathcal{E}$ and initial guess.

As this optimisation problem is non-convex, the above procedure might converge to a mere local maximum instead of the global one [87]. Nevertheless, in a Monte Carlo experiment, reported in the following subsection, this conjugate gradient method returns estimates that are consistently concentrated near the true parameter values (see also Subsection 3.5.1).

4.10.2 Uncertainty quantification

The uncertainty of ML estimates of the parameters of a state-dependent Hawkes process can be quantified using a parametric bootstrap procedure. The procedure entails simulating realisations of the state-dependent Hawkes process under the estimated parameter values using Algorithm 3.3.1 and then applying ML estimation again to each simulated realisation, producing a sample of estimates that approximates the distribution of the ML estimator. To exemplify the method, we apply it here to Model_{QI} estimates for INTC on 13 February 2018. The results are presented in Figure 4.11, and we observe

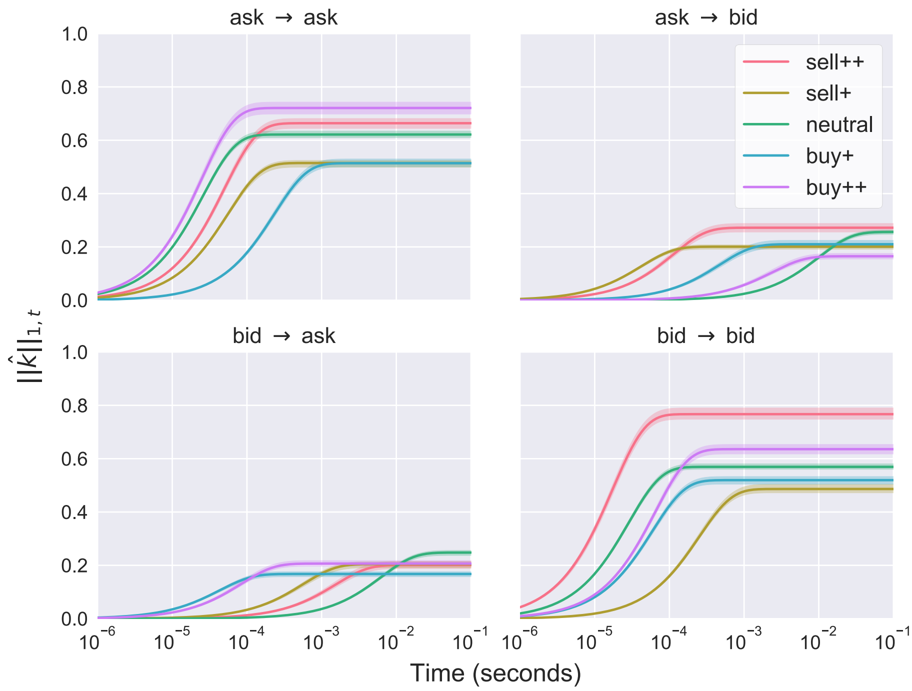


Figure 4.11: Uncertainty quantification of the estimated excitation profiles for INTC on 13 February 2018 under Model_{QI}. The parametric bootstrap procedure involves simulating 100 paths covering 2.5 hours of trading using the ML estimates of the parameters on the considered day and applying ML estimation again to each of the simulated paths. Three random sets of parameters are used as the initial guess in the optimisation procedure. We use the 100 estimates to compute a 99%-confidence interval for the truncated kernel norm (translucent area). The solid line corresponds to the ML estimates using the original INTC data.

that the uncertainty of the estimates is negligible compared to the estimated excitation patterns.

5

Framework for marked point processes

This chapter serves as a transition to the more theoretical part of this thesis by presenting a general framework for marked point processes. It sets the stage for the generalisation of state-dependent Hawkes processes to hybrid marked point processes (Chapter 6) and the resolution of the strong existence and uniqueness problem (Chapter 7).

In the following, $\mathcal{U}$ refers to a complete separable metric space and we denote by $\mathcal{B}(\mathcal{U})$ its Borel σ-algebra. We reserve the notation $\mathscr{M}$ for a complete separable metric space that represents the set of marks in the context of marked point processes. For most definitions, we follow closely Daley and Vere-Jones [41] along with Brémaud [19]. The former reference will be especially used to introduce (marked) point processes while the latter is essential when defining the intensity process.

Let ξ be a Borel measure on $\mathcal{U}$. We say that ξ is boundedly finite if $\xi(A) < \infty$ for every bounded Borel set $A \in \mathcal{B}(\mathcal{U})$. We denote by $\mathcal{N}_{\mathcal{U}}^{\infty}$ the space of Borel measures on $\mathcal{U}$ with values in $\mathbb{N} \cup \{\infty\}$. We denote by $\mathcal{N}_{\mathcal{U}}^{\#}$ the set of all $\xi \in \mathcal{N}_{\mathcal{U}}^{\infty}$ such that ξ is boundedly finite. We denote by $\mathcal{N}_{\mathbb{R}\times\mathscr{M}}^{\#g}$ the set of all $\xi \in \mathcal{N}_{\mathbb{R}\times\mathscr{M}}^{\#}$ such that their ground measure $\xi_g(\cdot) := \xi(\cdot \times \mathscr{M})$ satisfies:

(i) $\xi_g \in \mathcal{N}_{\mathbb{R}}^{\#}$;

(ii) $\xi_g(\{t\}) = 0$ or 1 for all $t \in \mathbb{R}$ (we say that the ground measure is simple).

Observe that $\mathcal{N}_{\mathbb{R}\times\mathscr{M}}^{\#g} \subset \mathcal{N}_{\mathbb{R}\times\mathscr{M}}^{\#} \subset \mathcal{N}_{\mathbb{R}\times\mathscr{M}}^{\infty}$. The space $\mathcal{N}_{\mathbb{R}\times\mathscr{M}}^{\infty}$ corresponds to the realisations of potentially explosive point processes, while the space $\mathcal{N}_{\mathbb{R}\times\mathscr{M}}^{\#}$ corresponds to the realisations of non-explosive point processes and contains all the realisations of potentially explosive marked point processes. Regarding the space $\mathcal{N}_{\mathbb{R}\times\mathscr{M}}^{\#g}$, each $\xi \in \mathcal{N}_{\mathbb{R}\times\mathscr{M}}^{\#g}$ is a realisation of a non-explosive marked point process. When $\xi(\{(t,m)\}) = 1$ for some $t \in \mathbb{R}$ and $m \in \mathscr{M}$, this should be interpreted as an event happening at time t with characteristics m. The boundedly finite property of the ground measure ensures that, in any finite amount of time, only finitely many events can occur (i.e., the marked point point process is non-explosive). The simpleness constraint on the ground measure means that there cannot be two events at the same time.

The so-called $w^{\#}$-distance $d^{\#}$ ("weak-hash") introduced by Daley and Vere-Jones [40, p. 403], makes $\mathcal{N}_{\mathcal{U}}^{\#}$ a complete separable metric space, see Theorem A2.6.III in Daley and Vere-Jones [40, p. 404]. The corresponding σ-algebra $\mathcal{B}(\mathcal{N}_{\mathcal{U}}^{\#})$ coincides with the one generated by all mappings

$$\xi \mapsto \xi(A), \quad \xi \in \mathcal{N}_{\mathcal{U}}^{\#}, A \in \mathcal{B}(\mathcal{U}).$$

Proposition A2.6.II of Daley and Vere-Jones [40, p. 403] characterises convergence in this topology, called the $w^{\#}$-topology. These properties of the space $\mathcal{N}_{\mathcal{U}}^{\#}$ play an important role in the subsequent chapters. Note that, in Chapter 8, we clarify the proofs of Proposition A2.6.II and Theorem A2.6.III of Daley and Vere-Jones [40]. Indeed, the original proofs assume a certain function to be monotonic. As this does not actually hold in general, we find alternative arguments where required. Besides, in Chapter 8, we show that $\mathcal{N}^{\#g}_{\mathbb{R}\times\mathscr{M}}$ is indeed a Borel set of $\mathcal{N}^{\#}_{\mathbb{R}\times\mathscr{M}}$, see Lemma 8.7.1.

Finally, for any $u \in \mathcal{U}$, we denote by δ_u the Dirac measure at u.

5.2 Non-explosive marked point processes

In the following, the notation $(\Omega, \mathcal{F}, \mathbb{P})$ refers to a probability space. For any σ-algebra $\mathcal{S}$, the trace of $A \in \mathcal{S}$ on $\mathcal{S}$ is defined by $A \cap \mathcal{S} := \{A \cap S \,:\, S \in \mathcal{S}\}$.

Definition 5.2.1 (Non-explosive point process)**.** A *non-explosive point process* on $\mathcal{U}$ is a measurable mapping from $(\Omega, \mathcal{F})$ into $(\mathcal{N}^{\#}_{\mathcal{U}}, \mathcal{B}(\mathcal{N}^{\#}_{\mathcal{U}}))$.

Definition 5.2.2 (Non-explosive marked point process)**.** A *non-explosive marked point process* N on $\mathbb{R} \times \mathscr{M}$ is a non-explosive point process N on $\mathbb{R} \times \mathscr{M}$ such that $N(\omega) \in \mathcal{N}^{\#g}_{\mathbb{R}\times\mathscr{M}}$ for all $\omega \in \Omega$.

Remark 5.2.3. By applying Lemma 1.6 in Kallenberg [74, p. 4], we obtain that $\mathcal{B}(\mathcal{N}^{\#g}_{\mathbb{R}\times\mathscr{M}}) = \mathcal{N}^{\#g}_{\mathbb{R}\times\mathscr{M}} \cap \mathcal{B}(\mathcal{N}^{\#}_{\mathbb{R}\times\mathscr{M}})$, where $\mathcal{N}^{\#g}_{\mathbb{R}\times\mathscr{M}}$ is also equipped with the $w^{\#}$-metric $d^{\#}$. This implies that Definition 5.2.2 is equivalent to saying that a non-explosive marked point process is a measurable mapping from $(\Omega, \mathcal{F})$ into $(\mathcal{N}^{\#g}_{\mathbb{R}\times\mathscr{M}}, \mathcal{B}(\mathcal{N}^{\#g}_{\mathbb{R}\times\mathscr{M}}))$.

Next, a non-explosive point process induces a probability measure on $\mathcal{N}^{\#}_{\mathcal{U}}$.

Definition 5.2.4 (Induced probability)**.** Let N be a non-explosive point process on $\mathcal{U}$. We define the *induced probability measure* $\mathcal{P}^N$ on the measurable space $(\mathcal{N}^{\#}_{\mathcal{U}}, \mathcal{B}(\mathcal{N}^{\#}_{\mathcal{U}}))$ through the relation

$$\mathcal{P}^N(A) := \mathbb{P}\left(N^{-1}(A)\right), \quad A \in \mathcal{B}(\mathcal{N}^{\#}_{\mathcal{U}}).$$

As we did in the previous chapters, it is common to define instead marked point processes on $\mathbb{R}_{\geq 0} \times \mathscr{M}$ as a sequence $(T_n, M_n)_{n\in\mathbb{N}}$ of random variables in $(0, \infty] \times \mathscr{M}$ such that $(T_n)_{n\in\mathbb{N}}$ is non-decreasing and $T_n < \infty$ implies

$T_n < T_{n+1}$ [68, 19]. We will call such a sequence an enumeration. Here, T_n is to be interpreted as the time when the n^{th} event occurs while M_n describes the characteristics of that event. Moreover, $T_n < \infty$ with $T_{n+1} = \infty$ means that there are no more events after time T_n. Note that this definition allows for explosion in the sense that $\lim_{n\to\infty} T_n < \infty$ is possible with positive probability. This is why we stress the non-explosive character of marked point processes in Definition 5.2.2. There is a one-to-one correspondence between non-explosive marked point processes on $\mathbb{R}_{\geq 0} \times \mathscr{M}$ and enumerations such that $\lim_{n\to\infty} T_n = \infty$ a.s. We will give a proof of this correspondence in Chapter 8 for completeness.

5.3 Pathwise integration

Let N be a non-explosive point process on $\mathcal{U}$. Let $H : \Omega \times \mathcal{U} \to \mathbb{R}_{\geq 0}$ be an $\mathcal{F} \otimes \mathcal{B}(\mathcal{U})$-measurable non-negative mapping. In particular, H is an $\mathbb{R}_{\geq 0}$-valued stochastic process on $\mathcal{U}$. One can define the integral of H against N in a pathwise fashion as

$$I(\omega) := \int_{\mathcal{U}} H(\omega, u) N(\omega, du), \quad \omega \in \Omega.$$

Besides, by a monotone class argument, one can check that $\omega \mapsto I(\omega)$ is $\mathcal{F}$-measurable. In the special case where N is actually a non-explosive marked point process on $\mathbb{R}_{\geq 0} \times \mathscr{M}$, the integral can be rewritten as

$$\iint_{\mathbb{R}_{\geq 0} \times \mathscr{M}} H(t, m) N(dt, dm) = \sum_{n \in \mathbb{N}} H(T_n, M_n) \mathbb{1}_{\{T_n < \infty\}}, \quad \text{a.s.},$$

where $(T_n, M_n)_{n\in\mathbb{N}}$ is the enumeration corresponding to N. For any realisation $\xi \in \mathcal{N}^{\#g}_{\mathbb{R}\times\mathscr{M}}$ and any $\tau \in \mathbb{R}$ such that $\xi(\{\tau\} \times \mathscr{M}) > 0$, we abuse the notation and define $\iint_{\{\tau\}\times\mathscr{M}} m\xi(dt, dm)$ as the unique element $m \in \mathscr{M}$ such that $\xi(\{\tau\} \times \{m\}) = 1$.

5.4 Shifts, restrictions, histories and predictability

For all $t \in \mathbb{R}$, define the shift operator $\theta_t : \mathcal{N}^{\#}_{\mathbb{R}\times\mathcal{U}} \to \mathcal{N}^{\#}_{\mathbb{R}\times\mathcal{U}}$ by

$$\theta_t\xi(A) := \xi(A+t), \quad A \in \mathcal{B}(\mathbb{R}\times\mathcal{U}),$$

where $A+t := \{(s+t,u) \in \mathbb{R}\times\mathcal{U} \,:\, (s,u) \in A\}$. Then, for any non-explosive point process N on $\mathbb{R}\times\mathcal{U}$, define $\theta_t N$ through $(\theta_t N)(\omega) := \theta_t(N(\omega))$, $\omega \in \Omega$. It will be useful to show that $\theta_t\xi$ is jointly continuous in t and ξ (Lemma 8.7.2).

Denote the restriction to the negative real line of any realisation $\xi \in \mathcal{N}^{\#}_{\mathbb{R}\times\mathcal{U}}$ by $\xi^{<0}$, which is defined by $\xi^{<0}(A) := \xi(A \cap \mathbb{R}_{<0} \times \mathcal{U})$, $A \in \mathcal{B}(\mathbb{R}\times\mathcal{U})$. We can then define the restriction to the negative real line of any non-explosive point process N on $\mathbb{R}\times\mathcal{U}$ by $N^{<0}(\omega) := (N(\omega))^{<0}$, $\omega \in \Omega$. Similarly, define the notations $\xi^{\leq 0}(A) := \xi(A \cap \mathbb{R}_{\leq 0} \times \mathcal{U})$, $N^{\leq 0}(\omega) := N(\omega)^{\leq 0}$,
$\xi^{\geq 0}(A) := \xi(A \cap \mathbb{R}_{\geq 0} \times \mathcal{U})$, $N^{\geq 0}(\omega) := N(\omega)^{\geq 0}$, $\xi^{>0}(A) := \xi(A \cap \mathbb{R}_{>0} \times \mathcal{U})$ and $N^{>0}(\omega) := N(\omega)^{>0}$.

These notations will allow us to refer to the internal history of N. For instance, for any $t \in \mathbb{R}$, $(\theta_t N)^{<0}$ contains the history of the process up to time t, excluding time t. To lighten these notations, we will use the conventions $\theta_t\xi^{<0} := (\theta_t\xi)^{<0}$, $\theta_t\xi^{\leq 0} := (\theta_t\xi)^{\leq 0}$, $\theta_t\xi^{>0} := (\theta_t\xi)^{>0}$ and $\theta_t\xi^{\geq 0} := (\theta_t\xi)^{\geq 0}$. It will be useful to note that these restriction mappings are measurable (Lemma 8.7.3) and that $\theta_t\xi^{<0}$ is left-continuous as a function of $t \in \mathbb{R}$ (Lemma 8.7.4).

Let N be a non-explosive point process on $\mathbb{R}\times\mathcal{U}$. We can define the filtration $\mathbb{F}^N = (\mathcal{F}^N_t)_{t\in\mathbb{R}}$ that corresponds to the internal history of N by

$$\mathcal{F}^N_t := \sigma\left\{N(A\times U) \,:\, A \in \mathcal{B}(\mathbb{R}),\, A \subset (-\infty,t],\, U \in \mathcal{B}(\mathcal{U})\right\} \quad \text{for all } t \in \mathbb{R}.$$

Using Lemma 1.4 in Kallenberg [74, p. 4] along with the characterisation of $\mathcal{B}(\mathcal{N}^{\#}_{\mathbb{R}\times\mathcal{U}})$ given in Theorem A2.6.III in Daley and Vere-Jones [40, p. 404], one can check that $\mathcal{F}^N_t = \sigma(\theta_t N^{\leq 0})$.

In the following, we call a *history* any filtration that contains the internal history of N, that is any filtration $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{R}}$ such that $\mathcal{F}^N_t \subset \mathcal{F}_t$, $t \in \mathbb{R}$.

Equivalently, one says that N is $\mathbb{F}$-adapted. The notation $\mathbb{F} = (\mathcal{F}_t)_{t \in \mathbb{R}}$ will always be used to refer to a history. We also need to define the predictable σ-algebra $\mathcal{F}^p$ on $\Omega \times \mathbb{R} \times \mathcal{U}$ corresponding to a history $\mathbb{F}$. The σ-algebra $\mathcal{F}^p$ is the one which is generated by all the sets of the form

$$A \times (s,t] \times U, \quad s,t \in \mathbb{R},\, s < t,\, U \in \mathcal{B}(\mathcal{U}),\, A \in \mathcal{F}_s.$$

Any mapping $H : \Omega \times \mathbb{R} \times \mathcal{U} \to \mathbb{R}$ that is $\mathcal{F}^p$-measurable is called an $\mathbb{F}$-predictable process. Any mapping H on the positive real line of the form $H : \Omega \times \mathbb{R}_{>0} \times \mathcal{U} \to \mathbb{R}$ that is $(\Omega \times \mathbb{R}_{>0} \times \mathcal{U}) \cap \mathcal{F}^p$-measurable is also called an $\mathbb{F}$-predictable process. Given an $\mathbb{F}$-stopping time τ, the strict past $\mathcal{F}_{\tau-}$ is defined as the σ-algebra generated by the all the classes $\{t < \tau\} \cap \mathcal{F}_t$, $t \in \mathbb{R}$.

5.5 Integration with respect to Poisson processes

Let ν be a boundedly finite measure on $(\mathcal{U}, \mathcal{B}(\mathcal{U}))$. We say that a non-explosive point process N on $\mathcal{U}$ is a Poisson process on $\mathcal{U}$ with parameter measure ν if $N(A_1), \ldots, N(A_n)$ are mutually independent for all disjoint and bounded sets $A_1, \ldots, A_n \in \mathcal{B}(\mathcal{U})$, $n \in \mathbb{N}$, and $N(A)$ follows a Poisson distribution with parameter $\nu(A)$ for all bounded sets $A \in \mathcal{B}(\mathcal{U})$. Their existence can be verified using Theorem 9.2.X in Daley and Vere-Jones [41, p. 30], see Example 9.2(b) on p. 31 therein.

We next clarify briefly the link between point processes and random measures. A random measure M on a measurable space $(S, \mathcal{S})$ is a mapping $M : \Omega \times \mathcal{S} \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ such that $M(\omega, \cdot)$ is a measure on $(S, \mathcal{S})$ for all $\omega \in \Omega$ and $M(\cdot, A)$ is a random variable for all $A \in \mathcal{S}$, see Kallenberg [74, p. 106] and Çinlar [30, Chapter 6, p. 243]. Note that the concepts of internal history and adaptedness of Subsection 5.4 can be directly extended to random measures. Not surprisingly, point processes are exactly the boundedly finite integer-valued random measures.

Proposition 5.5.1. *Let N be a random measure on $(\mathcal{U}, \mathcal{B}(\mathcal{U}))$ such that $N(\omega, \cdot) \in \mathcal{N}^{\#}_{\mathcal{U}}$ for all $\omega \in \Omega$. Then N is a non-explosive point process on $\mathcal{U}$. In return, any non-explosive point process N on $\mathcal{U}$ is a random measure on*

$(\mathcal{U}, \mathcal{B}(\mathcal{U}))$ *such that* $N(\omega, \cdot) \in \mathcal{N}_{\mathcal{U}}^{\#}$ *for all* $\omega \in \Omega$.

Proof. See Proposition 9.1.VIII in Daley and Vere-Jones [41, p. 8]. $\square$

One can show that Poisson processes are Poisson random measures in the sense of Çinlar [30, Chapter 6, p. 249]. This enables us to apply an important result on integration with respect to Poisson random measures. Before stating the result, we need to define what it means for a Poisson process to be Poisson relative to a filtration.

Definition 5.5.2. Let N be a Poisson process on $\mathbb{R} \times \mathcal{U}$ and $\mathbb{F} = (\mathcal{F}_t)_{t \in \mathbb{R}}$ be a filtration. We say that N is Poisson *relative* to $\mathbb{F}$ if for all $t \in \mathbb{R}$, the point process $\theta_t N^{\leq 0}$ is $\mathcal{F}_t$-measurable and $\sigma(\theta_t N^{>0})$ is independent of $\mathcal{F}_t$.

Trivially, a Poisson process N is always Poisson relative to its internal history $\mathbb{F}^N$. The next result plays a crucial role in the Poisson embedding technique, which is later used to construct marked point processes with given intensities.

Theorem 5.5.3. *Let* N *be a Poisson process on* $\mathbb{R} \times \mathcal{U}$ *with parameter measure* ν. *Let* $\mathbb{F} = (\mathcal{F}_t)_{t \in \mathbb{R}}$ *be a filtration and suppose that* N *is Poisson relative to* $\mathbb{F}$. *Then, for every non-negative* $\mathbb{F}$*-predictable process* H *of the form* $H : \Omega \times \mathbb{R} \times \mathcal{U} \to \mathbb{R}_{\geq 0}$, *we have that*

$$\mathbb{E}\left[\iint_{\mathbb{R} \times \mathcal{U}} H(t, u) N(dt, du)\right] = \mathbb{E}\left[\iint_{\mathbb{R} \times \mathcal{U}} H(t, u) \nu(dt, du)\right].$$

Proof. See Theorem 6.2 in Çinlar [30, Chapter 6, p. 299]. $\square$

5.6 Intensity process and functional

We equip the mark space $(\mathscr{M}, \mathcal{B}(\mathscr{M}))$ with a reference measure $\mu_{\mathscr{M}}$, allowing us to define the concept of intensity rigorously. Let N be a marked point process on $\mathbb{R} \times \mathscr{M}$ and $\mathbb{F} = (\mathcal{F}_t)_{t \in \mathbb{R}}$ a history. Let $\lambda : \Omega \times \mathbb{R}_{>0} \times \mathscr{M} \to \mathbb{R}_{\geq 0}$ be a non-negative $\mathbb{F}$-predictable process. We say that λ is the $\mathbb{F}$-intensity of

N relative to $\mu_{\mathscr{M}}$ if for every non-negative $\mathbb{F}$-predictable process of the form $H : \Omega \times \mathbb{R}_{>0} \times \mathscr{M} \to \mathbb{R}_{\geq 0}$,

$$\mathbb{E}\left[\iint_{\mathbb{R}_{>0}\times\mathscr{M}} H(t,m)N(dt,dm)\right] = \mathbb{E}\left[\iint_{\mathbb{R}_{>0}\times\mathscr{M}} H(t,m)\lambda(t,m)\mu_{\mathscr{M}}(dm)dt\right]. \tag{5.1}$$

Note that if an intensity exists, it is then unique up to $\mathbb{P}(d\omega)\mu_{\mathscr{M}}(dm)dt$-null sets thanks to the predictability requirement, see Brémaud [19, Section II.4] and Daley and Vere-Jones [41, p. 391]. In the following chapters, we will be particularly interested in intensities that are expressed in terms of a functional applied to the point process.

Definition 5.6.1 (Intensity functional). Let $\psi : \mathscr{M} \times \mathcal{N}^{\#}_{\mathbb{R}\times\mathscr{M}} \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ be a measurable functional. We say that a non-explosive marked point process $N : \Omega \to \mathcal{N}^{\#g}_{\mathbb{R}\times\mathscr{M}}$ admits ψ as its *intensity functional* if N admits an $\mathbb{F}^N$-intensity $\lambda : \Omega \times \mathbb{R}_{>0} \times \mathscr{M} \to \mathbb{R}_{\geq 0}$ relative to $\mu_{\mathscr{M}}$ such that

$$\lambda(\omega,t,m) = \psi(m \,|\, \theta_t N(\omega)^{<0}), \quad \mathbb{P}(d\omega)dt\mu_{\mathscr{M}}(dm)\text{-a.e.} \tag{5.2}$$

5.7 Initial condition

Let $(\Omega_{\leq 0}, \mathcal{F}_{\leq 0}, \mathbb{P}_{\leq 0})$ be a given probability space and $N_{\leq 0}$ be a given marked point process on $\mathbb{R} \times \mathscr{M}$ such that $N_{\leq 0}(\omega_{\leq 0})^{\leq 0} = N_{\leq 0}(\omega_{\leq 0})$ for all $\omega_{\leq 0} \in \Omega_{\leq 0}$ (i.e., there are no events on $\mathbb{R}_{>0}$). We will reserve the notation $N_{\leq 0}$ to refer to an initial condition.

Let $(\Omega_{>0}, \mathcal{F}_{>0}, \mathbb{P}_{>0})$ be another probability space that will correspond to the driving Poisson process in the SDE of Chapter 7. In the context of strong existence, we will work with the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ which we define as the completion [74, p. 13] of the product probability space given by

$$\Omega := \Omega_{\leq 0} \times \Omega_{>0}, \quad \tilde{\mathcal{F}} := \mathcal{F}_{\leq 0} \otimes \mathcal{F}_{>0}, \quad \tilde{\mathbb{P}} := \mathbb{P}_{\leq 0} \times \mathbb{P}_{>0}. \tag{5.3}$$

Such a structure of the probability space is motivated by the fact that the driving noise and the initial condition are independent.

Definition 5.7.1 (Strong initial condition)**.** Let $N : \Omega \to \mathcal{N}^{\#g}_{\mathbb{R}\times\mathscr{M}}$ be a non-explosive marked point process on $\mathbb{R} \times \mathscr{M}$. We say that N *satisfies a strong initial condition* $N_{\leq 0}$ if $N(\omega)^{\leq 0} = N_{\leq 0}(\omega_{\leq 0})$ a.s., where $\omega = (\omega_{\leq 0}, \omega_{>0}) \in \Omega$.

In the context of weak uniqueness, one needs a another concept of initial condition. Let $(\Omega', \mathcal{F}', \mathbb{P}')$ be another probability space potentially different from $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 5.7.2 (Weak initial condition)**.** Let $N' : \Omega' \to \mathcal{N}^{\#g}_{\mathbb{R}\times\mathscr{M}}$ be a non-explosive marked point process on $\mathbb{R} \times \mathscr{M}$. We say that N' *satisfies a weak initial condition* $N_{\leq 0}$ if the induced probability $\mathcal{P}^{N'^{\leq 0}}$ coincides with $\mathcal{P}^{N_{\leq 0}}$.

6

Generalisation to hybrid marked point processes

We generalise state-dependent Hawkes processes to a class of hybrid marked point processes that encompasses and extends continuous-time Markov chains and Hawkes processes. To do so, we view the point process and the state process as a single marked point process on a product mark space and allow the intensity of events to be any measurable functional of past events and states. Under this alternative viewpoint, these new hybrid marked point processes are actually defined implicitly via their intensity that takes a specific product form. The main result of this chapter is that the dynamics generated by this product form completely characterise the class of hybrid marked point processes (Theorem 6.2.5).

Offering an *event–state* viewpoint, this new class is well-suited to the joint modelling of events and the time evolution of the state of a system. As far as the modelling of LOBs is concerned, this class provides a framework that unifies the existing models that are based on Hawkes processes and continuous-time Markov chains. More importantly, it also contains new processes that bridge these two prominent modelling approaches, state-dependent Hawkes processes being just one example, and thus, resolves the HM dichotomy.

6.1 Mark space and state process

Let $(\mathcal{E}, \mathcal{B}(\mathcal{E}), \mu_{\mathcal{E}})$ and $(\mathcal{X}, \mathcal{B}(\mathcal{X}), \mu_{\mathcal{X}})$ be two measure spaces where both $\mathcal{E}$ and $\mathcal{X}$ are complete separable metric spaces and both $\mu_{\mathcal{E}}$ and $\mu_{\mathcal{X}}$ are a boundedly finite Borel measures. Each $e \in \mathcal{E}$ represents a type of event and we call $\mathcal{E}$ the event space. Each $x \in \mathcal{X}$ represents a possible state of a system and we call $\mathcal{X}$ the state space. Motivated by the need to jointly model events and the state of the system (HM dichotomy), we consider a mark space $(\mathscr{M}, \mathcal{B}(\mathscr{M}), \mu_{\mathscr{M}})$ of the form

$$\mathscr{M} := \mathcal{E} \times \mathcal{X}, \quad \mathcal{B}(\mathscr{M}) = \mathcal{B}(\mathcal{E}) \otimes \mathcal{B}(\mathcal{X}), \quad \mu_{\mathscr{M}} := \mu_{\mathcal{E}} \times \mu_{\mathcal{X}}. \tag{6.1}$$

Such a decomposition of the mark space admits the following interpretation. Let $\xi \in \mathcal{N}^{\#g}_{\mathbb{R}\times\mathscr{M}}$ be a realisation of a marked point process on $\mathbb{R} \times \mathscr{M}$. A point $t \in \mathbb{R}$ and a point $m = (e, x) \in \mathscr{M}$ such that $\xi(\{t, m\}) = 1$ can now be interpreted as an event of type e occurring at time t and moving the state of the system to x. To formalise this viewpoint, we define the state functional and the state process as follows.

Definition 6.1.1 (State functional and state process)**.** We define the measurable state functional
$F : \mathcal{N}^{\#}_{\mathbb{R}\times\mathscr{M}} \to \mathcal{X}$ by

$$F(\xi) := \begin{cases} \iint_{\{\kappa(\xi)\}\times\mathscr{M}} x\xi(dt, de, dx), & \text{if } \xi \in \mathcal{N}^{\#g}_{\mathbb{R}\times\mathscr{M}} \text{ and } \kappa(\xi) > -\infty, \\ x_0, & \text{otherwise,} \end{cases}$$

where $\kappa(\xi) := \inf\{t < 0 : \xi((t, 0) \times \mathscr{M}) = 0\}$ and $x_0 \in \mathcal{X}$ is an arbitrary initial state. Given a non-explosive marked point process N on $\mathscr{M}$, we define the state process $(X_t)_{t\in\mathbb{R}}$ by

$$X_t := F(\theta_t N^{<0}), \quad t \in \mathbb{R}.$$

Note that $\kappa(\theta_t N^{<0})$ is the time of the last event up to time t and, thus, X_t is the coordinate $x \in \mathcal{X}$ of the mark $m = (e, x) \in \mathscr{M}$ of the most recent event. As a consequence, we indeed have that a point $t \in \mathbb{R}$ and a point $m = (e, x) \in \mathscr{M}$ such that $N(\{t, m\}) = 1$ can be interpreted as an event of type e occurring at time t and moving the state of the system to x. With this viewpoint, a marked point process on $\mathbb{R} \times \mathscr{M}$ allows to jointly model the evolution of a system with state process $(X_t)_{t \in \mathbb{R}}$ in $\mathcal{X}$ and the arrival in time of the event types $\mathcal{E}$. To check that the state functional F is indeed measurable, one can adapt the proof of Lemma 8.7.6 in Chapter 8.

6.2 Definition, implied dynamics and characterisation

We can now introduce the class of hybrid marked point processes, which provides a unified treatment of various types of processes, including continuous-time Markov chains and Hawkes processes, whence the qualifier *hybrid*. Moreover, this class contains new types of processes, such as state-dependent Hawkes processes, which are applicable to the joint modelling of events and systems. This class is specified implicitly through a specific form of the intensity. We were inspired by Cartea et al. [23] who, in the context of a continuous-time Markov chain model, propose a decomposition of the intensity that is similar in spirit. We should also mention the connection to the decomposition of the rate kernel α of a continuous-time Markov chain into a rate function c and transition kernel μ (i.e., $\alpha = \mu c$), see Kallenberg [74, p. 238-239], even though this is slightly different as c is the total intensity and does not depend on the event variable $e \in \mathcal{E}$.

Definition 6.2.1 (Hybrid marked point processes)**.** Let $\phi : \mathcal{X} \times \mathcal{E} \times \mathcal{X} \to \mathbb{R}_{\geq 0}$ be a measurable non-negative function such that $\phi(\cdot \,|\, e, x)$ is a probability density over $(\mathcal{X}, \mathcal{B}(\mathcal{X}), \mu_{\mathcal{X}})$ for all $e \in \mathcal{E}$, $x \in \mathcal{X}$. Furthermore, let $\eta : \mathcal{E} \times \mathcal{N}^{\#}_{\mathbb{R} \times \mathscr{M}} \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ be a measurable non-negative functional. Define the measurable intensity functional $\psi : \mathscr{M} \times \mathcal{N}^{\#}_{\mathbb{R} \times \mathscr{M}} \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ by $\psi(m \,|\, \xi) := \phi(x \,|\, e, F(\xi))\eta(e \,|\, \xi)$ for all $m = (e, x) \in \mathscr{M}$, $\xi \in \mathcal{N}^{\#}_{\mathbb{R} \times \mathscr{M}}$. A *hybrid marked point process* with transition function ϕ and event functional η is a non-explosive marked point process $N : \Omega \to \mathcal{N}^{\#g}_{\mathbb{R} \times \mathscr{M}}$ that admits

ψ as its intensity functional. In other words, N admits an $\mathbb{F}^N$-intensity $\lambda : \Omega \times \mathbb{R}_{>0} \times \mathscr{M} \to \mathbb{R}_{\geq 0}$ relative to $\mu_{\mathscr{M}}$ that satisfies

$$\lambda(\omega, t, e, x) = \phi(x \,|\, e, X_t(\omega))\eta(e \,|\, \theta_t N(\omega)^{<0}), \quad \mathbb{P}(d\omega)dt\mu_{\mathscr{M}}(de, dx)\text{-a.e.} \tag{6.2}$$

To demonstrate the generality and flexibility of hybrid marked point processes, we give three examples of well-known processes that belong to the class. While at first these examples might be understood only at an intuitive level, the reader should become fully convinced of their validity once Theorem 6.2.5 is introduced.

Example 6.2.2 (Compound Poisson process)**.** Let $\mathcal{E} = \{0\}$ (i.e., just one type of event), $\mu_{\mathcal{E}} = \delta_0$, $\mathcal{X} = \mathbb{R}$, $\mu_{\mathcal{X}}(dx) = dx$ (i.e., the Lebesgue measure) and suppose that $f : \mathbb{R} \to \mathbb{R}_{\geq 0}$ is a probability density function. Consider a hybrid marked point process N with constant event functional $\eta \equiv \nu \in \mathbb{R}_{>0}$ and transition function given by $\phi(x' \,|\, x) = f(x' - x)$, $x', x \in \mathbb{R}$. Then, the $\mathbb{F}^N$-intensity of N satisfies $\lambda(t, x) = f(x - X_t)\nu$ and the state process $(X_t)_{t \in \mathbb{R}_{\geq 0}}$ is a compound Poisson process with rate ν and jump size distribution $f(x)dx$.

Example 6.2.3 (Continuous-time Markov chain)**.** Let $\mathcal{E}$, $\mu_{\mathcal{E}}$, $\mathcal{X}$ and $\mu_{\mathcal{X}}$ be as in Example 6.2.2. Suppose that a hybrid marked point process N has event functional of the form $\eta(\xi) = c(F(\xi))$, $\xi \in \mathcal{N}^{\#}_{\mathbb{R}^2}$, where c is a positive function. The $\mathbb{F}^N$-intensity of N is then given by $\lambda(t, x) = \phi(x \,|\, X_t)c(X_t)$ and the state process $(X_t)_{t \in \mathbb{R}_{\geq 0}}$ is a continuous-time Markov chain with rate function c and transition kernel $\mu(x, B) = \int_B \phi(y \,|\, x)dy$, $x \in \mathbb{R}$, $B \in \mathcal{B}(\mathbb{R})$.

Example 6.2.4 (Multivariate Hawkes process)**.** Let $\mathcal{E} = \{1, \ldots, d\}$, $d \in \mathbb{N}$, $\mu_{\mathcal{E}} = \sum_{n=1}^{d} \delta_n$, $\mathcal{X} = \{0\}$ (i.e., only one possible state), $\mu_{\mathcal{X}} = \delta_0$, $\nu = (\nu_1, \ldots, \nu_d) \in \mathbb{R}^d_{>0}$ and $k : \mathbb{R}_{>0} \times \mathcal{E}^2 \to \mathbb{R}_{\geq 0}$. Consider a hybrid marked point process N with event functional

$$\eta(e \,|\, \xi) = \nu_e + \iint_{(-\infty,0)\times\mathcal{E}} k(-t', e', e)\xi(dt', de'), \quad e = 1, \ldots, d,\, \xi \in \mathcal{N}^{\#}_{\mathbb{R}\times\mathcal{E}},$$

and note that the transition function must satisfy $\phi \equiv 1$. Then N is a multivariate Hawkes process with base rate ν and kernel k.

Let us next explain the dynamics that the intensity (6.2) implies. If one integrates out the state variable x by using the fact that $\phi(\cdot \,|\, e, X_t)$ is a probability density, one can see that the $\mathbb{F}^N$-intensity of the marked point process $N(\cdot \times \mathcal{X})$ on $\mathbb{R} \times \mathcal{E}$ is exactly $\eta(e \,|\, \theta_t N^{<0})$. In other words, $\eta(e \,|\, \theta_t N^{<0})$ is the intensity of the aggregation of events of type e (irrespectively of how they impact the state process X_t). Then, $\eta(e \,|\, \theta_t N^{<0})$ is distributed in the state space $\mathcal{X}$ according to $\phi(x \,|\, e, X_t)$, specifying the intensity of events with mark (e, x). This suggests that $\phi(\cdot \,|\, e, X_t)$ is the probability density of the next state of the system given that the next event is of type e and that the current state is X_t. This intuition is confirmed by Theorem 6.2.5, which actually goes further and states that these dynamics characterise hybrid marked point processes. The proof is presented in Subsection 6.4.2.

Theorem 6.2.5 (Implied dynamics and characterisation)**.** *Let ϕ and η be as in Definition 6.2.1. Moreover, suppose that N is a non-explosive marked point process on $\mathbb{R} \times \mathscr{M}$ with an $\mathbb{F}^N$-intensity relative to $\mu_{\mathscr{M}}$. Then, N is a hybrid marked point process with transition function ϕ and event functional η if and only if the following two statements hold.*

(i) *$N_{\mathcal{E}}(\cdot) := N(\cdot \times \mathcal{X})$ is a non-explosive marked point process on $\mathbb{R} \times \mathcal{E}$ that admits an $\mathbb{F}^N$-intensity $\lambda_{\mathcal{E}} : \Omega \times \mathbb{R}_{>0} \times \mathcal{E} \to \mathbb{R}_{\geq 0}$ relative to $\mu_{\mathcal{E}}$ such that $\lambda_{\mathcal{E}}(\omega, t, e) = \eta(e \,|\, \theta_t N(\omega)^{<0})$ holds $\mathbb{P}(d\omega)dt\mu_{\mathcal{E}}(de)$-a.e.*

(ii) *Let $t \in \mathbb{R}_{\geq 0}$ and define the stopping time*

$$\tau_t := \sup\{u > t \,:\, N((t, u) \times \mathscr{M}) = 0\}$$

and the random elements

$$(E, X) := \iint_{\{\tau_t\} \times \mathscr{M}} (e, x) N(du, de, dx)$$

such that τ_t is the time of the first event after time t and (E, X) is the corresponding mark. We have that

$$\mathbb{P}\left(X \in dx \,|\, \sigma(E) \vee \mathcal{F}^N_{\tau_t -}\right) \mathbb{1}_{\{\tau_t < \infty\}} = \phi(x \,|\, E, X_t) \mu_{\mathcal{X}}(dx) \mathbb{1}_{\{\tau_t < \infty\}}, \quad a.s. \tag{6.3}$$

Remark 6.2.6. As shown in the proof of Theorem 6.2.5, Equation (6.3) implies that

$$\mathbb{P}\left(X \in dx \,|\, \sigma(E) \vee \mathcal{F}_t^N, \{\tau_t < \infty\}\right) = \phi(x \,|\, E, X_t)\mu_{\mathcal{X}}(dx) \quad \text{a.s.}$$

We now add a fourth example to show that Definition 6.2.1 contains also new types of processes. This example is nothing else than the reformulation of state-dependent Hawkes processes as hybrid marked point processes. Together, the four examples demonstrate that hybrid marked point processes provide a common framework to construct and analyse various types of processes.

Example 6.2.7 (State-dependent Hawkes process)**.** Consider hybrid marked point processes with event functionals η of the form

$$\eta(e \,|\, \xi) = \nu(e) + \iint_{(-\infty,0)\times\mathscr{M}} k(-t', m', e)\xi(dt', dm'), \quad e \in \mathcal{E}, \xi \in \mathcal{N}^{\#}_{\mathbb{R}\times\mathscr{M}}, \tag{6.4}$$

where $\nu : \mathcal{E} \to \mathbb{R}_{\geq 0}$ and $k : \mathbb{R} \times \mathscr{M} \times \mathcal{E} \to \mathbb{R}_{\geq 0}$ are non-negative measurable functions. We will show later that such functionals are indeed measurable (Proposition 8.7.8). By Theorem 6.2.5, this gives rise to a marked point process $N_{\mathcal{E}}$ with marks in $\mathcal{E}$ and intensity η that interacts with a state process $(X_t)_{t\in\mathbb{R}}$ on $\mathcal{X}$ with transition probabilities ϕ. On the one hand, events in $N_{\mathcal{E}}$ occur like in a Hawkes process except that now the kernel depends also on the state process. For example, an event of type $e' \in \mathcal{E}$ might precipitate an event of type $e \in \mathcal{E}$ only if it moves the system to some specific state $x_0 \in \mathcal{X}$, i.e., $k(\cdot, e', x', e) \equiv 0$ as soon as $x' \neq x_0$. On the other hand, the occurence of an event in $N_{\mathcal{E}}$ prompts a state change according to the transition probabilities ϕ. Consequently, such a marked point process defines a state-dependent Hawkes process where the state process is fully coupled with the Hawkes process. Viewing $N_{\mathcal{E}}$ and $(X_t)_{t\in\mathbb{R}}$ as one single marked point process N on $\mathcal{E} \times \mathcal{X}$ with intensity $\phi\eta$ will allow us to prove the existence of such dynamics in the next chapter, see Corollary 7.4.2 and Example 7.4.3.

This subclass of hybrid marked point processes extends the regime-switching model of Vinkovskaya [121], where the state process triggering the regime switches is not modelled. Besides, since here the events drive the dynamics of the state process, this subclass is different from the Markov-modulated Hawkes processes considered by Cohen and Elliott [32] or Swishchuk [116], where the state process is a continuous-time Markov chain that jumps independently of the events. Moreover, the intensity in Cohen and Elliott [32] depends only on the current state whereas, here and in Swishchuk [116], it may depend on all past states.

Remark 6.2.8. The term *hybrid* also alludes to a connection with stochastic hybrid systems [24, 123]. These systems evolve continuously in time typically according to differential equations, switching between different regimes as new discrete events occur. The class of hybrid marked point processes introduced in this chapter could contribute to the modelling of such systems by considering more general state functionals F. For example, F could generate a state process $(X_t)_{t \in \mathbb{R}}$ such that, after each event (T_n, E_n, X_n), it follows a differential equation with initial condition X_n and parameters $\theta(E_n)$ that depend on the event type E_n.

6.3 A unifying framework for LOB modelling

Hybrid marked point processes provide us with a unifying framework that contains not only all the existing LOB models discussed in Section 1.6 (Hawkes processes) and Section 1.7 (continuous-time Markov chains) but also new models like state-dependent Hawkes processes that preserve the *event–state* structure of LOBs (cf. Sections 4.8 and 4.9). In fact, the present framework can be used as a *flexible and intuitive language* that simplifies the specification and comparison of a wide universe of models that are defined by four components:

- an event space $(\mathcal{E}, \mathcal{B}(\mathcal{E}), \mu_{\mathcal{E}})$ that classifies the event types in the order flow;
- a state space $(\mathcal{X}, \mathcal{B}(\mathcal{X}), \mu_{\mathcal{X}})$ that corresponds to the LOB state variables described by the model;

- an event functional η that determines how the arrival rates of events are linked to the history of the order flow and the LOB;
- a transition function ϕ that describes how order flow events impact the LOB state variables.

As an illustration, we reformulate the level-I reduced-form model of Cont and De Larrard [34] as a hybrid marked point process.

Example 6.3.1 (Reduced-form zero-intelligence model in [34])**.** The events in the level-I order flow are classified according to six different types:

$$\begin{aligned}\mathcal{E} =&\{\text{buy market order, buy limit order, buy cancellation,}\\ &\text{sell market order, sell limit order, sell cancellation}\}.\end{aligned}$$

The state process takes values in $\mathcal{X} = \mathbb{N}^3$, where the first component corresponds to the bid price, while the second and third component correspond to $D_t(B_t)$ and $D_t(A_t)$ (level-I depth profile), respectively:

$$X_t = (B_t/\rho, D_t(B_t), D_t(A_t)).$$

It is assumed that the spread always equals one tick so the ask price does not need to be modelled. The event and state spaces are here equipped with the canonical counting measures. The event functional does not depend on the history and is thus of the simple form:

$$\eta(e \mid \xi) = \nu(e), \quad e \in \mathcal{E}, \xi \in \mathcal{N}^{\#}_{\mathbb{R}\times\mathscr{M}}.$$

Finally, as it is assumed that all orders are of unit size, the LOB mechanism is embedded in the transition function ϕ as follows:

$$\begin{aligned}&\phi((n_1', n_2', n_3') \mid e, (n_1, n_2, n_3)) =\\ &\qquad \mathbb{1}(e = \text{buy limit}, n_1' = n_1, n_2' = n_2 + 1, n_3' = n_3)\\ &+ \qquad \mathbb{1}(e = \text{sell limit}, n_1' = n_1, n_2' = n_2, n_3' = n_3 + 1)\\ &+ \qquad \mathbb{1}(e \in \{\text{buy market, sell cancellation}\})\times\\ &\qquad [\mathbb{1}(n_3 > 1, n_1' = n_1, n_2' = n_2, n_3' = n_3 - 1)\end{aligned}$$

$$
\begin{aligned}
&\qquad + \mathbb{1}(n_3 = 1, n_1' = n_1 + 1) f(n_2', n_3')] \\
+ \quad & \mathbb{1}(e \in \{\text{sell market, buy cancellation}\}) \times \\
& [\mathbb{1}(n_2 > 1, n_1' = n_1, n_2' = n_2 - 1, n_3' = n_3) \\
&\qquad + \mathbb{1}(n_2 = 1, n_1' = \max(n_1 - 1, 0)) \tilde{f}(n_2', n_3')],
\end{aligned}
$$

where $e \in \mathcal{E}$, (n_1, n_2, n_3), $(n_1', n_2', n_3') \in \mathcal{X}$ and f and $\tilde{f}$ are the distributions of the new depth profile when the price jumps upwards or downwards, respectively.

We would also like to point out that the simulation algorithm and estimation framework introduced in Chapter 3 for state-dependent Hawkes processes naturally extends to hybrid marked point processes. The separability property of the likelihood function still holds and implies that the transition function ϕ and event functional η can be estimated independently of one another. The goodness-of-fit diagnostics using residuals would also still apply, providing a common framework to assess and compare different models. For instance, the goodness-of-fit of continuous-time Markov chains could be assessed through an analysis of residuals, something that has never been done so far in the context of zero-intelligence and queue-reactive models and that would allow for a direct comparison with Hawkes-process models.

While queue-reactive models specify the arrival rates as functions of the current shape of the LOB, we have with state-dependent Hawkes processes placed the state dependence in the excitation kernel, making the arrival rates depend on the history of the state process. Hybrid marked point processes provide us with new models where different kinds of state dependence can now be combined, as illustrated by this straightforward extension of state-dependent Hawkes processes already mentioned in Section 4.9.

Example 6.3.2 (State-dependent Hawkes process – *bis*)**.** In Example 6.2.7, change the event functional to

$$
\eta(e \,|\, \xi) = \nu(e, F(\xi)) + \iint_{(-\infty,0)\times\mathscr{M}} k(-t', m', e)\xi(dt', dm'), \quad e \in \mathcal{E}, \xi \in \mathcal{N}^{\#}_{\mathbb{R}\times\mathscr{M}},
$$

where now $\nu : \mathcal{E} \times \mathcal{X} \to \mathbb{R}_{\geq 0}$ is a non-negative measurable function.

Notice that this slight extension of state-dependent Hawkes processes now contains both Hawkes-process models and queue-reactive models of the LOB. Indeed, if $k \equiv 0$, the arrival rate of events is given by $\eta(e \mid \theta_t N^{<0}) = \nu(e, X_t)$. Hence, by estimating this model from market data, one could measure which of ν and k is the *most state-dependent*, shedding light on the true nature of the state dependence.

Finally, we have stressed several times that a key novelty of state-dependent Hawkes processes is that the state process is fully coupled to the point process, which distinguishes them from Markov-modulated Hawkes processes. However, we would like to clarify that this latter group of processes can also be formulated as hybrid marked point processes, which shows the flexibility of this new class.

Example 6.3.3 (Markov-modulated univariate Hawkes process)**.**
Set $\mathcal{X} = \mathcal{E} = \{0, 1\}$, $\mu_{\mathcal{X}} = \mu_{\mathcal{E}} = \delta_0 + \delta_1$ and let η and ϕ be of the form

$$\begin{aligned}
\eta(0 \mid \xi) &= \gamma(F(\xi)), \quad \xi \in \mathcal{N}^{\#}_{\mathbb{R}\times\mathscr{M}} \\
\eta(1 \mid \xi) &= \nu + \int_{(-\infty,0)} k(-t', F(\xi))\xi(dt', \{1\}, \mathcal{X}), \quad \xi \in \mathcal{N}^{\#}_{\mathbb{R}\times\mathscr{M}}, \\
\phi(x' \mid 1, x) &= \mathbb{1}(x' = x), \quad x', x \in \mathcal{X},
\end{aligned}$$

where $\nu \in \mathbb{R}_{>0}$ and $\gamma : \mathcal{X} \to \mathbb{R}_{>0}$ and $k : \mathbb{R} \times \mathcal{X} \to \mathbb{R}_{\geq 0}$ are measurable. Then, the state process can only jump at events of type 0 and the intensity at time t of these events is simply given by $\gamma(X_t)$. Consequently, $(X_t)_{t\in\mathbb{R}}$ is a continuous-time Markov chain and the component $N'(\cdot) := N(\cdot, \{1\}, \mathcal{X})$ of the hybrid marked point process N is a Markov-modulated univariate Hawkes process, with an intensity at time t of the form

$$\lambda(t) = \nu + \int_{(-\infty,t)} k(t - s, X_t)N'(ds).$$

Note that if ϕ is relaxed to a more general form, the state process can also jump when events of type 1 occur. In this case, N' is modulated by a state process that is partially exogenous, and is somehow an hybrid of a Markov-modulated Hawkes process and state-dependent Hawkes process.

6.4 Proofs

In this section, we prove Theorem 6.2.5, which characterises the dynamics of hybrid marked point processes.

6.4.1 Preliminaries

We first present a lemma that helps us reuse some results in the literature that require a specific form for the filtration. It simply says that the information up to time u is equal to the information up to time t to which we add the information between time t and u, where $t < u$.

Lemma 6.4.1. Let N be a non-explosive point process on $\mathbb{R}\times\mathcal{U}$. Let $t, u \in \mathbb{R}$ such that $u > t$. Then, we have that $\mathcal{F}_u^N = \mathcal{F}_t^N \vee \mathcal{F}_u^{\theta_{-t}[\theta_t N^{>0}]}$.

Proof. Note that

$$\mathcal{F}_u^{\theta_t N^{>0}} = \sigma\left\{N(A \times U) \,:\, A \in \mathcal{B}(\mathbb{R}),\, A \subset (t,u],\, U \in \mathcal{B}(\mathcal{U})\right\} \quad \text{for all } u > t.$$

Then, clearly $\mathcal{F}_u^{\theta_{-t}[\theta_t N^{>0}]} \subset \mathcal{F}_u^N$. Also, $\mathcal{F}_t^N \subset \mathcal{F}_u^N$ and, thus $\mathcal{F}_t^N \vee \mathcal{F}_u^{\theta_{-t}[\theta_t N^{>0}]} \subset \mathcal{F}_u^N$. On the other hand, let $A \in \mathcal{B}(\mathbb{R})$ be such that $A \subset (-\infty, u]$ and let $U \in \mathcal{B}(\mathcal{U})$. We have that

$$N(A \times U) = N(A \cap (-\infty, t]) \times U) + N(A \cap (t,u] \times U).$$

The first term is $\mathcal{F}_t^N$ measurable while the second term is $\mathcal{F}_u^{\theta_{-t}[\theta_t N^{>0}]}$ measurable. Hence, $N(A \times U)$ is $\mathcal{F}_t^N \vee \mathcal{F}_u^{\theta_{-t}[\theta_t N^{>0}]}$ measurable. Since, by definition, $\mathcal{F}_u^N$ is the smallest σ-algebra that makes all the $N(A \times U)$ measurable, this implies that $\mathcal{F}_u^N \subset \mathcal{F}_t^N \vee \mathcal{F}_u^{\theta_{-t}[\theta_t N^{>0}]}$, which concludes the proof. $\square$

As defined in Subsection 5.6, an intensity process has to always be finite. We verify that, if one finds a potentially infinite process that satisfies the definition of the intensity, then one can take a finite version of this process and identify it with the intensity.

Lemma 6.4.2. Let N be a non-explosive marked point process on $\mathbb{R} \times \mathscr{M}$ and let $\tilde{\lambda} : \Omega \times \mathbb{R}_{>0} \times \mathscr{M} \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ be an $\mathbb{F}$-predictable process that satisfies (5.1) for all non-negative $\mathbb{F}$-predictable processes H. Then N admits an $\mathbb{F}$-intensity $\lambda : \Omega \times \mathbb{R}_{>0} \times \mathscr{M} \mapsto \mathbb{R}_{\geq 0}$ relative to $\mu_{\mathscr{M}}$ such that $\lambda(\omega, t, m) = \tilde{\lambda}(\omega, t, m)$ holds $\mathbb{P}(d\omega)\mu_{\mathscr{M}}(dm)dt$-a.e.

Proof. Since the marked point process N is non-explosive, using similar arguments as in Lemma L2 of Brémaud [19, p. 24], one can show that, for all bounded sets $A \in \mathcal{B}(\mathbb{R}_{>0})$,

$$\iint_{A\times\mathscr{M}} \tilde{\lambda}(t, m)\mu_{\mathscr{M}}(dm)dt < \infty\,, \quad \text{a.s.},$$

which implies that $\tilde{\lambda}(\omega, t, m) < \infty$ holds $\mathbb{P}(d\omega)dt\mu_{\mathscr{M}}(dm)$-a.e. By a composition argument (see the beginning of the proof of Lemma 7.7.3), since $\tilde{\lambda}$ is $\mathbb{F}$-predictable, we have that $(\omega, t, m) \mapsto \mathbb{1}_{\{\tilde{\lambda}(\omega,t,m)<\infty\}}$ is also $\mathbb{F}$-predictable. It is then easy to check that $\lambda(\omega, t, m) := \mathbb{1}_{\{\tilde{\lambda}(\omega,t,m)<\infty\}}\tilde{\lambda}(\omega, t, m)$ is the $\mathbb{F}$-intensity of N where we use the convention $0 \times \infty = 0$. ☐

The next lemma says that by integrating the intensity against the state variable x, we obtain the intensity of the marked point process that tracks the event types, ignoring the state process.

Lemma 6.4.3. Let N be a marked point process on $\mathbb{R}\times\mathscr{M}$ with $\mathbb{F}$-intensity λ relative to $\mu_{\mathscr{M}}$. Then, $N_{\mathcal{E}}(\cdot) := N(\cdot \times \mathcal{X})$ is a non-explosive marked point process on $\mathbb{R} \times \mathcal{E}$ with $\mathbb{F}$-intensity $\lambda_{\mathcal{E}} : \Omega \times \mathbb{R}_{>0} \times \mathcal{E} \to \mathbb{R}_{\geq 0}$ relative to $\mu_{\mathcal{E}}$ such that $\lambda_{\mathcal{E}}(\omega, t, e) = \int_{\mathcal{X}} \lambda(\omega, t, e, x)\mu_{\mathcal{X}}(dx)$ holds $\mathbb{P}(d\omega)dt\mu_{\mathcal{E}}(de)$-a.e.

Proof. Let $H : \Omega\times\mathbb{R}_{>0}\times\mathcal{E} \to \mathbb{R}_{\geq 0}$ be an $\mathbb{F}$-predictable non-negative process. Then, by applying the definition of $N_{\mathcal{E}}$ and using Tonelli's theorem, we obtain that

$$\begin{aligned}\mathbb{E}\left[\iint_{\mathbb{R}_{>0}\times\mathcal{E}} H(t, e)N_{\mathcal{E}}(dt, de)\right] &= \mathbb{E}\left[\iiint_{\mathbb{R}_{>0}\times\mathcal{E}\times\mathcal{X}} H(t, e)N(dt, de, dx)\right]\\ &= \mathbb{E}\left[\iiint_{\mathbb{R}_{>0}\times\mathcal{E}\times\mathcal{X}} H(t, e)\lambda(t, e, x)\mu_{\mathcal{X}}(dx)\mu_{\mathcal{E}}(de)dt\right]\end{aligned}$$

$$= \mathbb{E}\left[\iint_{\mathbb{R}_{>0}\times\mathcal{E}} H(t,e)\left(\int_{\mathcal{X}} \lambda(t,e,x)\mu_{\mathcal{X}}(dx)\right)\mu_{\mathcal{E}}(de)dt\right].$$

The process $\int_{\mathcal{X}} \lambda(t,e,x)\mu_{\mathcal{X}}(dx)$, $t \in \mathbb{R}_{>0}$, $e \in \mathcal{E}$, is $\mathbb{F}$-predictable, see for example Lemma 25.23 in Kallenberg [74, p. 503] and we conclude using Lemma 6.4.2. □

We now check that an intensity functional applied to the history of a point process defines a predictable process.

Lemma 6.4.4. Let $\psi : \mathcal{U} \times \mathcal{N}^{\#}_{\mathbb{R}\times\mathcal{U}} \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ be a measurable functional and N be a non-explosive point process on $\mathbb{R}\times\mathcal{U}$ that is $\mathbb{F}$-adapted. Then, the process $\lambda : \Omega \times \mathbb{R} \times \mathcal{U} \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ defined by $\lambda(\omega, t, u) = \psi(u \,|\, \theta_t N(\omega)^{<0})$, $\omega \in \Omega, t \in \mathbb{R}, u \in \mathcal{U}$, is $\mathbb{F}$-predictable.

Proof. By Lemma 8.7.4, $\theta_t N(\omega)^{<0}$ is left-continuous in t and, by assumption, the process $(\theta_t N^{<0})_{t\in\mathbb{R}}$ is $\mathbb{F}$-adapted. As a consequence, the mapping $(\omega, t) \mapsto \theta_t N(\omega)^{<0}$ is $\mathbb{F}$-predictable, see for example Lemmas 25.1 and 1.10 in Kallenberg [74, p. 491, p. 6]. We then obtain that λ is $\mathbb{F}$-predictable by viewing it as the composition $(\omega, t, u) \mapsto (u, \theta_t N(\omega)^{<0}) \mapsto \psi(u \,|\, \theta_t N(\omega)^{<0})$ and using the measurability of ψ. □

The next lemma essentially says that if two predictable processes coincide at all event times of a marked point process, then they coincide everywhere under positive intensity. A less general variant of this result and its proof are suggested in Brémaud [19, Theorem T12, p. 31].

Lemma 6.4.5. Let $N_{\mathcal{E}}$ be a non-explosive marked point process on $\mathbb{R} \times \mathcal{E}$ with $\mathbb{F}$-intensity $\lambda_{\mathcal{E}}$ relative to $\mu_{\mathcal{E}}$. Let $H_1 : \Omega \times \mathbb{R}_{>0} \times \mathscr{M} \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ and $H_2 : \Omega \times \mathbb{R}_{>0} \times \mathscr{M} \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ be two non-negative $\mathbb{F}$-predictable processes. Then, $H_1 = H_2$ holds $\mathbb{P}(d\omega)N_{\mathcal{E}}(\omega, dt, de)\mu_{\mathcal{X}}(dx)$-a.e. if and only if $H_1 = H_2$ holds $\mathbb{P}(d\omega)\lambda_{\mathcal{E}}(\omega, t, e)dt\mu_{\mathscr{M}}(de, dx)$-a.e.

Proof. By a composition argument, since H_1 and H_2 are $\mathbb{F}$-predictable, we have that the function $(\omega, t, m) \mapsto \mathbb{1}_{\{H_1(\omega,t,m)\neq H_1(\omega,t,m)\}}$ is $\mathbb{F}$-predictable (see the beginning of the proof of Lemma 7.7.3). By Lemma 25.23 in Kallenberg

[74, p. 503], we also have that the process $\int_{\mathcal{X}} \mathbb{1}_{\{H_1(\cdot,\cdot,\cdot,x)\neq H_1(\cdot,\cdot,\cdot,x)\}}\mu_{\mathcal{X}}(dx)$ is $\mathbb{F}$-predictable. Using the definition of the intensity and Tonelli's theorem, we obtain that

$$\begin{aligned}
&\int_\Omega \int_{\mathbb{R}_{>0}\times\mathscr{M}} \mathbb{1}_{\{H_1(\omega,t,m)\neq H_1(\omega,t,m)\}}\mu_{\mathcal{X}}(dx)N_{\mathcal{E}}(\omega,dt,de)\mathbb{P}(d\omega)\\
&= \int_\Omega \int_{\mathbb{R}_{>0}\times\mathcal{E}} \left(\int_{\mathcal{X}} \mathbb{1}_{\{H_1(\omega,t,m)\neq H_1(\omega,t,m)\}}\mu_{\mathcal{X}}(dx)\right) N_{\mathcal{E}}(\omega,dt,de)\mathbb{P}(d\omega)\\
&= \int_\Omega \int_{\mathbb{R}_{>0}\times\mathscr{M}} \mathbb{1}_{\{H_1(\omega,t,m)\neq H_1(\omega,t,m)\}}\mu_{\mathcal{X}}(dx)\lambda_{\mathcal{E}}(\omega,t,e)\mu_{\mathcal{E}}(de)dt\mathbb{P}(d\omega),
\end{aligned}$$

from which the assertion follows. □

Finally, we show that the link between joint densities and conditional densities still holds when we pre-condition on a sub-σ-algebra. Since $\mathscr{M}$ is a complete separable metric space and, in particular, Borel, random elements in $\mathscr{M}$ always have regular conditional distributions [74, p. 106, Theorem A1.2, p. 561].

Lemma 6.4.6. Let (E, X) be a random element in $\mathscr{M}$. Let $\mathcal{G}$ be a sub-σ-algebra, i.e., $\mathcal{G} \subset \mathcal{F}$, and let $A \in \mathcal{G}$ such that $\mathbb{P}(A) > 0$. Moreover, let $f : \Omega \times \mathscr{M} \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ be a non-negative measurable function that is $\mathcal{G} \otimes \mathcal{B}(\mathscr{M})$-measurable. If we have

$$\mathbb{P}\left(E \in de, X \in dx \,|\, \mathcal{G}\right) \mathbb{1}_A = f(e,x)\mu_{\mathscr{M}}(de,dx)\mathbb{1}_A, \quad \text{a.s.,} \tag{6.5}$$

then

$$\mathbb{P}\left(X \in dx \,|\, \sigma(E) \vee \mathcal{G}\right) \mathbb{1}_A = \frac{f(E,x)}{\int_{\mathcal{X}} f(E,x')\mu_{\mathcal{X}}(dx')}\mu_{\mathcal{X}}(dx)\mathbb{1}_A, \quad \text{a.s.}$$

Proof. Let $B \in \mathcal{B}(\mathcal{X})$, $G \in \mathcal{G}$ and $H \in \sigma(E)$. On the one hand,

$$\begin{aligned}
\mathbb{E}\left[\mathbb{1}_G \mathbb{1}_H \mathbb{1}_{\{X\in B\}}\mathbb{1}_A\right] &= \mathbb{E}\left[\mathbb{1}_G h(E)\mathbb{1}_{\{X\in B\}}\mathbb{1}_A\right] = \mathbb{E}\left[\mathbb{1}_G \mathbb{E}\left[h(E)\mathbb{1}_{\{X\in B\}}\mathbb{1}_A \,|\, \mathcal{G}\right]\right]\\
&= \mathbb{E}\left[\mathbb{1}_G \mathbb{1}_A \int_{\mathcal{E}}\int_B h(e)f(e,x)\mu_{\mathcal{X}}(dx)\mu_{\mathcal{E}}(de)\right],
\end{aligned} \tag{6.6}$$

where we successively used Lemma 1.13 in Kallenberg [74, p. 7] to write

$\mathbb{1}_H = h(E)$ using a measurable function $h : \mathcal{E} \to \{0,1\}$, the Tower property, the disintegration theorem in Kallenberg [74, Theorem 6.4, p. 108] with the regular conditional distribution of (6.5) and, finally, the product form of $\mu_{\mathscr{M}}$. Note that, here, the disintegration theorem is applied to the probability measure $\mathbb{P}(\cdot \cap A)/\mathbb{P}(A)$ on the measurable space $(A, A \cap \mathcal{F})$. On the other hand, observe that (6.1) and (6.5) imply that

$$\mathbb{P}\left(E \in de \,|\, \mathcal{G}\right) \mathbb{1}_A = \int_{\mathcal{X}} f(e,x)\mu_{\mathcal{X}}(dx)\mu_{\mathcal{E}}(de)\mathbb{1}_A, \quad \text{a.s.}$$

Then, using similar arguments,

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{1}_G \mathbb{1}_H \int_B \frac{f(E,x)}{\int_{\mathcal{X}} f(E,x')\mu_{\mathcal{X}}(dx')}\mu_{\mathcal{X}}(dx)\mathbb{1}_A\right] \\
&= \mathbb{E}\left[\mathbb{1}_G h(E) \int_B \frac{f(E,x)}{\int_{\mathcal{X}} f(E,x')\mu_{\mathcal{X}}(dx')}\mu_{\mathcal{X}}(dx)\mathbb{1}_A\right] \\
&= \mathbb{E}\left[\mathbb{1}_G \mathbb{E}\left[h(E) \int_B \frac{f(E,x)}{\int_{\mathcal{X}} f(E,x')\mu_{\mathcal{X}}(dx')}\mu_{\mathcal{X}}(dx)\mathbb{1}_A \,\middle|\, \mathcal{G}\right]\right] \\
&= \mathbb{E}\left[\mathbb{1}_G \mathbb{1}_A \int_{\mathcal{E}} \left(h(e) \int_B \frac{f(e,x)}{\int_{\mathcal{X}} f(e,x')\mu_{\mathcal{X}}(dx')}\mu_{\mathcal{X}}(dx)\right) \int_{\mathcal{X}} f(e,x')\mu_{\mathcal{X}}(dx')\mu_{\mathcal{E}}(de)\right].
\end{aligned}$$

Tonelli's theorem and (6.6) then imply that

$$\mathbb{E}\left[\mathbb{1}_G \mathbb{1}_H \mathbb{1}_{\{X \in B\}} \mathbb{1}_A\right] = \mathbb{E}\left[\mathbb{1}_G \mathbb{1}_H \int_B \frac{f(E,x)}{\int_{\mathcal{X}} f(E,x')\mu_{\mathcal{X}}(dx')}\mu_{\mathcal{X}}(dx)\mathbb{1}_A\right]. \tag{6.7}$$

Using a monotone class argument, we show below that (6.7) can be extended to

$$\mathbb{E}\left[\mathbb{1}_F \mathbb{1}_{\{X \in B\}} \mathbb{1}_A\right] = \mathbb{E}\left[\mathbb{1}_F \int_B \frac{f(E,x)}{\int_{\mathcal{X}} f(E,x')\mu_{\mathcal{X}}(dx')}\mu_{\mathcal{X}}(dx)\mathbb{1}_A\right] \tag{6.8}$$

for all $F \in \sigma(E) \vee \mathcal{G}$, which means exactly that

$$\mathbb{P}\left(X \in B \,|\, \sigma(E) \vee \mathcal{G}\right) \mathbb{1}_A = \int_B \frac{f(E,x)}{\int_{\mathcal{X}} f(E,x')\mu_{\mathcal{X}}(dx')}\mu_{\mathcal{X}}(dx)\mathbb{1}_A, \quad \text{a.s.,}$$

as asserted.

To prove (6.8), define the functions

$$\begin{aligned}\mu_1 :\quad & \sigma(E) \vee \mathcal{G} \to [0,1] \\ & \qquad F \mapsto \mu_1(F) := \mathbb{E}\left[\mathbb{1}_F \mathbb{1}_{\{X \in B\}} \mathbb{1}_A\right], \\ \mu_2 :\quad & \sigma(E) \vee \mathcal{G} \to [0,1] \\ & \qquad F \mapsto \mu_2(F) := \mathbb{E}\left[\mathbb{1}_F \int_B \frac{f(E,x)}{\int_{\mathcal{X}} f(E,x')\mu_{\mathcal{X}}(dx')} \mu_{\mathcal{X}}(dx) \mathbb{1}_A\right].\end{aligned}$$

One can check that μ_1 and μ_2 are bounded measures on $(\Omega, \sigma(E) \vee \mathcal{G})$ (to swap an expectation with an infinite sum, use the monotone convergence theorem, see for example Theorem 1.19 in Kallenberg [74, p. 11]). Define also the class $\mathcal{C} := \{G \cap H \,:\, G \in \mathcal{G},\ H \in \sigma(E)\}$. Equation (6.7) means that $\mu_1(C) = \mu_2(C)$ for all $C \in \mathcal{C}$. Moreover, $\mathcal{C}$ is a π-system such that $\Omega \in \mathcal{C}$. Also, note that $\sigma(E) \cup \mathcal{G} \subset \mathcal{C} \subset \sigma(E) \vee \mathcal{G}$ and, thus, $\sigma(\mathcal{C}) = \sigma(E) \vee \mathcal{G}$. As a consequence, we can apply Lemma 1.17 in Kallenberg [74, p. 9] to conclude that $\mu_1(F) = \mu_2(F)$ for all $F \in \sigma(E) \vee \mathcal{G}$, meaning that (6.8) holds. □

6.4.2 Implied dynamics and characterisation

Proof of Theorem 6.2.5. Recall that we denote by λ the $\mathbb{F}^N$-intensity of N relative to $\mu_{\mathscr{M}}$ and by $\lambda_{\mathcal{E}}$ the $\mathbb{F}^N$-intensity of $N_{\mathcal{E}} := N(\cdot \times \mathcal{X})$ relative to $\mu_{\mathcal{E}}$.

Necessity. Assume that N is a hybrid marked point process with transition function ϕ and event functional η. We first observe that statement (i) holds simply by applying Lemma 6.4.3 and using the fact that $\phi(\cdot \mid e, x)$ is a probability density for all $e \in \mathcal{E}$ and $x \in \mathcal{X}$.

Next, we show that statement (ii) holds. When $\mathbb{P}(\tau_t < \infty) = 0$, this is clearly true and, thus, we assume that $\mathbb{P}(\tau_t < \infty) > 0$. By applying Theorem T6 in Brémaud [19, p. 236], we obtain that, for all $M \in \mathcal{B}(\mathscr{M})$,

$$\mathbb{P}\left((E,X) \in M \mid \mathcal{F}^N_{\tau_t -}\right) \mathbb{1}_{\{\tau_t < \infty\}} = \frac{\int_M \lambda(\tau_t, m)\mu_{\mathscr{M}}(dm)}{\int_{\mathscr{M}} \lambda(\tau_t, m')\mu_{\mathscr{M}}(dm')} \mathbb{1}_{\{\tau_t < \infty\}}, \quad \text{a.s.}$$

This is allowed since Lemma 6.4.1 tells us that the filtration $\mathbb{F}^N$ is within the framework of this result. Hence, we have identified the unique regular conditional distribution of (E, X) given $\mathcal{F}^N_{\tau_t-}$ on the measurable space

$$(\{\tau_t < \infty\}, \{\tau_t < \infty\} \cap \mathcal{F})$$

equipped with the measure $\mathbb{P}(\cdot \cap \{\tau_t < \infty\})/\mathbb{P}(\{\tau_t < \infty\})$ [74, Theorem 6.3, p. 107]. Besides, observe that the mapping $(\omega, m) \mapsto \lambda(\omega, \tau_t(\omega), m)\mathbb{1}_{\{\tau_t(\omega)<\infty\}}$ is $\mathcal{F}^N_{\tau_t-} \otimes \mathcal{B}(\mathscr{M})$-measurable, see for example Lemma 25.3 in Kallenberg [74, p. 492]. Using Lemma 1.26 in Kallenberg [74, p. 14], we obtain that the function f defined by

$$f(\omega, e, x) = \frac{\lambda(\omega, \tau_t(\omega), e, x)}{\int_{\mathscr{M}} \lambda(\omega, \tau_t(\omega), m')\mu_{\mathscr{M}}(dm')} \mathbb{1}_{\{\tau_t(\omega)<\infty\}}, \quad \omega \in \Omega, e \in \mathcal{E}, x \in \mathcal{X},$$

is $\mathcal{F}^N_{\tau_t-} \otimes \mathcal{B}(\mathscr{M})$-measurable. We can then apply Lemma 6.4.6 with $\mathcal{G} = \mathcal{F}^N_{\tau_t-}$ and $A = \{\tau_t < \infty\}$. This yields that

$$\begin{aligned} &\mathbb{P}\left(X \in dx \,|\, \sigma(E) \vee \mathcal{F}^N_{\tau_t-}\right) \mathbb{1}_{\{\tau_t<\infty\}} = \\ &\qquad \frac{\lambda(\tau_t, E, x)}{\int_{\mathcal{X}} \lambda(\tau_t, E, x')\mu_{\mathcal{X}}(dx')}\mu_{\mathcal{X}}(dx)\mathbb{1}_{\{\tau_t<\infty\}}, \quad \text{a.s.} \end{aligned} \tag{6.9}$$

By viewing the term $\phi(x \,|\, e, X_t)$ as a measurable function φ applied to the triplet $(e, x, \theta_t N^{<0})$ where $\varphi(x, e \,|\, \xi) = \phi(x \,|\, e, F(\xi))$, and using the measurability of the state functional F and the transition function ϕ, we obtain by Lemma 6.4.4 that $\phi(x \,|\, e, \theta_t N^{<0})$, $t \in \mathbb{R}$, $e \in \mathcal{E}$, $m \in \mathscr{M}$, is $\mathbb{F}^N$-predictable. Similarly, note that $\eta(e \,|\, \theta_t N^{<0})$, $t \in \mathbb{R}$, $e \in \mathcal{E}$, is also $\mathbb{F}^N$-predictable (this will be useful when proving sufficiency). Besides, thanks to the assumption on λ,

$$\frac{\lambda(\omega, t, e, x)}{\int_{\mathcal{X}} \lambda(\omega, t, e, x')\mu_{\mathcal{X}}(dx')} = \phi(x \,|\, e, X_t(\omega)), \quad \mathbb{P}(d\omega)dt\mu_{\mathscr{M}}(de, dx)\text{-a.e.}$$

Hence, using Lemma 6.4.5, (6.9) becomes

$$\mathbb{P}\left(X \in dx \,|\, \sigma(E) \vee \mathcal{F}^N_{\tau_t-}\right) \mathbb{1}_{\{\tau_t<\infty\}} = \phi(x \,|\, E, X_{\tau_t})\mu_{\mathcal{X}}(dx)\mathbb{1}_{\{\tau_t<\infty\}}, \quad \text{a.s.}$$

To obtain (6.3), it remains to notice that $X_{\tau_t} = X_{t+}$ on $\{\tau_t < \infty\}$ since there is no event on the time interval (t, τ_t) by definition of τ_t. Also, since the ground point process $N(\cdot \times \mathscr{M})$ admits an $\mathbb{F}^N$-intensity, we have that $N(\{t\} \times \mathscr{M}) = 0$ a.s., implying that $X_{t+} = X_t$ a.s. To show the statement in Remark 6.2.6, simply use (6.3) and the tower property to obtain that

$$\begin{aligned}
\mathbb{E}\left[\mathbb{1}_F \mathbb{1}_{\{\tau_t<\infty\}} \mathbb{1}_{X\in B}\right] &= \mathbb{E}\left[\mathbb{1}_F \mathbb{1}_{\{\tau_t<\infty\}} \mathbb{E}\left[\mathbb{1}_{X\in B} \,|\, \sigma(E) \vee \mathcal{F}^N_{\tau_t-}\right]\right] \\
&= \mathbb{E}\left[\mathbb{1}_F \mathbb{1}_{\{\tau_t<\infty\}} \int_B \phi(x \,|\, E, X_t) \mu_{\mathcal{X}}(dx)\right]
\end{aligned}$$

for all $F \in \sigma(E) \vee \mathcal{F}^N_t$, $B \in \mathcal{B}(\mathcal{X})$ and observe that $\int_B \phi(x \,|\, E, X_t)\mu_{\mathcal{X}}(dx)$ is $\sigma(E) \vee \mathcal{F}^N_t$-measurable.

Sufficiency. Assume that N is a non-explosive marked point process on $\mathbb{R} \times \mathscr{M}$ such that it admits an $\mathbb{F}^N$-intensity relative to $\mu_{\mathscr{M}}$ and such that statements (i) and (ii) hold. We want to show that

$$\lambda(\omega, t, e, x) = \phi(x \,|\, e, X_t(\omega))\eta(e \,|\, \theta_t N(\omega)^{<0})$$

holds $\mathbb{P}(d\omega)\mu_{\mathscr{M}}(de, dx)dt$-a.e. For all $t \in \mathbb{R}_{\geq 0}$, by using statement (ii), (6.9), Lemmas 6.4.3 and 6.4.5, and statement (i), we obtain that

$$\begin{aligned}
\phi(x \,|\, E, X_{\tau_t})\mu_{\mathcal{X}}(dx)\mathbb{1}_{\{\tau_t<\infty\}} &= \mathbb{P}\left(X \in dx \,|\, \sigma(E) \vee \mathcal{F}^N_{\tau_t-}\right) \mathbb{1}_{\{\tau_t<\infty\}} \\
&= \frac{\lambda(\tau_t, E, x)}{\int_{\mathcal{X}} \lambda(\tau_t, E, x')\mu_{\mathcal{X}}(dx')} \mu_{\mathcal{X}}(dx)\mathbb{1}_{\{\tau_t<\infty\}} \\
&= \frac{\lambda(\tau_t, E, x)}{\lambda_{\mathcal{E}}(\tau_t, E)} \mu_{\mathcal{X}}(dx)\mathbb{1}_{\{\tau_t<\infty\}} \\
&= \frac{\lambda(\tau_t, E, x)}{\eta(E \,|\, \theta_{\tau_t} N^{<0})} \mu_{\mathcal{X}}(dx)\mathbb{1}_{\{\tau_t<\infty\}}, \quad \text{a.s.}
\end{aligned}$$

This means that, for all $t \in \mathbb{R}_{\geq 0}$, we have that

$$\lambda(\tau_t, E, x)\mathbb{1}_{\{\tau_t<\infty\}} = \phi(x \,|\, E, X_{\tau_t})\eta(E \,|\, \theta_{\tau_t} N^{<0})\mathbb{1}_{\{\tau_t<\infty\}}, \quad \mu_{\mathcal{X}}(dx)\text{-a.e., a.s.}$$

This holds a.s. simultaneously for all $t \in \mathbb{Q} \cap \mathbb{R}_{\geq 0}$, whence, using that the number of events in N is countable and finite in any bounded time interval,

$$\lambda(\omega, t, e, x) = \phi(x \,|\, e, X_t(\omega))\eta(e \,|\, \theta_t N(\omega)^{<0}), \quad \mathbb{P}(d\omega)N_{\mathcal{E}}(\omega, dt, de)\mu_{\mathcal{X}}(dx)\text{-a.e.}$$

By Lemma 6.4.5, the above equality then implies that

$$\lambda(\omega, t, e, x) = \phi(x \mid e, X_t(\omega))\eta(e \mid \theta_t N(\omega)^{<0}), \qquad \mathbb{P}(d\omega)\lambda_{\mathcal{E}}(\omega, t, e)dt\mu_{\mathscr{M}}(de, dx)\text{-a.e.}$$

By noticing that, on $\lambda_{\mathcal{E}}(\omega, t, e) = 0$, we have that $\eta(e \mid \theta_t N(\omega)^{<0}) = 0$ holds $\mathbb{P}(d\omega)dt\mu_{\mathcal{E}}(de)$-a.e. and that $\lambda(\omega, t, e, x) = 0$ holds $\mathbb{P}(d\omega)dt\mu_{\mathcal{E}}(de)\mu_{\mathcal{X}}(dx)$-a.e. (using again Lemma 6.4.3), we conclude that the above equation actually holds $\mathbb{P}(d\omega)dt\mu_{\mathscr{M}}(de, dx)$-a.e. □

7

Strong existence and uniqueness

In this chapter, we prove the strong existence and uniqueness of non-explosive state-dependent Hawkes processes and, more generally, hybrid marked point processes. In fact, by dispensing with a Lipschitz condition, we extend the results currently available in the literature for marked point processes defined via their intensity.

It is known that a marked point process with an intensity λ expressed in terms of an intensity functional ψ can be formulated as a solution to a Poisson-driven stochastic differential equation (SDE) [91, 21]. However, the existence and uniqueness results available in these works cannot be applied to hybrid marked point processes because their intensity functional may fail to satisfy the Lipschitz condition imposed therein. We show that, under certain integrability or decay conditions, it is enough for ψ to be dominated by either a Hawkes functional or an increasing function of the total number of past events in order to obtain the existence of a strong solution to the Poisson-driven SDE (Theorem 7.4.1) and, in particular, the existence of hybrid marked point processes (Corollary 7.4.2). The solution is constructed piece by piece along the time axis in a pathwise manner, taking advantage of the discrete nature of the driving Poisson random measure. A domination argument is then used to show non-explosiveness. In the context of multivariate point processes, a similar construction has already been considered in [30], while a similar domination argument is given in [28]. We combine the two in a more general setting (i.e., general mark space, initial conditions and intensity functional). We are also able to obtain strong and weak uniqueness without any specific assumptions (Theorems 7.5.1 and 7.5.2) and, in particular, uniqueness of hybrid marked point processes (Corollary 7.5.3).

Note that, in this chapter, $\mathscr{M}$ is not required to be a product space as in Chapter 6 but can be again an arbitrary complete separable metric space.

7.1 The existence and uniqueness problem

A hybrid marked point processes (Definition 6.2.1) is defined implicitly via its intensity process, which, in turn, depends on the history of the hybrid marked point process. Due to the self-referential nature of the definition, it is not clear a priori that such a marked point process exists. More generally, given an initial condition $N_{\leq 0}$ (see Subsection 5.7) and a measurable intensity functional $\psi : \mathscr{M} \times \mathcal{N}^{\#}_{\mathbb{R}\times\mathscr{M}} \to \mathbb{R}_{\geq 0} \cup \{\infty\}$, one can ask if there exists a unique non-explosive marked point process N that satisfies the initial condition $N_{\leq 0}$ on $\mathbb{R}_{\leq 0}$ and admits ψ as its intensity functional on $\mathbb{R}_{>0}$.

Massoulié [91] tackles this question by reformulating the existence problem as a Poisson-driven SDE, extending the works of Brémaud and Massoulié [21], Grigelionis [56] and Kerstan [76]. Delattre et al. [43] also employ this Poisson embedding technique in the context of Hawkes processes on infinite directed graphs. However, in these papers, strong existence and uniqueness is obtained by imposing a Lipschitz condition on the intensity functional ψ. More precisely, it is assumed that there exists a non-negative kernel $\overline{k} : \mathbb{R}_{>0} \times \mathscr{M} \times \mathscr{M} \to \mathbb{R}_{\geq 0}$ such that

$$|\psi(m \,|\, \xi) - \psi(m \,|\, \xi')| \leq \iint_{\mathbb{R}_{<0}\times\mathscr{M}} \overline{k}(-t', m', m)|\xi - \xi'|(dt', dm'), \qquad (7.1)$$

for all $m \in \mathscr{M}$, $\xi,\xi' \in \mathcal{N}^{\#}_{\mathbb{R}\times\mathscr{M}}$. Unfortunately, this condition is too restrictive in the context of hybrid marked point processes. A simple, yet natural, example of a hybrid marked point process not satisfying (7.1) is given in Subsection 7.7.1. Hence, our goal is to construct a strong solution to a Poisson-driven SDE without imposing the Lipschitz condition (7.1) on the intensity functional ψ. We will in fact extend the existence result in Massoulié [91] by imposing only a weaker sublinearity condition on ψ. The idea to define a random measure as a strong solution to an SDE driven by another random measure was also studied by Jacod [69]. Similarly, a Lipschitz condition that does not seem to apply to Hawkes processes and hybrid marked point processes is required [69, Chapter 14, Section 1].

Let us also briefly review some weak existence and uniqueness results. Jacod [68] proved that there exists a unique probability measure on the canonical space of marked point processes such that the canonical marked point process admits a given compensator. However, this marked point process may be explosive a priori. Still, we will apply this result in the proof of Theorem 7.5.2 below to obtain weak uniqueness. A similar approach is followed by Jacobsen [67, Proposition 4.3.5, Corollary 4.4.4] who, furthermore, gives a domination condition on the intensity functional ensuring that the corresponding marked point process is non-explosive. Proposition 7.7.12 will be the counterpart of this result in the strong setting. These weak existence results are however limited to intensities with respect to the internal history $\mathbb{F}^N$. The advantage of the strong setting is that the results of Massoulié [91] and the pathwise construction of this chapter also hold when the intensity functional depends additionally on an auxiliary process, meaning that intensities with respect to larger filtrations can be considered. Besides, the Poisson-driven SDE representation of marked point processes considered in the strong setting directly suggests a simulation (thinning) algorithm. In fact, the Poisson embedding lemma (Lemma 7.7.3 below), which is a stepping stone to the strong setting, was first given in the simulation literature [83, 99].

Finally, there is a third approach to obtain existence, based on a change of measure, see Brémaud [19, Theorem 11, p. 242] and Sokol and Hansen [113]. While this technique also accommodates filtrations that are larger than the internal history, existence is generally obtained only on finite time intervals.

7.2 The Poisson-driven SDE

Let $(\Omega_{>0}, \mathcal{F}_{>0}, \mathbb{P}_{>0})$ be given and let $M_{>0} : \Omega_{>0} \to \mathcal{N}^{\#}_{\mathbb{R}\times\mathscr{M}\times\mathbb{R}}$ be a Poisson process on $\mathbb{R} \times \mathscr{M} \times \mathbb{R}$ with mean measure $dt\mu_{\mathscr{M}}(dm)dz$. As usual, denote by $(\mathcal{F}^{M_{>0}}_t)_{t\in\mathbb{R}}$ the internal history of $M_{>0}$ on $\Omega_{>0}$. In this Subsection, we work under the assumption that underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is the completion of the product probability space defined by (5.3). In particular, $\Omega := \Omega_{\leq 0} \times \Omega_{>0}$, where $\Omega_{\leq 0}$ corresponds to the probability space of an initial condition $N_{\leq 0}$, see Subsection 5.7. We extend $M_{>0}$ to a mapping $M : \Omega \to \mathcal{N}^{\#}_{\mathbb{R}\times\mathscr{M}\times\mathbb{R}}$ by simply setting

$$M(\omega) := M_{>0}(\omega_{>0}), \quad \omega = (\omega_{\leq 0}, \omega_{>0}) \in \Omega. \tag{7.2}$$

Let $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{R}}$ be the filtration on Ω such that, for all $t \in \mathbb{R}$, $\mathcal{F}_t$ is the $\mathbb{P}$-completion of $\mathcal{F}^{N_{\leq 0}}_t \otimes \mathcal{F}^{M_{>0}}_t$ in $\mathcal{F}$. In particular, the filtration $\mathbb{F}$ is complete [74, p. 123]. Similarly to Massoulié [91], we want to solve the following Poisson-driven SDE.

Definition 7.2.1 (The Poisson-driven SDE)**.**
Let $\psi : \mathscr{M} \times \mathcal{N}^{\#}_{\mathbb{R}\times\mathscr{M}} \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ be a given measurable functional. By a solution to the *Poisson-driven SDE*, we mean an $\mathbb{F}$-adapted non-explosive marked-point process $N : \Omega \to \mathcal{N}^{\#g}_{\mathbb{R}\times\mathscr{M}}$ that solves

$$\begin{cases} N(dt, dm) = M(dt, dm, (0, \lambda(t, m)]), & t \in \mathbb{R}_{>0}, \text{ a.s.}, \\ \lambda(\omega, t, m) = \psi(m \,|\, \theta_t N(\omega)^{<0}), & t \in \mathbb{R}_{>0}, m \in \mathscr{M}, \omega \in \Omega, \\ N^{\leq 0}(\omega) = N_{\leq 0}(\omega_{\leq 0}), & \omega = (\omega_{\leq 0}, \omega_{>0}) \in \Omega, \text{a.s.}, \end{cases} \tag{7.3}$$

where $N_{\leq 0}$ is a given initial condition (see Subsection 5.7).

Still, notice that our problem differs slightly as we only search for solutions in the space of non-explosive marked point processes, a smaller space than the one considered in Massoulié [91].

7.3 Assumptions

The following assumptions are only required for the strong existence result (Theorem 7.4.1 below). We first need to assume that the mark space $\mathscr{M}$ has finite total mass.

Assumption A. *The reference measure $\mu_{\mathscr{M}}$ is finite, i.e., $\mu_{\mathscr{M}}(\mathscr{M}) < \infty$.*

Next, we need to control for both the intensity functional ψ and the initial condition $N_{\leq 0}$. We will prove Theorem 7.4.1 for two different scenarios. In the first scenario, the intensity is dominated by an increasing function of the total number of past events, while the number of events before time 0 is finite.

Assumption B. *There exists a non-decreasing function $a : \mathbb{N} \cup \{\infty\} \to \mathbb{N} \cup \{\infty\}$ with $a(n) < \infty$ for all $n \in \mathbb{N}$ and $a(\infty) = \infty$ such that:*

(i) $\psi(m \,|\, \xi) \leq a(\xi((-\infty, 0) \times \mathscr{M}))$, $m \in \mathscr{M}$, $\xi \in \mathcal{N}^{\#}_{\mathbb{R}\times\mathscr{M}}$;

(ii) $\sum_{n=0}^{\infty} a(n)^{-1} = \infty$.

Assumption C. *The initial condition satisfies $N_{\leq 0}(\omega_{\leq 0}, (-\infty, 0] \times \mathscr{M})) < \infty$ for all $\omega_{\leq 0} \in \Omega_{\leq 0}$.*

In the second scenario, the intensity functional ψ is dominated by a Hawkes functional. Note that this requirement is weaker than the Lipschitz condition (7.1) in Massoulié [91].

Assumption D. *There exists $\lambda_0 \in \mathbb{R}_{\geq 0}$ and a measurable function $\overline{k} : \mathbb{R}_{>0} \times \mathscr{M} \times \mathscr{M} \to \mathbb{R}_{\geq 0}$ such that:*

(i) $\psi(m \,|\, \xi) \leq \lambda_0 + \iint_{(-\infty,0)\times\mathscr{M}} \overline{k}(-t', m', m)\xi(dt', dm')$, *for all* $m \in \mathscr{M}$, $\xi \in \mathcal{N}^{\#}_{\mathbb{R}\times\mathscr{M}}$;

(ii) $\rho := \sup_{m\in\mathscr{M}} \iint_{(0,\infty)\times\mathscr{M}} \overline{k}(t', m', m)\mu_{\mathscr{M}}(dm')dt' < 1$;

(iii) $\sup_{m\in\mathscr{M}} k(t', m', m) < \infty$ *for all* $t' \in \mathbb{R}_{>0}$, $m' \in \mathscr{M}$.

Assumption E. *The initial condition $N_{\leq 0}$ satisfies:*

(i) $\sup_{t>0,\, m\in\mathscr{M}} \mathbb{E}\left[\iint_{(-\infty,0]\times\mathscr{M}} \overline{k}(t-t',m',m)N_{\leq 0}(dt',dm')\right] < \infty$;

(ii) $\tilde{\lambda}_{\leq 0}(\omega_{\leq 0},t) := \sup_{m\in\mathscr{M}} \iint_{(-\infty,0]\times\mathscr{M}} \overline{k}(t-t',m',m)N_{\leq 0}(\omega_{\leq 0},dt',dm')$ *is finite for all* $\omega_{\leq 0}\in\Omega_{\leq 0}$, $t\in\mathbb{R}_{>0}$.

Note that Assumptions D.(ii) and E.(i) are needed in order to reuse Theorem 2 in Massoulié [91]. It will allow us to dominate the marked point process N by a Hawkes process with kernel $\overline{k}$.

7.4 Existence

We construct a solution to the Poisson-driven SDE in two mains steps. First, by taking advantage of the discrete nature of the driving Poisson process, we construct in a pathwise fashion a mapping $N:\Omega\to\mathcal{N}^{\infty}_{\mathbb{R}\times\mathscr{M}}$ that solves (7.3) up to each event time, generalising the construction in Çinlar [30, Chapter 6, p. 302-306] and Lindvall [86, p. 127]. Second, we dominate N by a non-explosive marked point process to show that N is itself non-explosive, generalising the argument in Chevallier [28, Lemma B.1, p. 30]. When working under Assumptions D and E, these two steps must actually be performed concurrently. Then, it turns out that this constructed N admits ψ as its intensity functional on $\mathbb{R}_{>0}$ and, thus, solves the existence problem. The proof of the following theorem, which extends the existence result in Massoulié [91], is given in Subsection 7.7.2.

Theorem 7.4.1 (Strong existence)**.** *Under either Assumptions A, B, C or Assumptions A, D, E, there exists a non-explosive marked point process, denoted by $N:\Omega\to\mathcal{N}^{\#g}_{\mathbb{R}\times\mathscr{M}}$, that solves the Poisson-driven SDE (Definition 7.2.1). Any such N satisfies the strong initial condition $N_{\leq 0}$ and admits ψ as its intensity functional on $\mathbb{R}_{>0}$.*

As a corollary, we obtain conditions that ensure the existence of hybrid marked point processes.

Corollary 7.4.2 (Existence of hybrid marked point processes)**.** *Suppose that Assumption A holds and $\|\phi\|_\infty < \infty$. Moreover, suppose that either Assumptions B and C or Assumptions D and E hold with $\psi(m \mid \xi)$ replaced by $\eta(e \mid \xi)$, where the dominating kernel $\overline{k}$ is now a function $\overline{k} : \mathbb{R}_{>0} \times \mathscr{M} \times \mathcal{E} \to \mathbb{R}_{\geq 0}$, and with the constraint $\rho < \|\phi\|_\infty^{-1}$. Then, there exists a hybrid marked point process $N : \Omega \to \mathcal{N}^{\#g}_{\mathbb{R}\times\mathscr{M}}$ with transition function ϕ and event functional η that satisfies the strong initial condition $N_{\leq 0}$.*

Example 7.4.3 (Existence of state-dependent Hawkes processes)**.** When the transition function ϕ is bounded, the above corollary encompasses the case of state-dependent Hawkes processes (Example 6.2.7) for either bounded kernels with no integrability constraint or unbounded kernels (up to the constraint D.(iii)) with an integrability constraint.

7.5 Uniqueness

As Massoulié [91] considers point processes on $\mathbb{R}\times\mathscr{M}$ that are not necessarily non-explosive marked point processes, he uses the Lipschitz condition (7.1) to obtain strong uniqueness in a space of regular point processes. Here, since we restrict ourselves to non-explosive marked point processes, the enumeration representation allows us to prove strong uniqueness more easily without any specific assumptions. The proof is deferred until Subsection 7.7.3.

Theorem 7.5.1 (Strong uniqueness)**.** *Let $N : \Omega \to \mathcal{N}^{\#g}_{\mathbb{R}\times\mathscr{M}}$ and $N' : \Omega \to \mathcal{N}^{\#g}_{\mathbb{R}\times\mathscr{M}}$ be two non-explosive marked point processes solving the Poisson-Driven SDE (Definition 7.2.1). Then $N = N'$ a.s.*

By applying Theorem 3.4 in Jacod [68], we can also obtain weak uniqueness. Alternatively, we could also have applied Theorem 14.2.IV in Daley and Vere-Jones [41, p. 381]. The idea is that the intensity and the conditional distributions $\mathbb{P}((T_{n+1}, M_{n+1}) \in \cdot \mid \mathcal{F}^N_{T_n})$ uniquely determine each other, see also Last and Brandt [80] and Jacobsen [67, Theorem 4.3.2, p. 54]. Another approach, as suggested by Massoulié [91], could be to use the fact that any marked point process with an intensity functional can be represented as the strong solution to a Poisson-driven SDE like in Definition 7.2.1, see Jacod

[69, Theorem 14.56, p. 472], and use the strong uniqueness result. We prove the following result in Subsection 7.7.3.

Theorem 7.5.2 (Weak uniqueness)**.** *Let N_1 and N_2 be two non-explosive marked point processes (possibly on distinct probability spaces) that admit the same intensity functional ψ on $\mathbb{R}_{>0}$. Assume also that both N_1 and N_2 satisfy the weak initial condition $N_{\leq 0}$. Then, we have that $\mathcal{P}^{N_1} = \mathcal{P}^{N_2}$, i.e., the induced probabilities measures on $\mathcal{N}^{\#}_{\mathbb{R}\times\mathscr{M}}$ coincide.*

As a corollary, we obtain the weak uniqueness of hybrid marked point processes.

Corollary 7.5.3 (Uniqueness of hybrid marked point processes)**.** *All hybrid marked point processes with transition function ϕ and event functional η that satisfy the weak initial condition $N_{\leq 0}$ induce the same probability measure on $\mathcal{N}^{\#}_{\mathbb{R}\times\mathscr{M}}$.*

Remark 7.5.4. Note that weak uniqueness might not hold for a general history $\mathbb{F}$. Given an $\mathbb{F}$-predictable process λ, there could be two marked point processes N and N' that both admit λ as their $\mathbb{F}$-intensity, but such that $\mathcal{P}^N \neq \mathcal{P}^{N'}$, see Proposition 9.54.(ii) in Kallenberg [75] for such an example. The fact the we restrict ourselves to the natural filtration $\mathbb{F}^N$ is crucial here.

7.6 Discussion on stationarity

We say that a non-explosive (marked) point process N on $\mathbb{R} \times \mathscr{M}$ is *stationary* if its distribution is invariant with respect to time shifts, that is if $\mathcal{P}^{\theta_t N} = \mathcal{P}^N$ for all $t \in \mathbb{R}$. In addition of being an interesting topic in its own right, stationarity is important for applications. Indeed, stationarity may be required by estimation procedures [12, 77] and is crucial when applying ergodic theorems [41, Section 12.2], [111, Section 2.5].

However, there is no reason to expect a non-explosive marked point process with general intensity functional ψ to admit a stationary version. For example, by contradiction, one can check that there is no stationary point process

N on $\mathbb{R}$ with the intensity functional $\psi(\xi) = \mathbb{1}(\xi((-\infty, 0)) < 1)$, $\xi \in \mathcal{N}^{\#}_{\mathbb{R}}$ (intuitively, the probability of an event occurring has to decrease as time tends to infinity). Instead, it seems that stationarity of marked point processes ought to be studied in a more case-specific manner.

Going back to hybrid marked point processes and considering the above remark, we do not expect them in general to admit stationary versions. Another limitation of stationarity is that, in some applications, the state process $(X_t)_{t\in\mathbb{R}}$ is expected to have some non-stationary components (think of any diffusive quantity like a stock price, say), implying that the corresponding hybrid marked point process N on $\mathbb{R}\times\mathcal{E}\times\mathcal{X}$ cannot be stationary. If however the event functional η does not depend on these non-stationary components, one could try to find a stationary version of $N_{\mathcal{E}}(\cdot) := N(\cdot \times \mathcal{X})$, the marked point process that only keeps track of the event types.

Let us now consider the hybrid marked point processes of Example 6.2.7 (state-dependent Hawkes processes). We will briefly review some main results in the literature on stationarity to see if they can be applied to this special class. First, in the case of Lipschitz intensity functionals ψ, Massoulié [91] constructs a stationary version of the corresponding marked point process using a Picard iteration. Here, this technique cannot be applied since ψ will generally not have the required Lipschitz property as shown in Example 7.7.1 below. Second, Brémaud and Massoulié [20] find a condition for stationarity when the intensity functional ψ has random finite memory, generalising the case of (A, m) processes in Lindvall [86]. Here, this result could be applied when the kernel k has bounded support in the time domain. Third, another approach to stationarity consists in using the theory of Markov processes. Indeed, in some specific cases, the point process can actually be represented as an $\mathbb{R}^d$-valued Markov process, $d \in \mathbb{N}$, which is easier to show to be stationary [79]. Here, as in the case of linear Hawkes processes [48, 42], one could try this approach when each $k(\cdot, m', e)$, $m' \in \mathscr{M}$, $e \in \mathcal{E}$, is of an exponential form. Finally, we note that a marked point process on $\mathbb{R} \times \mathscr{M}$ with an intensity functional ψ can be seen as a Markov process in $\mathcal{N}^{\#}_{\mathbb{R}\times\mathscr{M}}$ [41, p. 429]. However, the study of a stationarity from this general viewpoint does not seem to be tackled in the literature.

7.7 Proofs

In this section, we prove the strong existence result (Theorem 7.4.1) by means of a Poisson embedding lemma given below. Subsequently, we also prove the strong and weak uniqueness results (Theorems 7.5.1 and 7.5.2).

7.7.1 Preliminaries

Example violating the Lipschitz condition

We give here an example of a hybrid marked point process that does not satisfy the Lipschitz condition (7.1), implying that the existence and uniqueness results in Massoulié [91] do not apply.

Example 7.7.1. Set $\mathcal{E} = \{0,1\}$ and $\mathcal{X} = \{0,1\}$ with $\mu_{\mathcal{E}} = \delta_0 + \delta_1$ and $\mu_{\mathcal{X}} = \delta_0 + \delta_1$. Consider an intensity functional ψ that corresponds to a hybrid marked point process with transition function ϕ and event functional η (see Definition 6.2.1). Take η to be a Hawkes functional of the form

$$\eta(e \,|\, \xi) = \nu + \iint_{\mathbb{R}_{<0}\times\mathscr{M}} k(-t', x', e)\xi(dt', de', dx'),$$

where $\nu \in \mathbb{R}_{>0}$ and $k : \mathbb{R}_{>0} \times \mathcal{X} \times \mathcal{E} \to \mathbb{R}_{>0}$ is continuous in time and strictly positive. Let $t_0 \in \mathbb{R}_{<0}$ and choose $\xi_0, \xi_1 \in \mathcal{N}^{\#g}_{\mathbb{R}\times\mathscr{M}}$ such that ξ_0 and ξ_1 coincide on $(-\infty, t_0]$ (i.e., $\theta_{t_0}\xi_0^{\leq 0} = \theta_{t_0}\xi_1^{\leq 0}$) but $F(\xi_0) = 0$ and $F(\xi_1) = 1$ (thus, ξ_0 and ξ_1 do not coincide on $(t_0, 0)$). Assume also that $\phi(0 \,|\, 0, 1) > \phi(0 \,|\, 0, 0)$, $\eta(0 \,|\, \xi_1) < \infty$, and

$$\iint_{(t_0,0)\times\mathscr{M}} k(-t', x', 0)\xi_1(dt', de', dx') > \iint_{(t_0,0)\times\mathscr{M}} k(-t', x', 0)\xi_0(dt', de', dx').$$

Then, following some computations that are left to the reader,

$$\begin{aligned} &|\psi(0, 0 \,|\, \xi_1) - \psi(0, 0 \,|\, \xi_0)| \geq \\ &\qquad (\phi(0 \,|\, 0, 1) - \phi(0 \,|\, 0, 0)) \iint_{(-\infty,t_0]\times\mathscr{M}} k(-t, x, 0)\xi_0(dt, dx). \end{aligned}$$

Next, consider any non-negative kernel $\overline{k} : \mathbb{R}_{>0} \times \mathscr{M} \times \mathscr{M} \to \mathbb{R}_{\geq 0}$. We have that

$$\iint_{\mathbb{R}_{<0} \times \mathscr{M}} \overline{k}(-t, m, 0, 0)|\xi_1 - \xi_0|(dt, dm) = \iint_{(t_0, 0) \times \mathscr{M}} \overline{k}(-t, m, 0, 0)|\xi_1 - \xi_0|(dt, dm).$$

We can now add as many points as necessary to ξ_0 and ξ_1 on $(-\infty, t_0]$ to guarantee that

$$|\psi(0, 0 \,|\, \xi_1) - \psi(0, 0 \,|\, \xi_0)| > \iint_{\mathbb{R}_{<0} \times \mathscr{M}} \overline{k}(-t, m, 0, 0)|\xi_1 - \xi_0|(dt, dm).$$

Consequently, the intensity functional ψ does not satisfy the Lipschitz condition (7.1).

Driving Poisson process

We prove that the mapping $M : \Omega \to \mathcal{N}^{\#}_{\mathbb{R} \times \mathscr{M} \times \mathbb{R}}$ defined by (7.2) is still a Poisson process.

Lemma 7.7.2. *The mapping $M : \Omega \to \mathcal{N}^{\#}_{\mathbb{R} \times \mathscr{M} \times \mathbb{R}}$ is a Poisson process on $\mathbb{R} \times \mathscr{M} \times \mathbb{R}$ with parameter measure $dt\mu_{\mathscr{M}}(dm)dz$. Moreover, M is Poisson relative to $\mathbb{F}$.*

Proof. By composition, using the measurability of $M_{>0}$, it is easy to check that M is a measurable mapping and, thus, it is a non-explosive point process. To show that M is a Poisson process with parameter measure $dt\mu_{\mathscr{M}}(dm)dz$, it is enough notice that, for any $n \in \mathbb{N}$, for every family of bounded sets $(A_i)_{i \in \{1, \dots, n\}}$, for all $k_1, \dots, k_n \in \mathbb{N}$,

$$\mathbb{P}(M(A_i) = k_i,\, i = 1, \dots, n) = \mathbb{P}_{>0}(M_{>0}(A_i) = k_i,\, i = 1, \dots, n),$$

and use the fact that $M_{>0}$ is a Poisson process with parameter measure $dt\mu_{\mathscr{M}}(dm)dz$. To show that $\theta_t M^{\leq 0}$ is $\mathcal{F}_t$-measurable for any $t \in \mathbb{R}$, use the fact that $\theta_t M^{\leq 0}_{>0}$ is $\mathcal{F}^{M_{>0}}_t$-measurable (since a Poisson process is always

Poisson relative to its internal history) along with a composition argument. Similarly, one can show that $\sigma(\theta_t M^{>0}) \subset \{\varnothing, \Omega_{\leq 0}\} \otimes \sigma(\theta_t M^{\geq 0}_{>0})$ and, thus, to show that $\sigma(\theta_t M^{>0})$ is independent of $\mathcal{F}_t$, it is enough to show that $\{\varnothing, \Omega_{\leq 0}\} \otimes \sigma(\theta_t M^{\geq 0}_{>0})$ is independent of $\mathcal{F}_t$. For this, let $A_{\leq 0} \in \{\varnothing, \Omega_{\leq 0}\}$, $A_{>0} \in \sigma(\theta_t M^{\geq 0}_{>0})$, $B_{\leq 0} \in \mathcal{F}_t^{N_{\leq 0}}$ and $B_{>0} \in \mathcal{F}_t^{M_{>0}}$. Then, using the fact that $M_{>0}$ is Poisson relative to $\mathcal{F}^{M_{>0}}$, we have that

$$\begin{aligned}
\mathbb{P}(A_{\leq 0} \times A_{>0} \cap B_{\leq 0} \times B_{>0}) &= \mathbb{P}(A_{\leq 0} \cap B_{\leq 0} \times A_{>0} \cap B_{>0}) \\
&= \mathbb{P}_{\leq 0}(A_{\leq 0} \cap B_{\leq 0}) \mathbb{P}_{>0}(A_{>0} \cap B_{>0}) \\
&= \mathbb{P}_{\leq 0}(A_{\leq 0}) \mathbb{P}_{\leq 0}(B_{\leq 0}) \mathbb{P}_{>0}(A_{>0}) \mathbb{P}_{>0}(B_{>0}) \\
&= \mathbb{P}(A_{\leq 0} \times A_{>0}) \mathbb{P}(B_{\leq 0} \times B_{>0}).
\end{aligned}$$

This shows that two π-systems generating $\{\varnothing, \Omega_{\leq 0}\} \otimes \sigma(\theta_t M^{\geq 0}_{>0})$ and $\mathcal{F}_t^{N_{\leq 0}} \otimes \mathcal{F}_t^{M_{>0}}$, respectively, are independent. We conclude using Lemma 3.6 in Kallenberg [74, p. 50] that $\{\varnothing, \Omega_{\leq 0}\} \otimes \sigma(\theta_t M^{\geq 0}_{>0})$ and $\mathcal{F}_t^{N_{\leq 0}} \otimes \mathcal{F}_t^{M_{>0}}$ are independent. We can then verify that $\{\varnothing, \Omega_{\leq 0}\} \otimes \sigma(\theta_t M^{\geq 0}_{>0})$ remains independent of the completion of $\mathcal{F}_t^{N_{\leq 0}} \otimes \mathcal{F}_t^{M_{>0}}$, which by definition is $\mathcal{F}_t$. Indeed, remember that $\mathcal{F}_t := \sigma(\mathcal{C})$ with $\mathcal{C} := (\mathcal{F}_t^{N_{\leq 0}} \otimes \mathcal{F}_t^{M_{>0}}) \cup \mathcal{A}$ and where $\mathcal{A}$ denotes the class of all subsets of $\mathbb{P}$-null sets in $\mathcal{F}$. It then suffices to notice that $\mathcal{C}$ is a π-system and that $\{\varnothing, \Omega_{\leq 0}\} \otimes \sigma(\theta_t M^{\geq 0}_{>0})$ remains independent of $\mathcal{C}$. $\square$

Poisson-embedding lemma

We are now able to show the following key lemma which demonstrates how the extra-dimension of the Poisson process M allows us to generate a marked point process with a given intensity.

Lemma 7.7.3 (Poisson embedding)**.** *Let $\lambda : \Omega \times \mathbb{R}_{>0} \times \mathscr{M} \to \mathbb{R}_{\geq 0}$ be an $\mathbb{F}$-predictable process. Then, the mapping*

$$\begin{aligned}
N : \quad & \Omega \times \mathcal{B}(\mathbb{R}_{>0} \times \mathscr{M}) \to \mathbb{R}_{\geq 0} \cup \{\infty\} \\
& (\omega, A) \mapsto N(\omega, A) := \iint_A \int_{(0, \lambda(\omega, t, m)]} M(\omega, dt, dm, dz)
\end{aligned} \tag{7.4}$$

is an $\mathbb{F}$-adapted integer-valued random measure on $\mathbb{R}_{>0} \times \mathscr{M}$. Moreover, for every non-negative $\mathbb{F}$-predictable process $H : \Omega \times \mathbb{R}_{>0} \times \mathscr{M} \to \mathbb{R}_{\geq 0}$, we have that

$$\mathbb{E}\left[\iint_{\mathbb{R}_{>0}\times\mathscr{M}} H(t,m)N(dt,dm)\right] = \mathbb{E}\left[\iint_{\mathbb{R}_{>0}\times\mathscr{M}} H(t,m)\lambda(t,m)\mu_{\mathscr{M}}(dm)dt\right].$$

Proof. First, let $A \in \mathcal{B}(\mathbb{R}_{>0} \times \mathscr{M})$ and consider the following composition

$$(\omega, t, m, z) \mapsto (\lambda(\omega, t, m), z) \mapsto \mathbb{1}_{(0,\lambda(\omega,t,m)]}(z)$$

to notice that $\mathbb{1}_{(0,\lambda(\omega,t,m)]}(z)$ is $\mathbb{F}$-predictable by means of Lemma 1.7 and Lemma 1.8 in Kallenberg [74, p. 5]. Then, the product $\mathbb{1}_A(t,m)\mathbb{1}_{(0,\lambda(\omega,t,m)]}(z)$ of two $\mathbb{F}$-predictable processes is also $\mathbb{F}$-predictable by Lemma 1.12 in Kallenberg [74, p. 7]. This ensures that the integral

$$\iint_A \int_{(0,\lambda(\omega,t,m)]} M(\omega, dt, dm, dz) = \iiint_{\mathbb{R}_{>0}\times\mathscr{M}\times\mathbb{R}} \mathbb{1}_A(t,m)\mathbb{1}_{(0,\lambda(\omega,t,m)]}(z)M(\omega, dt, dm, dz)$$

is well defined for all $\omega \in \Omega$ and that $N(\cdot, A)$ is a random variable (see Subsection 5.3).

Second, let $\omega \in \Omega$. For any finite family of disjoint sets $A_1, \ldots, A_n$ in $\mathcal{B}(\mathbb{R}_{>0} \times \mathscr{M})$, $n \in \mathbb{N}$, we clearly have that $N(\omega, \bigcup_{i\leq n} A_i) = \sum_{i\leq n} N(\omega, A_i)$, which means that $N(\omega, \cdot)$ is finitely additive. To prove that $N(\omega, \cdot)$ is countably additive, invoke finite additivity and apply the monotone convergence theorem. These first two steps show that N is indeed a random measure.

Third, to show that N is $\mathbb{F}$-adapted, first consider processes of the form $\lambda : \Omega \times \mathbb{R}_{>0} \times \mathscr{M} \to \mathbb{R}_{\geq 0}$, $\lambda(\omega, t, m) = \mathbb{1}_F(\omega)\mathbb{1}_{(s,u]}(t)\mathbb{1}_C(m)$ where $F \in \mathcal{F}_s$, $s, u \in \mathbb{R}_{>0}$, $s < u$, $C \in \mathcal{B}(\mathscr{M})$. For any $t \in \mathbb{R}_{>0}$, any $A \in \mathcal{B}(\mathbb{R}_{>0})$ such that $A \subset (0, t]$ and any $B \in \mathcal{B}(\mathscr{M})$, we obtain that

$$N(\omega, A \times B) = \mathbb{1}_F(\omega)M(\omega, A \cap (s, u] \times B \cap C \times (0, 1]),$$

which is $\mathcal{F}_t$-measurable since M is $\mathbb{F}$-adapted by Lemma 7.7.2. Hence, N is

$\mathbb{F}$-adapted. To extend this result to any $\mathbb{F}$-predictable process λ, one can use a monotone class argument like in the proof of Proposition 8.7.8 for example.

Fourth, let $\omega \in \Omega$. By the definition of N and by the linearity of the integral, for all simple non-negative functions f on $\mathbb{R}_{>0} \times \mathscr{M}$, we have that

$$\iint_{\mathbb{R}_{>0}\times\mathscr{M}} f(t,m) N(\omega, dt, dm) = \\ \iiint_{\mathbb{R}_{>0}\times\mathscr{M}\times\mathbb{R}} f(t,m) \mathbb{1}_{(0,\lambda(\omega,t,m)]}(z) M(\omega, dt, dm, dz).$$

Then, by Lemma 1.11 in Kallenberg [74, p. 7] and the monotone convergence theorem, we have that the above equality holds for any $\mathcal{B}(\mathbb{R}_{>0} \times \mathscr{M})$-measurable non-negative function f. In particular, we have that, for all $\omega \in \Omega$,

$$\iint_{\mathbb{R}_{>0}\times\mathscr{M}} H(\omega,t,m) N(\omega, dt, dm) = \\ \iiint_{\mathbb{R}_{>0}\times\mathscr{M}\times\mathbb{R}} H(\omega,t,m) \mathbb{1}_{(0,\lambda(\omega,t,m)]}(z) M(\omega, dt, dm, dz).$$

Fifth, using Lemma 7.7.2 and Theorem 5.5.3, we deduce that

$$\begin{aligned}
&\mathbb{E}\left[\iint_{\mathbb{R}_{>0}\times\mathscr{M}} H(t,m) N(dt, dm)\right] \\
&\quad = \mathbb{E}\left[\iiint_{\mathbb{R}_{>0}\times\mathscr{M}\times\mathbb{R}} H(t,m) \mathbb{1}_{(0,\lambda(t,m)]}(z) M(dt, dm, dz)\right] \\
&\quad = \mathbb{E}\left[\iiint_{\mathbb{R}_{>0}\times\mathscr{M}\times\mathbb{R}} H(t,m) \mathbb{1}_{(0,\lambda(t,m)]}(z) dt \mu_{\mathscr{M}}(dm) dz\right] \\
&\quad = \mathbb{E}\left[\iint_{\mathbb{R}_{>0}\times\mathscr{M}} H(t,m) \lambda(t,m) \mu_{\mathscr{M}}(dm) dt\right]. \qquad \square
\end{aligned}$$

Remark 7.7.4. Similar results are given by Brémaud and Massoulié [21, Lemma 3, p. 1571], Massoulié [91, Lemma 1, p. 3] and Torrisi [119, Lemma 2.1, p. 4]. They refer to Lewis and Shedler [83] and Ogata [99] for proofs. The fifth part of our proof follows Daley and Vere-Jones [41, Proposition 14.7.I, p. 427], but we could not find the first four parts anywhere. For the special case $\mathscr{M} = \{0\}$ (i.e., for univariate point processes), a similar proof is given by Çinlar [30, Theorem 6.11, p. 303] while an alternative proof is given by Chevallier et al. [29, Theorem B.11]. Besides, our version of this lemma does not impose any local integrability condition on λ and, thus, does not say if the obtained random measure N is boundedly finite. Finally, note that (7.4) can be rewritten using the compact notation of Massoulié [91] as $N(dt, dm) = M(dt, dm, (0, \lambda(t, m)])$, $t \in \mathbb{R}_{>0}$.

We can now prove the final statement in Thoerem 7.4.1, which we restate here as a corollary.

Corollary 7.7.5. *Let $N : \Omega \to \mathcal{N}^{\#g}_{\mathbb{R}\times\mathscr{M}}$ be a solution to the Poisson-driven SDE (Definition 7.2.1) under either Assumptions A, B, C, or Assumptions A, D, E. Then, N admits ψ as its intensity functional on $\mathbb{R}_{>0}$.*

Proof. Let $G \in \mathcal{F}$ be the almost sure event that (7.3) holds. Consider the following modifications of N and λ, where λ is defined as in (7.3):

$$\tilde{N}(\omega) := N(\omega)\mathbb{1}_G(\omega), \quad \omega \in \Omega, \quad \text{and} \quad \tilde{\lambda}(\omega, t, m) := \lambda(\omega, t, m)\mathbb{1}_G(\omega),$$

for all $\omega \in \Omega$, $t \in \mathbb{R}_{>0}$, $m \in \mathscr{M}$. Then, $\tilde{N}$ and $\tilde{\lambda}$ satisfy (7.4) and, using either Assumptions B.(i) and C or Assumptions D.(i) and E.(ii), one can check that $\tilde{\lambda}(\omega, t, m) < \infty$ for all $\omega \in \Omega$, $t \in \mathbb{R}_{>0}$, $m \in \mathscr{M}$. Moreover, by Lemma 6.4.4, λ is $\mathbb{F}^N$-predictable, and, thus, $\mathbb{F}$-predictable as N is $\mathbb{F}$-adapted. Since the filtration $\mathbb{F}$ is complete, this implies that $\tilde{\lambda}$ is also $\mathbb{F}$-predictable. Now, consider any non-negative $\mathbb{F}^N$-predictable process $H : \Omega \times \mathbb{R}_{>0} \times \mathscr{M} \to \mathbb{R}_{\geq 0}$ and apply Lemma 7.7.3 to obtain

$$\begin{aligned}
\mathbb{E}\left[\iint_{\mathbb{R}_{>0}\times\mathscr{M}} H(t, m)N(dt, dm)\right] &= \mathbb{E}\left[\iint_{\mathbb{R}_{>0}\times\mathscr{M}} H(t, m)\tilde{N}(dt, dm)\right] \\
&= \mathbb{E}\left[\iint_{\mathbb{R}_{>0}\times\mathscr{M}} H(t, m)\tilde{\lambda}(t, m)\mu_{\mathscr{M}}(dm)dt\right] \\
&= \mathbb{E}\left[\iint_{\mathbb{R}_{>0}\times\mathscr{M}} H(t, m)\lambda(t, m)\mu_{\mathscr{M}}(dm)dt\right].
\end{aligned}$$

We conclude that N admits ψ as its intensity functional using Lemma 6.4.2. $\square$

Given a non-explosive point process N on $\mathbb{R} \times \mathscr{M}$ that solves (7.3) or is defined through a Poisson embedding as in Lemma 7.7.3, one can ask when N is in fact a non-explosive marked point process. To this end, it is useful to define the following random measures induced by the driving Poisson process M:

$$L_n(\omega, \cdot) := M(\omega, \cdot \times \mathscr{M} \times (0, n]), \quad \omega \in \Omega, n \in \mathbb{N}.$$

We are then able to find the following sufficient condition on λ.

Lemma 7.7.6 (Simple ground measure)**.** *Let $\lambda : \Omega \times \mathbb{R}_{>0} \times \mathscr{M} \to \mathbb{R}_{\geq 0}$ be an $\mathbb{F}$-predictable process and let N be the $\mathbb{F}$-adapted integer-valued random measure on $\mathbb{R}_{>0} \times \mathscr{M}$ defined by (7.4). Then, if Assumption A holds and if $\sup_{m \in \mathscr{M}} \lambda(t, m) < \infty$ for all $t \in \mathbb{R}_{>0}$, a.s., we have that $N(\{t\} \times \mathscr{M}) \leq 1$ for all $t \in \mathbb{R}_{>0}$, a.s.*

Proof. Each L_n is a Poisson random measure on $\mathbb{R}$ in the sense of Çinlar [30, Chapter 6, p. 249] with boundedly finite parameter measure $n\mu_{\mathscr{M}}(\mathscr{M})dt$. Applying Theorem 2.17 in Çinlar [30, Chapter 6, p. 256] for each $n \in \mathbb{N}$, there exists a set $B \in \mathcal{F}$ such that $\mathbb{P}(B) = 1$ and such that, for all $\omega \in B$ and $n \in \mathbb{N}$, $L_n(\omega) \in \mathcal{N}_{\mathbb{R}}^{\#g}$ (i.e., L_n, $n \in \mathbb{N}$, are simultaneously simple). Next, let A be the almost sure event that $\sup_{m \in \mathscr{M}} \lambda(t, m) < \infty$ for all $t \in \mathbb{R}_{>0}$. Fix $\omega \in A \cap B$ and use the assumption on λ to find that

$$\begin{aligned}
N(\omega, \{t\} \times \mathscr{M}) &= \int_{\{t\}} \int_{\mathscr{M}} \int_{(0, \lambda(\omega, s, m)]} M(\omega, ds, dm, dz) \\
&\leq M\left(\omega, \{t\} \times \mathscr{M} \times \left(0, \sup_{m \in \mathscr{M}} \lambda(\omega, t, m)\right]\right) \\
&\leq M\left(\omega, \{t\} \times \mathscr{M} \times (0, p(\omega, t)]\right) = L_{p(\omega,t)}(\omega, \{t\}) \leq 1,
\end{aligned}$$

where $p(\omega, t) \in \mathbb{N}$ is such that $\sup_{m \in \mathscr{M}} \lambda(\omega, t, m) \leq p(\omega, t)$. □

7.7.2 Strong existence: pathwise construction via Poisson embedding

Existence under Assumptions A, B, C

We begin by proposing a construction of a candidate solution $N : \Omega \to \mathcal{N}^{\infty}_{\mathbb{R}\times\mathscr{M}}$ to (7.3). We proceed in a pathwise fashion. Under Assumption A, using the definition of a Poisson process, it is not difficult to see that, given $n \in \mathbb{N}$, $L_n \in \mathcal{N}^{\#}_{\mathbb{R}}$ a.s. This implies that

$$F_1 := \{\omega \in \Omega \,|\, L_n(\omega) \in \mathcal{N}^{\#}_{\mathbb{R}}, n \in \mathbb{N}\} \in \mathcal{F}$$

is an almost sure event, which plays a key role in our pathwise construction.

Algorithm 7.7.7. Construct the mapping $N : \Omega \to \mathcal{N}^{\infty}_{\mathbb{R}\times\mathscr{M}}$ as follows. For all $\omega = (\omega_{\leq 0}, \omega_{>0}) \in F_1$, set $N_0(\omega) := N_{\leq 0}(\omega_{\leq 0})$, $T_0(\omega) := 0$, $\mathbb{M}_0(\omega) := \varnothing$, and $\lambda_0(\omega, t, m) := \psi(m \,|\, \theta_t N_0(\omega)^{<0})$ for all $t \in \mathbb{R}_{>0}$, $m \in \mathscr{M}$. Define recursively the sequences $(N_n)_{n\in\mathbb{N}}$, $(T_n)_{n\in\mathbb{N}}$, $(\mathbb{M}_n)_{n\in\mathbb{N}}$, and $(\lambda_n)_{n\in\mathbb{N}}$ as follows. For all $n \in \mathbb{N}$,

- if $T_n(\omega) < \infty$, then

$$T_{n+1}(\omega) := \sup\left\{u > T_n(\omega) \;:\; \iint_{(T_n(\omega),u)\times\mathscr{M}} \int_{(0,\lambda_n(\omega,t,m)]} M(\omega, dt, dm, dz) = 0\right\}; \tag{7.5}$$

  - if $T_{n+1}(\omega) < \infty$, then

$$\mathbb{M}_{n+1}(\omega) := \{m \in \mathscr{M} \;:\; M(\omega, \{T_{n+1}(\omega)\} \times \{m\} \times (0, \lambda_n(\omega, T_{n+1}(\omega), m)]) > 0\};$$

$$N_{n+1}(\omega) := \sum_{i=1}^{n+1} \sum_{m\in\mathbb{M}_i(\omega)} M(\omega, \{T_i(\omega)\} \times \{m\}\times (0, \lambda_{i-1}(\omega, T_i(\omega), m)])\delta_{(T_i(\omega),m)}; \tag{7.6}$$

$$\lambda_{n+1}(\omega,t,m) := \psi(m \,|\, \theta_t N_{n+1}(\omega)^{<0}), \quad t \in \mathbb{R}_{>0}, m \in \mathscr{M}\,; \quad (7.7)$$

– if $T_{n+1}(\omega) = \infty$, then

$$\begin{aligned} \mathbb{M}_{n+1}(\omega) &:= \varnothing\,; \\ N_{n+1}(\omega) &:= N_n(\omega)\,; \\ \lambda_{n+1}(\omega,t,m) &:= \lambda_n(\omega,t,m), \quad t \in \mathbb{R}_{>0}, m \in \mathscr{M}\,; \end{aligned}$$

- if $T_n(\omega) = \infty$, then

$$\begin{aligned} T_{n+1}(\omega) &:= \infty\,; \\ \mathbb{M}_{n+1}(\omega) &:= \varnothing\,; \\ N_{n+1}(\omega) &:= N_n(\omega)\,; \\ \lambda_{n+1}(\omega,t,m) &:= \lambda_n(\omega,t,m), \quad t \in \mathbb{R}_{>0}, m \in \mathscr{M}\,. \end{aligned}$$

For all $\omega = (\omega_{\leq 0}, \omega_{>0}) \in \Omega \setminus F_1$, set $N_n(\omega) := N_{\leq 0}(\omega_{\leq 0})$, $T_n(\omega) := \infty$, $\mathbb{M}_n(\omega) := \varnothing$, $\lambda_n(\omega,t,m) := 0$, $t \in \mathbb{R}_{>0}$, $m \in \mathscr{M}$, for all $n \in \mathbb{N}$. Then, for all $\omega \in \Omega$, for all $n \in \mathbb{N}$, define $N(\omega)$ on $(-\infty, T_{n+1}(\omega))$ by

$$\theta_{T_{n+1}(\omega)} N(\omega)^{<0} := \theta_{T_{n+1}(\omega)} N_n(\omega)^{<0}.$$

Define also the explosion time $T_\infty(\omega) := \lim_{n\to\infty} T_n(\omega)$. If $T_\infty(\omega) < \infty$, extend $N(\omega)$ to $[T_\infty(\omega), \infty)$ by $\theta_{T_\infty(\omega)} N(\omega)^{\geq 0} := 0$. This is equivalent to defining $N(\omega)$ as

$$\begin{aligned} &N(\omega) := \lim_{n\to\infty} N_n(\omega) = \\ &\sum_{n=1}^{\infty} \sum_{m \in \mathbb{M}_n(\omega)} M(\omega, \{T_n(\omega)\} \times \{m\} \times \\ &\qquad\qquad (0, \lambda_{n-1}(\omega, T_n(\omega), m)])\,\delta_{(T_n(\omega),m)} \mathbb{1}_{\{T_n(\omega)<\infty\}}. \end{aligned}$$

Algorithm 7.7.7 would be ill-defined if the set in (7.5) were empty. This would mean that there are infinitely many events just after the time T_n. The following proposition shows that this actually never happens and, thus, ensures that Algorithm 7.7.7 is well-defined. We also need to prove that the set $\mathbb{M}_n$ is finite and that $N_n(\omega) \in \mathcal{N}^{\#}_{\mathbb{R}\times\mathscr{M}}$ for all $n \in \mathbb{N}$, because otherwise $\lambda_n(\omega, t, m)$ might be ill-defined (ψ is a functional on $\mathscr{M} \times \mathcal{N}^{\#}_{\mathbb{R}\times\mathscr{M}}$).

Proposition 7.7.8. *In Algorithm 7.7.7, under Assumptions A, B.(i) and C, we have that, for every $\omega \in F_1$, $\mathrm{card}(\mathbb{M}_n(\omega)) < \infty$, $N_n(\omega) \in \mathcal{N}^{\#}_{\mathbb{R}\times\mathscr{M}}$, $\|\lambda_i\|(\omega) := \sup_{t>0,\, m\in\mathscr{M}} \lambda_i(\omega, t, m) < \infty$, for all $n \in \mathbb{N}$, and*

$$\left\{ u > T_n(\omega) \,:\, \iint_{(T_n(\omega),u)\times\mathscr{M}} \int_{(0,\lambda_n(\omega,t,m)]} M(\omega, dt, dm, dz) = 0 \right\} \neq \varnothing$$

for all $n \in \mathbb{N}$ s.t. $T_n(\omega) < \infty$. Hence, Algorithm 7.7.7 is well-defined under these assumptions.

Proof. We show the desired result by induction. Fix $\omega = (\omega_{\leq 0}, \omega_{>0}) \in F_1$. Let $n \in \mathbb{N}$ and for all $i \in \mathbb{N}$ such that $i < n$ and $T_i(\omega) < \infty$, assume that

$$\left\{ u > T_i(\omega) \,:\, \iint_{(T_i(\omega),u)\times\mathscr{M}} \int_{(0,\lambda_i(\omega,t,m)]} M(\omega, dt, dm, dz) = 0 \right\} \neq \varnothing. \quad (7.8)$$

For all $i \in \mathbb{N}$ such that $i \leq n$, assume that $N_i(\omega) \in \mathcal{N}^{\#}_{\mathbb{R}\times\mathscr{M}}$ and that $\|\lambda_i\|(\omega) := \sup_{t>0,\, m\in\mathscr{M}} \lambda_i(\omega, t, m) < \infty$. If $T_n(\omega) = \infty$, then, by construction, this is also true for $n + 1$.

Now, assume that $T_n(\omega) < \infty$. We first show that (7.8) holds also for $i = n$. Take any $\varepsilon > 0$. We have that

$$\begin{aligned}
&\iint_{(T_n(\omega),T_n(\omega)+\varepsilon)\times\mathscr{M}} \int_{(0,\lambda_n(\omega,t,m)]} M(\omega, dt, dm, dz) \\
&\leq M(\omega, (T_n(\omega), T_n(\omega) + \varepsilon) \times \mathscr{M} \times (0, \|\lambda_n\|(\omega)]) \\
&\leq L_{p_n(\omega)}(\omega, (T_n(\omega), T_n(\omega) + \varepsilon)) =: U_n(\omega, \varepsilon) < \infty,
\end{aligned}$$

where $p_n(\omega) \in \mathbb{N}$ is such that $\|\lambda_n\|(\omega) \leq p_n(\omega)$ and we used the fact that $L_{p_n(\omega)}(\omega) \in \mathcal{N}^{\#}_{\mathbb{R}}$. If $U_n(\omega, \varepsilon) = 0$, then clearly (7.8) is satisfied for $i = n$. If not, $M(\omega)$ has a finite number of points in the set

$$(T_n(\omega), T_n(\omega) + \varepsilon) \times \mathscr{M} \times (0, \|\lambda_n\|(\omega)]$$

and there exists $0 < \varepsilon' < \varepsilon$ such that $U_n(\omega, \varepsilon') = 0$, in which case (7.8) is again satisfied for $i = n$. Note that the integral in (7.8) is well-defined since $\lambda_i(\omega, \cdot, \cdot)$ is a measurable function on $\mathbb{R}_{>0} \times \mathscr{M}$ for all $\omega \in \Omega$. To see this, consider the composition $(t, m) \mapsto (m, \theta_t N_i(\omega)^{<0}) \mapsto \psi(m \,|\, \theta_t N_i(\omega)^{<0})$ and use Lemma 8.7.4, the measurability of ψ and Lemma 1.8 in Kallenberg [74].

Second, we show that $\mathrm{card}(\mathbb{M}_{n+1}) < \infty$ and $N_{n+1}(\omega) \in \mathcal{N}^{\#}_{\mathbb{R}\times\mathscr{M}}$. If $T_{n+1}(\omega) = \infty$, then this is immediate. If not, using again that $L_{p_n(\omega)}(\omega) \in \mathcal{N}^{\#}_{\mathbb{R}}$,

$$\begin{aligned}
&\sum_{m \in \mathbb{M}_{n+1}} M(\omega, \{T_{n+1}(\omega)\} \times \{m\} \times (0, \lambda_n(\omega, T_{n+1}(\omega), m)]) \\
&\leq M(\omega, \{T_{n+1}(\omega)\} \times \mathscr{M} \times (0, \|\lambda_n\|(\omega)]) \\
&\leq L_{p_n(\omega)}(\omega, \{T_{n+1}(\omega)\}) < \infty,
\end{aligned}$$

which implies that the set $\mathbb{M}_{n+1}(\omega)$ is finite and, in view of (7.6), that $N_{n+1}(\omega) \in \mathcal{N}^{\#}_{\mathbb{R}\times\mathscr{M}}$. Note that this also proves that $N_{n+1}(\omega, (0, T_{n+1}(\omega)) \times \mathscr{M}) < \infty$.

Third, we show that $\|\lambda_{n+1}\|(\omega) < \infty$. If $T_{n+1}(\omega) = \infty$, then this is immediate. If not, by (7.7) and using Assumptions B.(i) and (7.6), we have that for all $t > 0$, $m = (x, e) \in \mathscr{M}$,

$$\begin{aligned}
\lambda_{n+1}(\omega, t, m) &\leq a\left(N_{n+1}(\omega, (-\infty, t) \times \mathscr{M})\right) \\
&= a\left(N_{n+1}(\omega, (-\infty, 0] \times \mathscr{M}) + N_{n+1}(\omega, (0, t) \times \mathscr{M})\right) \\
&\leq a\left(N_{\leq 0}(\omega_{\leq 0}, (-\infty, 0] \times \mathscr{M}) + N_{n+1}(\omega, (0, T_{n+1}(\omega)) \times \mathscr{M})\right).
\end{aligned} \tag{7.9}$$

Since $N_{n+1}(\omega, (0, T_{n+1}(\omega)) \times \mathscr{M}) < \infty$ and, by Assumption C,

$$N_{\leq 0}(\omega_{\leq 0}, (-\infty, 0] \times \mathscr{M}) < \infty,$$

this implies that $\|\lambda_{n+1}\|(\omega) < \infty$.

Regarding the basis of this induction, it is immediate that $N_0(\omega) = N_{\leq 0}(\omega_{\leq 0}) \in \mathcal{N}^{\#}_{\mathbb{R}\times\mathscr{M}}$. To see that $\|\lambda_0\|(\omega) < \infty$, simply set $n = -1$

in (7.9). $\square$

We show that the constructed mapping $N : \Omega \to \mathcal{N}^{\infty}_{\mathbb{R}\times\mathscr{M}}$ satisfies indeed (7.3) up to each event time.

Proposition 7.7.9. *Under Assumptions A, B.(i) and C, the mapping $N : \Omega \to \mathcal{N}^{\infty}_{\mathbb{R}\times\mathscr{M}}$ given by Algorithm 7.7.7 is such that $N(\omega)$ solves* (7.3) *on $(-\infty, T_n(\omega))$ for all $n \in \mathbb{N}$, for all $\omega \in F_1$.*

Proof. Define the process $\lambda(\omega, t, m) := \psi(m \mid \theta_t N(\omega)^{<0})$, $\omega \in \Omega$, $t \in (0, T_\infty(\omega))$, $m \in \mathscr{M}$. Take any $\omega = (\omega_{\leq 0}, \omega_{>0}) \in F_1$. By construction, we have $N(\omega)^{\leq 0} = N_0(\omega) = N_{\leq 0}(\omega_{\leq 0})$ and, thus, N satisfies the strong initial condition $N_{\leq 0}$. Take any $n \in \mathbb{N}$ such that $T_{n+1}(\omega) < \infty$ in Algorithm 7.7.7 and consider the time interval $(T_n(\omega), T_{n+1}(\omega)]$. By construction, we have that

$$N_{n+1}(\omega, dt, dm) = M(\omega, dt, dm, (0, \lambda_n(\omega, t, m)]) \tag{7.10}$$

for all $t \in (T_n(\omega), T_{n+1}(\omega)]$. But, by definition, on $(-\infty, T_{n+1}(\omega)]$, we have $N(\omega) = N_{n+1}(\omega)$ and thus, for all $t \in (0, T_{n+1}(\omega)]$, $m \in \mathscr{M}$,

$$\begin{aligned}\lambda(\omega, t, m) &= \psi(m \mid \theta_t N(\omega)^{<0}) = \psi(m \mid \theta_t N_{n+1}(\omega)^{<0}) = \psi(m \mid \theta_t N_n(\omega)^{<0}) \\ &= \lambda_n(\omega, t, m),\end{aligned}$$

by the definition (7.7) of λ_n, since $N_{n+1}(\omega)$ and $N_n(\omega)$ can only differ by a mass at time $T_{n+1}(\omega)$. Consequently, (7.10) can be rewritten on the interval $(T_n(\omega), T_{n+1}(\omega)]$ as

$$N(\omega, dt, dm) = M(\omega, dt, dm, (0, \lambda(\omega, t, m)]). \tag{7.11}$$

This shows that the constructed $N(\omega)$ solves (7.3) on $(-\infty, T_n(\omega)]$ for all $n \in \mathbb{N}$ such that $T_n(\omega) < \infty$. Now, if there is $n \in \mathbb{N}$ such that $T_n(\omega) < \infty$ and $T_{n+1}(\omega) = \infty$, then clearly the constructed $N(\omega)$ is null on $(T_n(\omega), \infty)$ and by similar arguments, (7.11) holds on $(T_n(\omega), \infty)$. This now allows us to conclude that $N(\omega)$ solves (7.3) on $(-\infty, T_n(\omega))$ for all $n \in \mathbb{N}$ in both cases $T_n(\omega) < \infty$ and $T_n(\omega) = \infty$. $\square$

It will also be crucial for the strong existence proof to show that, for all $n \in \mathbb{N}$, N_n is adapted to the filtration $\mathbb{F}$ and λ_n is $\mathbb{F}$-predictable.

Proposition 7.7.10. *In Algorithm 7.7.7, for all $n \in \mathbb{N}$, λ_n is $\mathbb{F}$-predictable, N_n is an $\mathbb{F}$-adapted non-explosive point process and T_n is an $\mathbb{F}$-stopping time.*

Proof. We proceed by induction. Regarding the basis, as the filtration $\mathbb{F}$ is complete, clearly N_0 is $\mathbb{F}$-adapted and T_0 is an $\mathbb{F}$-stopping time. Now assume that N_n is $\mathbb{F}$-adapted and T_n is an $\mathbb{F}$-stopping time for some $n \in \mathbb{N}$. First, observe that this implies that λ_n is $\mathbb{F}$-predictable by simply using the identity $\lambda_n(\omega, t, m) = \psi(m \,|\, \theta_t N_n(\omega)^{<0}) \mathbb{1}_{F_1}(\omega)$ and invoking Lemma 6.4.4, the fact that $\mathcal{F}_t^{N_n} \subset \mathcal{F}_t$, $t \in \mathbb{R}$, and the assumption that $\mathbb{F}$ is complete. Second, let $t \in \mathbb{R}$ and notice that

$$\{T_{n+1} \le t\} = \tag{7.12}$$
$$\left\{ \iiint_{\mathbb{R} \times \mathscr{M} \times \mathbb{R}} \mathbb{1}_{(T_n, t]}(s) \mathbb{1}_{(0, \lambda_n(s,m))}(z) M(ds, dm, dz) > 0 \right\} \cap \{T_n \le t\} \cap F_1.$$

Because T_n is an $\mathbb{F}$-stopping time, we have that $(\mathbb{1}_{(T_n,t]}(s))_{s \in \mathbb{R}}$ is $\mathbb{F}$-adapted and left-continuous, implying that it is $\mathbb{F}$-predictable, see for example Lemma 25.1 in Kallenberg [74, p. 491]. Adapting the arguments of the third part of the proof of Lemma 7.7.3, we deduce that the first event on the right-hand side of (7.12) belongs to $\mathcal{F}_t$ and so T_{n+1} is an $\mathbb{F}$-stopping time. Third, using Proposition 7.7.9 and looking at Algorithm 7.7.7, notice that N_{n+1} satisfies

$$\begin{cases} N_{n+1}(\omega, dt, dm) = M(\omega, dt, dm, (0, \lambda_n(\omega, t, m) \mathbb{1}_{\{t \le T_{n+1}(\omega)\}}]), \\ N^{\le 0}(\omega) = N_{\le 0}(\omega_{\le 0}). \end{cases}$$

for all $\omega = (\omega_{\le 0}, \omega_{>0}) \in \Omega$, $t \in \mathbb{R}_{>0}$, where $\lambda_n(t, m) \mathbb{1}_{\{t \le T_{n+1}\}}$ is $\mathbb{F}$-predictable as a product of $\mathbb{F}$-predictable processes, note that $\mathbb{1}_{\{t \le T_{n+1}\}}$ is $\mathbb{F}$-adapted and left-continuous since T_{n+1} is an $\mathbb{F}$-stopping time. Now, applying Lemma 7.7.3, it follows that N_{n+1} is indeed $\mathbb{F}$-adapted. □

We are now in a position to prove Theorem 7.4.1 under Assumption A, B and C for the following intensity functional:

$$\psi' : \mathscr{M} \times \mathcal{N}^{\#}_{\mathbb{R}\times\mathscr{M}} \to \mathbb{R}_{\geq 0} \cup \{\infty\}$$
$$(m, \xi) \mapsto \psi'(m \,|\, \xi) := a\left(\xi((-\infty, 0) \times \mathscr{M})\right).$$

Still, note that the first step of the following proof remains true for general intensity functionals ψ that satisfy Assumption B and will be reused in other parts of the proof of Theorem 7.4.1.

Proof of Theorem 7.4.1, Part 1. Let $N : \Omega \to \mathcal{N}^{\infty}_{\mathbb{R}\times\mathscr{M}}$ be given by Algorithm 7.7.7 under Assumptions A, B and C, which is well-defined by Proposition 7.7.8, and consider here the special case $\psi = \psi'$. We will prove that N admits a version that solves the Poisson-driven SDE. We proceed in four steps.

First, notice that for all $\omega \in \Omega$, $t < T_\infty(\omega)$, there exists $n \in \mathbb{N}$ such that $\theta_t N(\omega)^{<0} = \theta_t N_n(\omega)^{<0}$, which implies by Proposition 7.7.8 that the process

$$\lambda(\omega, t, m) := \psi(m \,|\, \theta_t N(\omega)^{<0}) \mathbb{1}_{F_1}(\omega) \mathbb{1}_{\{t<T_\infty(\omega)\}}, \quad \omega \in \Omega, t \in \mathbb{R}_{>0}, m \in \mathscr{M},$$

is well-defined and finite, and that, for all $\omega \in \Omega$, $t \in \mathbb{R}_{>0}$, $m \in \mathscr{M}$,

$$\begin{aligned}\lambda(\omega, t, m) &= \lim_{n\to\infty} \psi(m \,|\, \theta_t N_n(\omega)^{<0}) \mathbb{1}_{F_1}(\omega) \mathbb{1}_{\{t<T_\infty(\omega)\}} \\ &= \lim_{n\to\infty} \lambda_n(\omega, t, m) \mathbb{1}_{\{t<T_\infty(\omega)\}}.\end{aligned}$$

By Proposition 7.7.9, and because of the way we constructed N, we have that N and λ satisfy (7.4). By Proposition 7.7.10, for all $n \in \mathbb{N}$, λ_n is $\mathbb{F}$-predictable and T_n is an $\mathbb{F}$-stopping time. Since $T_\infty = \lim_{n\to\infty} T_n$, we have that T_∞ is an $\mathbb{F}$-predictable time, which implies by Lemma 25.3.(ii) in Kallenberg [74, p. 492] that $\mathbb{1}_{\{t<T_\infty\}}$ is $\mathbb{F}$-predictable. As λ is a limit of $\mathbb{F}$-predictable processes, we have that λ is also $\mathbb{F}$-predictable by Lemma 1.9 in Kallenberg [74, p. 6]. Consequently, we can apply Lemma 7.7.3 to obtain that N is an $\mathbb{F}$-adapted integer-valued random measure. The main goal of the next steps is to show that $T_\infty = \infty$ a.s.

Second, following Proposition 7.7.8, we can see that $\sup_{m\in\mathscr{M}} \lambda(\omega, t, m) < \infty$ for all $t \in \mathbb{R}_{>0}$, $\omega \in \Omega$. Hence, by Lemma 7.7.6, there exists and almost sure event $G \in \mathcal{F}$ on which $N(\{t\} \times \mathscr{M}\}) \leq 1$ for all $t \in \mathbb{R}$. Let $\tilde{N}$, $(\tilde{N}_n)_{n\in\mathbb{N}}$, $(\tilde{T}_n)_{n\in\mathbb{N}}$ and $\tilde{T}_\infty$ coincide with N, $(N_n)_{n\in\mathbb{N}}$, $(T_n)_{n\in\mathbb{N}}$ and T_∞ on G. Outside G, set $\tilde{N} := 0$, $\tilde{N}_n := 0$, $\tilde{T}_n := \infty$, for all $n \in \mathbb{N}$, and $\tilde{T}_\infty := \infty$. Define the random measures on $\mathbb{R}$

$$\tilde{N}_{\mathscr{M}}(\cdot) := \tilde{N}(\cdot \times \mathscr{M}), \quad \tilde{N}_{\mathscr{M},n}(\cdot) := \tilde{N}_n(\cdot \times \mathscr{M}), \quad n \in \mathbb{N},$$

and define the process

$$\begin{aligned}\tilde{\lambda}(\omega, t) :&= \lim_{n\to\infty} a(\tilde{N}_{\mathscr{M},n}(\omega, (-\infty, t)))\mathbb{1}_{\{t<\tilde{T}_\infty(\omega)\}} \\ &= a(\tilde{N}_{\mathscr{M}}(\omega, (-\infty, t)))\mathbb{1}_{\{t<\tilde{T}_\infty(\omega)\}}, \quad \omega \in \Omega, t \in \mathbb{R}_{>0}.\end{aligned}$$

Since $\{\tilde{T}_n \leq t\} = \{\tilde{N}_{\mathscr{M}}((0, t]) \geq n\}$, $\tilde{T}_n$ is in fact an $\mathbb{F}^{\tilde{N}_{\mathscr{M}}}$-stopping time and, thus, reusing the argument in the first step, we have that $\mathbb{1}_{\{t<\tilde{T}_\infty\}}$ is $\mathbb{F}^{\tilde{N}_{\mathscr{M}}}$-predictable. Moreover, by Lemma 6.4.4, we have that $(a(\tilde{N}_{\mathscr{M},n}((-\infty, t))))_{t>0}$ is $\mathbb{F}^{\tilde{N}_{\mathscr{M},n}}$-predictable and, thus, $\mathbb{F}^{\tilde{N}_{\mathscr{M}}}$-predictable. Hence, using again Lemma 1.9 in Kallenberg [74, p. 6], we have that $\tilde{\lambda}$ is also $\mathbb{F}^{\tilde{N}_{\mathscr{M}}}$-predictable. Next, because $\tilde{N} = N$ a.s. and $\tilde{\lambda}(t) = \lambda(t, m)$ for all $t \in \mathbb{R}_{>0}$, $m \in \mathscr{M}$, a.s., and because Lemma 7.7.3 applies to N and λ, we have that, for any non-negative $\mathbb{F}^{\tilde{N}_{\mathscr{M}}}$-predictable process $H : \Omega \times \mathbb{R}_{>0} \to \mathbb{R}_{\geq 0}$,

$$\begin{aligned}\mathbb{E}\left[\int_{\mathbb{R}_{>0}} H(t)\tilde{N}_{\mathscr{M}}(dt)\right] &= \mathbb{E}\left[\iint_{\mathbb{R}_{>0}\times\mathscr{M}} H(t)N(dt, dm)\right] \\ &= \mathbb{E}\left[\iint_{\mathbb{R}_{>0}\times\mathscr{M}} H(t)\lambda(t, m)\mu_{\mathscr{M}}(dm)dt\right] \\ &= \mathbb{E}\left[\int_{\mathbb{R}_{>0}} H(t)\tilde{\lambda}(t)\mu_{\mathscr{M}}(\mathscr{M})dt\right].\end{aligned}$$

Consequently, $\tilde{N}_{\mathscr{M}}$, or equivalently $(\tilde{T}_n)_{n\in\mathbb{N}}$, defines a simple point process on $\mathbb{R}_{>0}$ with $\mathbb{F}^{\tilde{N}_{\mathscr{M}}}$-predictable projection $(\mu_{\mathscr{M}}(\mathscr{M}) \int_0^t \tilde{\lambda}(s)ds)_{t>0}$ in the sense of Jacod [68].

Third, by Lemma 6.4.1, $\mathcal{F}_t^{\tilde{N}_{\mathscr{M}}} = \mathcal{F}_0^{\tilde{N}_{\mathscr{M}}} \vee \mathcal{F}_t^{\tilde{N}_{\mathscr{M}}^{>0}}$, and, thus, Assumption A.1 of Jacod [68] holds, see also the proof of Theorem 7.5.2 and Remark 7.7.14. Then, by Proposition 3.1 in Jacod [68], we have that, conditional on

$N_{\leq 0}((-\infty, 0]) = n_0 \in \mathbb{N}$, $S_n := \tilde{T}_{n+1} - \tilde{T}_n$, $n \in \mathbb{N}$, follows an exponential distribution with parameter $a(n+n_0)\mu_{\mathscr{M}}(\mathscr{M})$ and the $(S_n)_{n\in\mathbb{N}}$ are independent. Thanks to Assumption B.(ii), by Example 3.1.4 in Jacobsen [67, p.20], we deduce that, conditional on $N_{\leq 0}((-\infty, 0]) = n_0$, $\tilde{T}_\infty = \lim_{n\to\infty} \tilde{T}_n = \infty$ a.s., see also Proposition 12.19 in Kallenberg [74, p.240]. Consequently, it holds that $\tilde{T}_\infty = \infty$ a.s. unconditionally.

Fourth, following these first three steps, we have proved that there exists a version of N such that $N \in \mathcal{N}^{\#g}_{\mathbb{R}\times\mathscr{M}}$, i.e., this version of N is a non-explosive marked point process, see Proposition 5.5.1, and such that N solves the Poisson-driven SDE. We conclude by Corollary 7.7.5. $\square$

To prove Theorem 7.4.1 under Assumptions A, B and C in the general case, we will use a solution to the special case $\psi = \psi'$ to show that the constructed mapping $N : \Omega \to \mathcal{N}^{\infty}_{\mathbb{R}\times\mathscr{M}}$ actually takes values in $\mathcal{N}^{\#g}_{\mathbb{R}\times\mathscr{M}}$. First, we need to define what we mean for a marked point process N to be dominated by another marked point process $\overline{N}$.

Definition 7.7.11. Let $\xi, \overline{\xi} \in \mathcal{N}^{\infty}_{\mathbb{R}\times\mathscr{M}}$. We say that ξ is *dominated* by $\overline{\xi}$ and write $\xi \prec \overline{\xi}$ if, for all $A \in \mathcal{B}(\mathbb{R}\times\mathscr{M})$, $\xi(A) \leq \overline{\xi}(A)$. Let $T \in \mathbb{R}$. We say that ξ is dominated by $\overline{\xi}$ on $(-\infty, T]$ if $\theta_T \xi^{\leq 0} \prec \theta_T \overline{\xi}^{\leq 0}$. Consider two mappings $N : \Omega \to \mathcal{N}^{\infty}_{\mathbb{R}\times\mathscr{M}}$ and $\overline{N} : \Omega \to \mathcal{N}^{\infty}_{\mathbb{R}\times\mathscr{M}}$. We say that N is *dominated* by $\overline{N}$ if $N \prec \overline{N}$ a.s.

When $N \prec \overline{N}$ a.s., one could also say that N is a thinning of $\overline{N}$. Indeed, notice that $\xi \prec \overline{\xi}$ implies that all the atoms of ξ are also atoms of $\overline{\xi}$.

We will now show that the constructed mapping $N : \Omega \to \mathcal{N}^{\infty}_{\mathbb{R}\times\mathscr{M}}$ is dominated by any solution to the special case $\psi = \psi'$.

Proposition 7.7.12. *Let $N' : \Omega \to \mathcal{N}^{\#g}_{\mathbb{R}\times\mathscr{M}}$ be a solution to the Poisson-driven SDE with intensity functional ψ'. Then, under Assumptions A, B.(i) and C, the mapping $N : \Omega \to \mathcal{N}^{\infty}_{\mathbb{R}\times\mathscr{M}}$ obtained from Algorithm 7.7.7 satisfies $N \prec N'$ a.s.*

Proof. Fix $\omega = (\omega_{\leq 0}, \omega_{>0}) \in A \cap F_1$, where $A \in \mathcal{F}$ is the almost sure event that N' solves (7.3), where ψ is replaced by ψ'. Clearly, $N(\omega) \prec N'(\omega)$

on $(-\infty, 0]$. Now take any $n \in \mathbb{N}$ such that $T_n(\omega) < \infty$ and assume that $N(\omega) \prec N'(\omega)$ on $(-\infty, T_n(\omega)]$. If $T_{n+1}(\omega) = \infty$, then $N(\omega)$ is null on (T_n, ∞) and we have $N(\omega) \prec N'(\omega)$. If $T_{n+1}(\omega) < \infty$, we have that for all $t \in (T_n(\omega), T_{n+1}(\omega)]$, $m \in \mathscr{M}$,

$$\begin{aligned}
\lambda(\omega, t, m) &= \psi(m \,|\, \theta_t N(\omega)^{<0}) \\
\text{(by construction)} \quad &= \psi(m \,|\, \theta_t N_n(\omega)^{<0}) \\
\text{(by Assumption B.(i))} \quad &\leq \psi'(m \,|\, \theta_t N_n(\omega)^{<0}) \\
&\leq \psi'(m \,|\, \theta_t N'(\omega)^{<0}) =: \lambda'(\omega, t, m),
\end{aligned}$$

where the last inequality holds by the definition of ψ', Assumption B and since $N_n(\omega) \prec N'(\omega)$. By Proposition 7.7.9 for $N(\omega)$ and by assumption for $N'(\omega)$, we have that $N(\omega)$ and $N'(\omega)$ both satisfy (7.3) on $(-\infty, T_{n+1}(\omega)]$, where λ is replaced by λ' for $N'(\omega)$. Consequently, we have $N(\omega) \prec N'(\omega)$ on $(-\infty, T_{n+1}(\omega)]$. As, by construction, $N(\omega)$ has mass on $\mathbb{R}_{>0}$ only at the times $T_1(\omega) < T_2(\omega) < \ldots < \infty$, we have shown that $N(\omega) \prec N'(\omega)$. This implies that $N \prec N'$ a.s. $\square$

This allows us to conclude the proof of Theorem 7.4.1 under Assumptions A, B and C.

Proof of Theorem 7.4.1, Part 2. Repeat the first step of Part 1. Then, by Proposition 7.7.12, we deduce that $N \in \mathcal{N}^{\#g}_{\mathbb{R}\times\mathscr{M}}$ and $T_\infty = \infty$ a.s. We then conclude by repeating the fourth step of Part 1. $\square$

Existence under Assumptions A, D, E

To prove Theorem 7.4.1 under Assumptions A, D and E, we will also use Algorithm 7.7.7 to construct a candidate solution, but the almost sure event F_1 needs to be replaced by another almost sure event F_2 that guarantees that the algorithm is well-defined under these new assumptions. Whereas under Assumptions B and C, we were able to first construct the candidate solution and then dominate it by a solution to the special case $\psi = \psi'$, here we will dominate the candidate solution while constructing it. The dominating non-explosive marked point process is nothing else than a solution to the Poisson-driven SDE with the Hawkes intensity functional

$$\overline{\psi} : \mathscr{M} \times \mathcal{N}^{\#}_{\mathbb{R}\times\mathscr{M}} \to \mathbb{R}_{>0} \cup \{\infty\}$$
$$(m, \xi) \mapsto \overline{\psi}(m \mid \xi) := \lambda_0 + \iint_{(-\infty,0)\times\mathscr{M}} \overline{k}(-t', m', m)\xi(dt', dm'),$$

where λ_0 and $\overline{k}$ are as in Assumption D. Indeed, by applying the results of Massoulié [91] and Lemma 7.7.6, we can prove Theorem 7.4.1 under Assumptions A, D and E for the special case $\psi = \overline{\psi}$.

Proof of Theorem 7.4.1, Part 3. Clearly, $\overline{\psi}$ satisfies the Lipschitz condition (7.1) with the kernel $\overline{k}$. Under Assumptions A, D.(ii) and E.(i), by Theorem 2 of Massoulié [91], we know that there exists a non-explosive point process $\overline{N} : \Omega \to \mathcal{N}^{\#}_{\mathbb{R}\times\mathscr{M}}$ that solves (7.3), where ψ is replaced by $\overline{\psi}$, and such that $\overline{N}(\cdot \times \mathscr{M}) \in \mathcal{N}^{\#}_{\mathbb{R}}$. Moreover, applying Assumptions D.(iii) and E.(ii), we obtain that

$$\begin{aligned}
\overline{\lambda}(\omega, t, m) &:= \overline{\psi}(m \mid \theta_t \overline{N}(\omega)^{<0}) \\
&= \lambda_0 + \iint_{(-\infty,t)\times\mathscr{M}} \overline{k}(t - t', m', m)\overline{N}(\omega, dt', dm') \\
&\leq \lambda_0 + \tilde{\lambda}_{\leq 0}(\omega_{\leq 0}, t) + \iint_{(0,t)\times\mathscr{M}} \sup_{m''\in\mathscr{M}} \overline{k}(t - t', m', m'')\overline{N}(\omega, dt', dm') \\
&< \infty, \quad \omega \in \Omega, t \in \mathbb{R}_{>0}, m \in \mathscr{M},
\end{aligned}$$

which proves that $\sup_{m\in\mathscr{M}} \overline{\lambda}(t, m) < \infty$, for all $t \in \mathbb{R}_{>0}$, a.s. Hence, by Lemma 7.7.6, we conclude that $\overline{N}$ admits a version such that $\overline{N}(\omega) \in \mathcal{N}^{\#g}_{\mathbb{R}\times\mathscr{M}}$, $\omega \in \Omega$, meaning that this version solves the Poisson-driven SDE. Conclude by Corollary 7.7.5. □

From now on, denote by $\overline{N}$ a solution to the Poisson-driven SDE in the special case $\psi = \overline{\psi}$ and by $F_2 \in \mathcal{F}$ the almost sure event that $\overline{N}$ solves (7.3), where ψ is replaced by $\overline{\psi}$. The following statement is the analogue of Proposition 7.7.8 and ensures that Algorithm 7.7.7 is well-defined under this different set of assumptions.

Proposition 7.7.13. *In Algorithm 7.7.7, where F_1 is replaced by F_2, under Assumptions A, D and E, we have that, for every $\omega \in F_2$, $\operatorname{card}(\mathbb{M}_n(\omega)) \leq 1$, $N_n(\omega) \prec \overline{N}(\omega)$, for all $n \in \mathbb{N}$, and*

$$\left\{ u > T_n(\omega) \,:\, \iint_{(T_n(\omega),u)\times\mathscr{M}} \int_{(0,\lambda_n(\omega,t,m)]} M(\omega, dt, dm, dz) = 0 \right\} \neq \varnothing$$

for all $n \in \mathbb{N}$ s.t. $T_n(\omega) < \infty$. Hence, Algorithm 7.7.7 is well-defined under these assumptions.

Proof. We show the assertion by induction. Take any $\omega = (\omega_{\leq 0}, \omega_{>0}) \in F_2$. Let $n \in \mathbb{N}$ and, for all $i \in \mathbb{N}$ such that $i < n$ and $T_i(\omega) < \infty$, assume that

$$\left\{ u > T_i(\omega) \,:\, \iint_{(T_i(\omega),u)\times\mathscr{M}} \int_{(0,\lambda_i(\omega,t,m)]} M(\omega, dt, dm, dz) = 0 \right\} \neq \varnothing. \quad (7.13)$$

For all $i \in \mathbb{N}$ such that $i \leq n$, assume that $N_i(\omega) \prec \overline{N}$. If $T_n(\omega) = \infty$, then, by construction, this is also true for $n + 1$.

Now, assume that $T_n(\omega) < \infty$. We first show that (7.13) holds also for $i = n$. By adapting the proof of Proposition 7.7.12 and using Assumption D.(i), we get that $\lambda_n(\omega, t, m) \leq \overline{\lambda}(\omega, t, m)$, for all $t > T_n(\omega)$, $m \in \mathscr{M}$. Hence, for any $\varepsilon > 0$, we have that

$$\begin{aligned} &\iint_{(T_n(\omega),T_n(\omega)+\varepsilon)\times\mathscr{M}} \int_{(0,\lambda_n(\omega,t,m)]} M(\omega, dt, dm, dz) \leq \\ &\quad \iint_{(T_n(\omega),T_n(\omega)+\varepsilon)\times\mathscr{M}} \int_{(0,\overline{\lambda}(\omega,t,m)]} M(\omega, dt, dm, dz) \\ &\quad = \overline{N}(\omega, (T_n(\omega), T_n(\omega) + \varepsilon) \times \mathscr{M}) =: \overline{U}_n(\omega, \varepsilon) < \infty, \end{aligned}$$

since $\overline{N}(\omega, \cdot \times \mathscr{M}) \in \mathcal{N}^{\#}_{\mathbb{R}}$. If $\overline{U}_n(\omega, \varepsilon) = 0$, then clearly (7.13) is satisfied for $i = n$. If not, $\overline{N}(\omega, \cdot \times \mathscr{M})$ has a finite number of points in $(T_n(\omega), T_n(\omega) + \varepsilon)$ and there exists $0 < \varepsilon' < \varepsilon$ such that $\overline{U}_n(\omega, \varepsilon') = 0$, in which case (7.13) is again satisfied for $i = n$.

Second, we show that $\operatorname{card}(\mathbb{M}_{n+1}) \leq 1$ and $N_{n+1}(\omega) \prec \overline{N}(\omega)$. If $T_{n+1}(\omega) = \infty$, then this is immediate. If not, as $\lambda_n(\omega, t, m) \leq \overline{\lambda}(\omega, t, m)$, for all $t > T_n(\omega)$, it is enough to notice that

$$\begin{aligned}
&\mathbb{M}_{n+1}(\omega) := \\
&\qquad \{m \in \mathscr{M} \,:\, M(\omega, \{T_{n+1}(\omega)\} \times \{m\} \times \ (0, \lambda_n(\omega, T_{n+1}(\omega), m)]) > 0\} \\
&\qquad \subset \big\{m \in \mathscr{M} \,:\, M(\omega, \{T_{n+1}(\omega)\} \times \{m\} \times \ (0, \overline{\lambda}(\omega, T_{n+1}(\omega), m)]) > 0\big\} \\
&\qquad = \big\{m \in \mathscr{M} \,:\, \overline{N}(\omega, \{T_{n+1}(\omega)\} \times \{m\}) > 0\big\} \leq 1,
\end{aligned}$$

since $\overline{N}(\omega) \in \mathcal{N}^{\#g}_{\mathbb{R}\times\mathscr{M}}$. As we already know that $N_n(\omega) \prec \overline{N}(\omega)$, looking at (7.6) and observing that $N_{n+1}(\omega)$ and $N_n(\omega)$ only differ by a mass at time $T_{n+1}(\omega)$, we further deduce that $N_{n+1}(\omega) \prec \overline{N}(\omega)$.

Regarding the basis of this induction, it is immediate that $N_0(\omega) \prec \overline{N}(\omega)$ since $N_0(\omega) = N_{\leq 0}(\omega_{\leq 0}) = \overline{N}^{\leq 0}(\omega)$. □

We are now in a position to finish the proof of Theorem 7.4.1 under Assumption A, D and E.

Proof of Theorem 7.4.1, Part 4. Let $N : \Omega \to \mathcal{N}^{\infty}_{\mathbb{R}\times\mathscr{M}}$ be given by Algorithm 7.7.7 under Assumptions A, D and E, where F_1 is replaced by F_2. By Proposition 7.7.13, this mapping is well defined. Moreover, we notice that Propositions 7.7.9 and 7.7.10 still hold under the present assumptions. Hence, we can repeat the first step of Part 1 of the proof. By Proposition 7.7.13, we know that $N_n(\omega) \prec \overline{N}(\omega)$ for all $\omega \in F_2$, which implies by construction that $N(\omega) \prec \overline{N}(\omega)$ on $\mathbb{R}_{>0}$ for all $\omega \in \Omega$. Consequently, we have that $N(\omega) \in \mathcal{N}^{\#g}_{\mathbb{R}\times\mathscr{M}}$ and $T_\infty(\omega) = \infty$ for all $\omega \in \Omega$, which implies that N solves the Poisson-driven SDE. We conclude again by Corollary 7.7.5. □

7.7.3 Strong and weak uniqueness

Proof of Theorem 7.5.1. Let $\tilde{\Omega} \in \mathcal{F}$ be the almost sure event that both N and N' solve (7.3). Let $(T_n, M_n)_{n\in\mathbb{N}}$ and $(T'_n, M'_n)_{n\in\mathbb{N}}$ be the enumerations in $(0, \infty] \times \mathscr{M}$ to which N and N' are respectively equivalent. Now fix arbitrary $\omega \in \tilde{\Omega}$. We show by strong induction that $T_n(\omega) = T'_n(\omega)$ and $M_n(\omega) = M'_n(\omega)$ for all $n \in \mathbb{N}$.

Let $n \in \mathbb{N}$ and assume that $T_i(\omega) = T'_i(\omega)$ and $M_i(\omega) = M'_i(\omega)$ for all $i = 1, \ldots, n-1$. By contradiction, assume that $T_n(\omega) \neq T'_n(\omega)$ and, moreover, without loss of generality, that $T_n(\omega) < T'_n(\omega)$. Then, this implies that

$$N(\omega, (0, T_n(\omega)] \times \mathscr{M}) = \int_{(0,T_n(\omega)]} \int_{\mathscr{M}} \int_{(0,\lambda(\omega,t,m)]} M(\omega, dt, dm, dz) = n,$$
$$N'(\omega, (0, T_n(\omega)] \times \mathscr{M}) = \int_{(0,T_n(\omega)]} \int_{\mathscr{M}} \int_{(0,\lambda'(\omega,t,m)]} M(\omega, dt, dm, dz) = n-1,$$

where $\lambda(\omega, t, m) = \psi(m \,|\, \theta_t N(\omega)^{<0})$ and $\lambda'(\omega, t, m) = \psi(m \,|\, \theta_t N'(\omega)^{<0})$. But since $N(\omega)^{\leq 0} = N'(\omega)^{\leq 0}$ and also $T_i(\omega) = T'_i(\omega)$ and $M_i(\omega) = M'_i(\omega)$ for all $i = 1, \ldots, n-1$, we have that $\theta_t N(\omega)^{<0} = \theta_t N'(\omega)^{<0}$ for all $t \leq T_n(\omega)$ and, thus, $\lambda(\omega, t, m) = \lambda'(\omega, t, m)$ for all $t \leq T_n(\omega), m \in \mathscr{M}$. This implies that $n = n-1$ which is a contradiction and, thus, necessarily, $T_n(\omega) = T'_n(\omega)$.

Similarly, if we assume that $M_n(\omega) \neq M'_n(\omega)$, then this implies that

$$\begin{aligned} N(\omega, \{T_n(\omega)\} \times \{M_n(\omega)\}) &= \int_{\{T_n(\omega)\}} \int_{\{M_n(\omega)\}} \int_{(0,\lambda(\omega,t,m)]} M(\omega, dt, dm, dz) \\ &= 1, \\ N'(\omega, \{T_n(\omega)\} \times \{M_n(\omega)\}) &= \int_{\{T_n(\omega)\}} \int_{\{M_n(\omega)\}} \int_{(0,\lambda'(\omega,t,m)]} M(\omega, dt, dm, dz) \\ &= 0. \end{aligned}$$

But again, since $\lambda(\omega, t, m) = \lambda'(\omega, t, m)$ for all $t \leq T_n(\omega), m \in \mathscr{M}$, this leads to the contradiction $1 = 0$ and, thus, it follows that $M_n(\omega) = M'_n(\omega)$. The same reasoning allows us to prove the basis of the strong induction (i.e., to show that $T_1(\omega) = T'_1(\omega)$ and $M_1(\omega) = M'_1(\omega)$). □

Proof of Theorem 7.5.2. Consider the canonical measurable space $(\mathcal{N}^{\#g}_{\mathbb{R}\times\mathscr{M}}, \mathcal{B}(\mathcal{N}^{\#g}_{\mathbb{R}\times\mathscr{M}}))$, where $\mathcal{B}(\mathcal{N}^{\#g}_{\mathbb{R}\times\mathscr{M}}) = \mathcal{N}^{\#g}_{\mathbb{R}\times\mathscr{M}} \cap \mathcal{B}(\mathcal{N}^{\#}_{\mathbb{R}\times\mathscr{M}})$, and the canonical non-explosive marked point process N defined by $N(\omega) = \omega$, $\omega \in \mathcal{N}^{\#g}_{\mathbb{R}\times\mathscr{M}}$. Under both $\mathcal{P}^{N_1}$ and $\mathcal{P}^{N_2}$ (they only charge $\mathcal{N}^{\#g}_{\mathbb{R}\times\mathscr{M}}$), N satisfies the weak initial condition $N_{\leq 0}$ and admits an intensity given by (5.2). We will now apply Theorem 3.4 in Jacod [68, p. 242] to show that $\mathcal{P}^{N_1} = \mathcal{P}^{N_2}$. By Lemma

6.4.1, we have that

$$\mathcal{F}_t^N = \mathcal{F}_0^N \vee \mathcal{F}_t^{\theta_0 N^{>0}} = \mathcal{F}_0^{N^{\leq 0}} \vee \mathcal{F}_t^{N^{>0}}, \quad t \in \mathbb{R}_{\geq 0},$$

and thus, Assumption A.1 of Jacod [68] is satisfied (see Remark 7.7.14).

To apply Theorem 3.4 in Jacod [68], it remains to verify that the restrictions of $\mathcal{P}^{N_1}$ and $\mathcal{P}^{N_2}$ coincide on $\mathcal{F}_0^N$. Note that $\mathcal{F}_0^N$ is generated by the π-system $\mathcal{C}$ of sets of the form

$$\{N \in \mathcal{N}_{\mathbb{R}\times\mathscr{M}}^{\#g} : N(A_1 \times M_1) \geq n_1, \ldots, N(A_k \times M_k) \geq n_k\}, \\ n_1, \ldots, n_k \in \mathbb{N},\ k \in \mathbb{N},$$

where $A_1, \ldots, A_k, \in \mathcal{B}(\mathbb{R}_{\leq 0})$ and $M_1, \ldots, M_k \in \mathcal{B}(\mathscr{M})$. For any such set $F \in \mathcal{C}$, setting $B_i := A_i \times M_i$ for $i = 1, \ldots, k$ and invoking the fact that both N_1 and N_2 satisfy the weak initial condition $N_{\leq 0}$, we deduce that

$$\begin{aligned}
\mathcal{P}^{N_1}(F) &= \mathcal{P}^{N_1}(N(B_1) \geq n_1, \ldots, N(B_k) \geq n_k) \\
&= \mathcal{P}^{N_1}(N^{\leq 0}(B_1) \geq n_1, \ldots, N^{\leq 0}(B_k) \geq n_k) \\
&= \mathcal{P}^{N_{\leq 0}}(N^{\leq 0}(B_1) \geq n_1, \ldots, N^{\leq 0}(B_k) \geq n_k) \\
&= \mathcal{P}^{N_2}(N^{\leq 0}(B_1) \geq n_1, \ldots, N^{\leq 0}(B_k) \geq n_k) \\
&= \mathcal{P}^{N_2}(N(B_1) \geq n_1, \ldots, N(B_k) \geq n_k) = \mathcal{P}^{N_2}(F).
\end{aligned}$$

Hence, $\mathcal{P}^{N_1}$ and $\mathcal{P}^{N_2}$ coincide on $\mathcal{C}$, a π-system that contains $\mathcal{N}_{\mathbb{R}\times\mathscr{M}}^{\#g}$. As a consequence, $\mathcal{P}^{N_1} = \mathcal{P}^{N_2}$ on $\mathcal{F}_0^N$, see for example Lemma 1.17 in Kallenberg [74, p. 9], and we can apply Theorem 3.4 in Jacod [68, p. 242] to deduce that $\mathcal{P}^{N_1} = \mathcal{P}^{N_2}$ on $(\mathcal{N}_{\mathbb{R}\times\mathscr{M}}^{\#g}, \mathcal{B}(\mathcal{N}_{\mathbb{R}\times\mathscr{M}}^{\#g}))$. □

Remark 7.7.14. Let us clarify the relationship between our notations and those in Jacod [68]. Our canonical measurable space $(\mathcal{N}_{\mathbb{R}\times\mathscr{M}}^{\#g}, \mathcal{B}(\mathcal{N}_{\mathbb{R}\times\mathscr{M}}^{\#g}))$ plays the role of his measurable space $(\Omega, \mathcal{F}_\infty)$. Our marked point process $N^{>0}$ corresponds to his marked point process μ. Our probability measures $\mathcal{P}^{N_1}$ and $\mathcal{P}^{N_2}$ are the counterparts of P and P', respectively. Our filtrations $\mathbb{F}^N$ and $\mathbb{F}^{N^{>0}}$ correspond to his filtrations $(\mathcal{F}_t)_{t\geq 0}$ and $(\mathcal{G}_t)_{t\geq 0}$, respectively.

8

On the weak-hash metric

As already mentioned in Chapter 5, it is known that the space of boundedly finite integer-valued measures on a complete separable metric space becomes itself a complete separable metric space when endowed with the weak-hash metric. It is also known that convergence under this topology can be characterised in a way that is similar to the weak convergence of totally finite measures. However, the original proofs of these two fundamental results assume that a certain term is monotonic, which is not always the case as we show by a counterexample. We manage to clarify these original proofs by addressing the parts that rely on this assumption and finding alternative arguments.

The results of Chapters 6 and 7 relied on some lemmas and propositions that derive from these fundamental properties of the weak-hash metric and that we prove in the last section of this chapter.

8.1 Notations and problem

Let $\mathcal{U}$ be a complete separable metric space and $x_0 \in \mathcal{U}$ be a fixed origin. Let $B_r(x)$ denote the open ball with radius $r \in \mathbb{R}_{\geq 0}$ and centre $x \in \mathcal{U}$; write

$B_r := B_r(x_0)$ for the open balls centred at x_0. For any subset $A \subset \mathcal{U}$ and $\varepsilon \in \mathbb{R}_{>0}$, the ε-neighbourhood of A is defined by $A^\varepsilon := \bigcup_{a \in A} B_\varepsilon(a)$, the boundary of A is denoted by ∂A and the closure of A is denoted by $\overline{A}$. For any Borel measure ξ on $\mathcal{U}$ and any $r \in \mathbb{R}_{\geq 0}$, we use the notation $\xi^{(r)}$ to refer to the restriction of ξ to the open ball B_r, that is $\xi^{(r)}(A) = \xi(A \cap B_r)$ for all $A \in \mathcal{B}(\mathcal{U})$. A Borel measure ξ on $\mathcal{U}$ is called totally finite if $\xi(\mathcal{U}) < \infty$. Let $\mathcal{M}_\mathcal{U}$ denote the space of totally finite measures on $\mathcal{U}$, and define the Prohorov distance d on $\mathcal{M}_\mathcal{U}$ by

$$\begin{aligned} d\, : \mathcal{M}_\mathcal{U} \times \mathcal{M}_\mathcal{U} &\to \mathbb{R}_{\geq 0} \\ (\mu, \nu) &\mapsto d(\mu, \nu) := \inf\{\varepsilon \in \mathbb{R}_{\geq 0} \,:\, \mu(A) \leq \nu(A^\varepsilon) + \varepsilon \text{ and} \\ &\qquad\qquad \nu(A) \leq \mu(A^\varepsilon) + \varepsilon, \text{for all closed } A \subset \mathcal{U}\}. \end{aligned}$$

It is known that d makes $\mathcal{M}_\mathcal{U}$ a complete separable metric space, see for example Section A2.5 in Daley and Vere-Jones [40, p. 398–402].

In this chapter, we are interested in boundedly finite integer-valued measures. Recall that a Borel measure ξ on $\mathcal{U}$ is called boundedly finite if $\xi(A) < \infty$ for all bounded Borel sets $A \in \mathcal{B}(\mathcal{U})$ and that we denote by $\mathcal{N}_\mathcal{U}^\#$ the space of boundedly finite measures on $\mathcal{U}$ with values in $\mathbb{N} \cup \{\infty\}$. Note that such measures are always atomic (i.e., a superposition of Dirac measures), see for example Proposition 9.1.III.(ii) in Daley and Vere-Jones [41, p. 4]. One might ask if the Prohorov distance d on the space $\mathcal{M}_\mathcal{U}$ has a counterpart on the space $\mathcal{N}_\mathcal{U}^\#$. Daley and Vere-Jones [40, p. 403] tackle this question by considering the distance function

$$\begin{aligned} d^\#\, : \mathcal{N}_\mathcal{U}^\# \times \mathcal{N}_\mathcal{U}^\# &\to \mathbb{R}_{\geq 0} \\ (\mu, \nu) &\mapsto d^\#(\mu, \nu) := \int_0^\infty e^{-r} \frac{d(\mu^{(r)}, \nu^{(r)})}{1 + d(\mu^{(r)}, \nu^{(r)})} dr. \end{aligned} \tag{8.1}$$

The core idea is to use the Prohorov metric on the restrictions to the open balls and compute a weighted average. They name the corresponding topology the $w^\#$-topology ("weak-hash") and refer to $d^\#$ as the $w^\#$-distance. They then obtain the following two fundamental results. The first one is a characterisation of convergence under this metric.

Theorem 8.1.1 (Characterisation of convergence)**.** *Let $(\mu_k)_{k\in\mathbb{N}}$ be a sequence in $\mathcal{N}^{\#}_{\mathcal{U}}$ and $\mu \in \mathcal{N}^{\#}_{\mathcal{U}}$. Then, the following statements are equivalent:*

(i) $d^{\#}(\mu_k, \mu) \to 0$ *as* $k \to \infty$;

(ii) $\int_{\mathcal{U}} f(x)\mu_k(dx) \to \int_{\mathcal{U}} f(x)\mu(dx)$ *as* $k \to \infty$ *for all bounded continuous functions* f *on* $\mathcal{U}$ *vanishing outside a bounded set;*

(iii) *there exists an increasing sequence* $(r_n)_{n\in\mathbb{N}}$ *with* $r_n \to \infty$ *as* $n \to \infty$ *such that*
$d(\mu_k^{(r_n)}, \mu^{(r_n)}) \to 0$ *as* $k \to \infty$ *for all* $n \in \mathbb{N}$;

(iv) $\mu_k(A) \to \mu(A)$ *as* $k \to \infty$ *for all bounded sets* $A \in \mathcal{B}(\mathcal{U})$ *such that* $\mu(\partial A) = 0$.

The second result confirms that $d^{\#}$ is indeed the counterpart of d, that is $\mathcal{N}^{\#}_{\mathcal{U}}$ inherits the completeness and separability properties of $\mathcal{U}$ under the metric $d^{\#}$. This second result also provides us with a characterisation of the Borel σ-algebra $\mathcal{B}(\mathcal{N}^{\#}_{\mathcal{U}})$.

Theorem 8.1.2 (Metric properties of $\mathcal{N}^{\#}_{\mathcal{U}}$)**.**

(i) *The space* $\mathcal{N}^{\#}_{\mathcal{U}}$ *is a complete separable metric space when it is equipped with the distance function* $d^{\#}$.

(ii) *The corresponding Borel* σ*-algebra* $\mathcal{B}(\mathcal{N}^{\#}_{\mathcal{U}})$ *is the smallest* σ*-algebra that makes all mappings* $\Phi_A : \mathcal{N}^{\#}_{\mathcal{U}} \to \mathbb{N}\cup\{\infty\}$, $A \in \mathcal{B}(\mathcal{U})$, *measurable, where* $\Phi_A(\xi) = \xi(A)$.

Theorem 8.1.1 and Theorem 8.1.2 in this chapter are Proposition A2.6.II and Theorem A2.6.III in Daley and Vere-Jones [40, p. 403–405], respectively.

Regarding the motivation of this chapter, we have seen that the metric space $(\mathcal{N}^{\#}_{\mathcal{U}}, d^{\#})$ is a stepping stone to the theory of point processes as presented by Daley and Vere-Jones [41], who define a point process as a random element in $\mathcal{N}^{\#}_{\mathcal{U}}$. Moreover, the above theorems are crucial in our framework and proofs, in Chapters 5, 6 and 7.

We now turn to the precise purpose of this chapter. To argue that the integrand in (8.1) is measurable and prove the above properties of the metric $d^{\#}$, Daley and Vere-Jones [40, p. 403–405] assume that $d(\mu^{(r)}, \nu^{(r)})$ is non-decreasing as a function of $r \in \mathbb{R}_{\geq 0}$. Example 8.1.3 below shows that this need not necessarily be true.

Example 8.1.3. Set $\mathcal{U} = \mathbb{R}$, $x_0 = 0$, $\mu = \delta_0$ and $\nu = \delta_{0.5}$, where, for any $x \in \mathcal{U}$, δ_x denotes the Dirac measure at x. Then, as long as $r < 0.5$, $d(\mu^{(r)}, \nu^{(r)}) = 1$. However, as soon as $r > 0.5$, $d(\mu^{(r)}, \nu^{(r)}) = 0.5$.

Consequently, our goal is to clarify the original proofs of Theorems 8.1.1 and 8.1.2 given in Daley and Vere-Jones [40] by addressing specifically the parts that rely on the assumed monotonicity of $d(\mu^{(r)}, \nu^{(r)})$. Note that Daley and Vere-Jones [40] consider the larger space $\mathcal{M}_{\mathcal{U}}^{\#}$ of boundedly finite measures, i.e., not necessarily integer-valued. The proofs we develop here (except in Section 8.3) are specialised to the subspace $\mathcal{N}_{\mathcal{U}}^{\#}$ and take advantage of the discrete nature of its elements. Besides, we should add that an alternative metrization of $\mathcal{M}_{\mathcal{U}}^{\#}$, leading to the same properties, is presented in Kallenberg [75, Section 4.1, p. 111–117]. According to Kallenberg [75, Historical and bibliographical notes, p. 638], this extension from totally finite measures to boundedly finite measures under this alternative metric was first developed by Matthes et al. [92].

The chapter is organised as follows. Section 8.2 gives some preliminary results on the Prohorov metric. Section 8.3 shows that the distance function in (8.1) is well defined. Section 8.4 deals with the proof of Theorem 8.1.1. Sections 8.5 and 8.6 address the proof of Theorem 8.1.2.

Remark 8.1.4. We would like to stress that this chapter focuses on the parts of the original proofs that assume that $r \mapsto d(\mu^{(r)}, \nu^{(r)})$ is non-decreasing (with the exception of Section 8.6). Our main objective is to find alternative arguments for these parts specifically. To understand the proofs of Theorems 8.1.1 and 8.1.2 in their entirety and the details of the other parts that are not treated here, we refer the reader to the original text [40, p. 403–405].

As the Prohorov metric d is the main building block of the $w^{\#}$-distance $d^{\#}$, it is not surprising that we need to study its behaviour. In particular, we will apply the following lemmas.

Lemma 8.2.1. *Let $\mu \in \mathcal{M}^{\#}_{\mathcal{X}}$ and $p, r \in \mathbb{R}_{\geq 0}$ be such that $p \leq r$. Then $d(\mu^{(p)}, \mu^{(r)}) \leq \mu(B_r \setminus B_p)$.*

Proof. Let $\varepsilon > \mu(B_r \setminus B_p)$. Let $F \in \mathcal{B}(\mathcal{U})$ be a closed set. Then, clearly

$$\mu^{(p)}(F) = \mu(F \cap B_p) \leq \mu(F^{\varepsilon} \cap B_r) + \varepsilon = \mu^{(r)}(F^{\varepsilon}) + \varepsilon.$$

Moreover, we have that

$$\begin{aligned} \mu^{(r)}(F) &= \mu(F \cap B_p) + \mu(F \cap B_r \setminus B_p) \\ &\leq \mu^{(p)}(F) + \mu(B_r \setminus B_p) \\ &\leq \mu^{(p)}(F^{\varepsilon}) + \varepsilon. \end{aligned}$$

This means exactly that $d(\mu^{(p)}, \mu^{(r)}) \leq \mu(B_r \setminus B_p)$ by definition of the Prohorov distance d. □

Lemma 8.2.2. *Let $\mu, \nu \in \mathcal{N}^{\#}_{\mathcal{U}}$ be such that $\mu(\mathcal{U}) < \infty$, $\nu(\mathcal{U}) < \infty$. Let $\underline{r}, \bar{r}, \varepsilon \in \mathbb{R}_{>0}$ be such that $\underline{r} < \bar{r}$ and $\varepsilon < (\bar{r} - \underline{r})/2 < 1$. If $\mu(B_{\bar{r}} \setminus B_{\underline{r}}) = 0$ and $\nu(B_{\bar{r}-\varepsilon} \setminus B_{\underline{r}+\varepsilon}) > 0$, then $d(\mu, \nu) \geq \varepsilon$.*

Proof. Let $0 \leq \delta < \varepsilon$ and $u \in B_{\bar{r}-\varepsilon} \setminus B_{\underline{r}+\varepsilon}$ be such that $\nu(\{u\}) \geq 1$. Then,

$$\nu(\{u\}) \geq 1 > \delta = \mu(\{u\}^{\delta}) + \delta,$$

which implies that $d(\mu, \nu) \geq \delta$ by definition of the Prohorov distance. Consequently, $d(\mu, \nu) \geq \varepsilon$. □

Lemma 8.2.3.
Let $r \in \mathbb{R}_{\geq 0}$ and $\mu, \nu \in \mathcal{N}^{\#}_{\mathcal{U}}$. Then $d(\mu^{(r)}, \nu^{(r)}) \geq |\mu(B_r) - \nu(B_r)|$.

Proof. Without loss of generality, we can assume that $\mu(B_r) > \nu(B_r)$. Let $\varepsilon \in [0, \mu(B_r) - \nu(B_r))$ and let $\delta \in [0, \mu(B_r) - \nu(B_r) - \varepsilon)$. By Proposition A2.2.II in Daley and Vere-Jones [40, p. 386], there exists a closed set $F \subset B_r$ such that $\mu^{(r)}(B_r \setminus F) < \delta$. Then,

$$\begin{aligned}\mu^{(r)}(F) = \mu^{(r)}(B_r) - \mu^{(r)}(B_r \setminus F) > \mu^{(r)}(B_r) - \delta \\ > \mu^{(r)}(B_r) + \varepsilon + \nu(B_r) - \mu(B_r) \geq \nu^{(r)}(F^{\varepsilon}) + \varepsilon.\end{aligned}$$

Again, this implies that $d(\mu^{(r)}, \nu^{(r)}) \geq |\mu(B_r) - \nu(B_r)|$ by definition of the Prohorov distance. $\square$

8.3 The metric $d^{\#}$ is well defined

In this section, we address the proof in Daley and Vere-Jones [40, p. 403] that shows that $d^{\#}$ is indeed a well-defined metric. We have to check that the integral in (8.1) is well defined and, in particular, that $r \mapsto d(\mu^{(r)}, \nu^{(r)})$ is measurable. To achieve this, it suffices to notice that this function is actually piecewise constant since μ and ν are atomic with finitely many atoms in any bounded set. In fact, for any $R \in \mathbb{R}_{>0}$, as r goes from 0 to R, the restricted measures $\mu^{(r)}$ and $\nu^{(r)}$ change only a finite number of times and so does $d(\mu^{(r)}, \nu^{(r)})$. The other arguments in Daley and Vere-Jones [40, p. 403] are then enough to obtain that $d^{\#}$ satisfies all the conditions of a distance function.

As a side note, for the general case where $\mu, \nu \in \mathcal{M}^{\#}_{\mathcal{U}}$, we can prove that $r \mapsto d(\mu^{(r)}, \nu^{(r)})$ is measurable by showing that it is of finite variation.

Proposition 8.3.1. *Let $\mu, \nu \in \mathcal{M}^{\#}_{\mathcal{X}}$ and $R \in \mathbb{R}_{\geq 0}$. Then, as a function of $r \in \mathbb{R}_{\geq 0}$, the variation of $d(\mu^{(r)}, \nu^{(r)})$ over $[0, R]$ is bounded by $\mu(S_R) + \nu(S_R)$. In particular, $r \mapsto d(\mu^{(r)}, \nu^{(r)})$ is of bounded variation and, thus, measurable.*

Proof. Let $r \in \mathbb{R}_{\geq 0}$ and $\delta > 0$. Applying the triangle inequality to the Prohorov distance, we obtain the following two inequalities:

$$d(\mu^{(r+\delta)}, \nu^{(r+\delta)}) \leq d(\mu^{(r+\delta)}, \mu^{(r)}) + d(\mu^{(r)}, \nu^{(r)}) + d(\nu^{(r)}, \nu^{(r+\delta)})\,;$$

$$d(\mu^{(r)}, \nu^{(r)}) \le d(\mu^{(r)}, \mu^{(r+\delta)}) + d(\mu^{(r+\delta)}, \nu^{(r+\delta)}) + d(\nu^{(r+\delta)}, \nu^{(r)}).$$

This implies that

$$|d(\mu^{(r+\delta)}, \nu^{(r+\delta)}) - d(\mu^{(r)}, \nu^{(r)})| \le d(\mu^{(r)}, \mu^{(r+\delta)}) + d(\nu^{(r)}, \nu^{(r+\delta)}).$$

Using Lemma 8.2.1, we can go further and conclude that

$$|d(\mu^{(r+\delta)}, \nu^{(r+\delta)}) - d(\mu^{(r)}, \nu^{(r)})| \le \mu(S_{r+\delta}) - \mu(S_r) + \nu(S_{r+\delta}) - \nu(S_r).$$

Since $\mu(S_r)$ and $\nu(S_r)$ are non-decreasing in r and always finite (because μ and ν are boundedly finite), they are of bounded variation, which concludes the proof. □

8.4 Characterisation of convergence in the $w^{\#}$-topology

In this section, we address the proof of Theorem 8.1.1, which characterises the convergence of boundedly finite integer-valued measures.

Proof of Theorem 8.1.1. We need to show only the implication (i) $\Longrightarrow$ (iii) because this is the only part in Daley and Vere-Jones [40, p. 403] which relies on the assumption that $d(\mu^{(r)}, \nu^{(r)})$ is non-decreasing in $r \in \mathbb{R}_{\ge 0}$. The rest of the proof of Proposition A2.6.II in Daley and Vere-Jones [40, p. 403-404] can be used to show that (iii) $\Longrightarrow$ (ii) $\Longrightarrow$ (iv) $\Longrightarrow$ (i).

Let $n \in \mathbb{N}$ and $\underline{r}_n, \bar{r}_n \in \mathbb{R}_{\ge 0}$ be such that $n < \underline{r}_n < \bar{r}_n < n+1$ and $\mu(B_{\bar{r}_n} \setminus B_{\underline{r}_n}) = 0$. Let $0 < \varepsilon < (\bar{r}_n - \underline{r}_n)/2$. By contradiction, assume that for any $K \in \mathbb{N}$, there exists $k > K$ such that $\mu_k(B_{\bar{r}_n - \varepsilon} \setminus B_{\underline{r}_n + \varepsilon}) \ge 1$. Then, by Lemma 8.2.2, there exists a subsequence $(k_p)_{p \in \mathbb{N}}$ such that $d(\mu_{k_p}^{(r)}, \mu^{(r)}) \ge \varepsilon$ for all $r \ge n+1$, $p \in \mathbb{N}$. Thus, along this subsequence, we must have

$$\begin{aligned} d^{\#}(\mu_{k_p}, \mu) &= \int_0^\infty e^{-r} \frac{d(\mu_{k_p}^{(r)}, \mu^{(r)})}{1 + d(\mu_{k_p}^{(r)}, \mu^{(r)})} dr \\ &\ge \int_{n+1}^\infty e^{-r} \frac{\varepsilon}{1+\varepsilon} dr = \frac{\varepsilon}{1+\varepsilon} e^{-n-1} > 0, \end{aligned}$$

which contradicts the assumption that $d^{\#}(\mu_k, \mu) \to 0$ as $k \to \infty$. Consequently, there exists a $K \in \mathbb{N}$ such that, for all $k \geq K$, $\mu_k(S_{\bar{r}_n-\varepsilon} \setminus S_{\underline{r}_n+\varepsilon}) = 0$. This means that, for all $k \geq K$, neither μ_k nor μ can have any atom in $S_{\bar{r}_n-\varepsilon} \setminus S_{\underline{r}_n+\varepsilon}$, whence there is a constant $d_k \in \mathbb{R}_{\geq 0}$ such that $d(\mu_k^{(r)}, \mu^{(r)}) = d_k$ for all $r \in (\underline{r}_n + \varepsilon, \bar{r}_n - \varepsilon)$. This implies that, for all $k \geq K$,

$$\begin{aligned} d^{\#}(\mu_k, \mu) &= \int_0^\infty e^{-r} \frac{d(\mu_k^{(r)}, \mu^{(r)})}{1 + d(\mu_k^{(r)}, \mu^{(r)})} dr \geq \int_{\underline{r}_n+\varepsilon}^{\bar{r}_n-\varepsilon} e^{-r} \frac{d_k}{1+d_k} dr \\ &\geq \frac{d_k}{1+d_k} e^{-\underline{r}_n-\varepsilon}(1 - e^{-(\bar{r}_n-\underline{r}_n-2\varepsilon)}), \end{aligned}$$

and, thus, $d_k \to 0$ as $k \to \infty$. If we set $r_n = (\underline{r}_n + \bar{r}_n)/2$, we finally have that $d(\mu_k^{(r_n)}, \mu^{(r_n)}) \to 0$ as $k \to \infty$. □

8.5 Completeness and separability of $\mathcal{N}_{\mathcal{U}}^{\#}$

In this section, we address the proof of the first part of Theorem 8.1.2, which states that $\mathcal{N}_{\mathcal{U}}^{\#}$ is complete and separable when it is endowed with the $w^{\#}$-metric $d^{\#}$.

8.5.1 Completeness

To begin with, we show that if a sequence $(\mu_k)_{k\in\mathbb{N}}$ in $(\mathcal{N}_{\mathcal{U}}^{\#}, d^{\#})$ is Cauchy, then the restrictions along an increasing sequence of balls are also Cauchy for the Prohorov metric d.

Proposition 8.5.1. *Let $(\mu_k)_{k\in\mathbb{N}}$ be a Cauchy sequence in $\mathcal{N}_{\mathcal{U}}^{\#}$ for the $w^{\#}$-metric $d^{\#}$. Then, there exists an increasing sequence $(r_n)_{n\in\mathbb{N}}$ in $\mathbb{R}_{>0}$ with $r_n \to \infty$ as $n \to \infty$ such that, for each $n \in \mathbb{N}$, $(\mu_k^{(r_n)})_{k\in\mathbb{N}}$ is a Cauchy sequence in $\mathcal{M}_{\mathcal{U}}$ for the Prohorov metric d.*

Proof. **Step 1.** We show that $\mu_k(B_r)$ is bounded in $k \in \mathbb{N}$ for all $r \in \mathbb{R}_{\geq 0}$. By contradiction, assume that this is not the case. Then, there exists a subsequence such that $\mu_{k_p}(B_r) \to \infty$. Along this subsequence, for p large enough and any fixed $q \in \mathbb{N}$, we have that

$$\begin{aligned}\int_r^{r+1} e^{-s} \frac{d(\mu_{k_p}^{(s)}, \mu_{k_q}^{(s)})}{1 + d(\mu_{k_p}^{(s)}, \mu_{k_q}^{(s)})} ds &\geq \int_r^{r+1} e^{-s} \frac{|\mu_{k_p}(B_s) - \mu_{k_q}(B_s)|}{1 + |\mu_{k_p}(B_s) - \mu_{k_q}(B_s)|} ds \\ &\geq \int_r^{r+1} e^{-s} \frac{\mu_{k_p}(B_r) - \mu_{k_q}(B_{r+1})}{1 + \mu_{k_p}(B_r) - \mu_{k_q}(B_{r+1})} ds \\ &\to e^{-r}(1 - e^{-1}), \quad p \to \infty,\end{aligned}$$

where we used Lemma 8.2.3 and the fact that $\mu_{k_p}(B_s)$ and $\mu_{k_q}(B_s)$ are non-decreasing in s. But this is incompatible with the Cauchy assumption on $(\mu_k)_{k \in \mathbb{N}}$. Indeed, let $\varepsilon < e^{-r}(1 - e^{-1})$. Then, the Cauchy assumption implies that there exists $K \in \mathbb{N}$ such that, for all $k, k' \geq K$,

$$d^{\#}(\mu_k, \mu_{k'}) = \int_0^{\infty} e^{-s} \frac{d(\mu_k^{(s)}, \mu_{k'}^{(s)})}{1 + d(\mu_k^{(s)}, \mu_{k'}^{(s)})} ds \leq \varepsilon.$$

But then, for $p, q \in \mathbb{N}$ large enough, we must have

$$\varepsilon \geq \int_0^{\infty} e^{-s} \frac{d(\mu_{k_p}^{(s)}, \mu_{k_q}^{(s)})}{1 + d(\mu_{k_p}^{(s)}, \mu_{k_q}^{(s)})} ds \geq \int_r^{r+1} e^{-s} \frac{d(\mu_{k_p}^{(s)}, \mu_{k_q}^{(s)})}{1 + d(\mu_{k_p}^{(s)}, \mu_{k_q}^{(s)})} ds > \varepsilon.$$

Step 2. Let $n \in \mathbb{N}$. We show that for $k, p \in \mathbb{N}$ large enough, there is a subinterval of $[n, n+1]$ on which the functions $r \mapsto d(\mu_k^{(r)}, \mu_p^{(r)})$ are constant. Define $M := \sup_{k \in \mathbb{N}} \mu_k(B_{n+1})$, which is finite by the first step and can be understood as a bound on the number of points in the ball B_{n+1} among all measures μ_k. Let $\varepsilon_1, \varepsilon_2 \in \mathbb{R}_{>0}$ be such that $\varepsilon_1 < \varepsilon_2 < 1/2(M+1)$ and $\varepsilon_1 < \varepsilon_2 e^{-n-1}/(1 + \varepsilon_2)$. Let $K \in \mathbb{N}$ be such that, for all $k, k' \geq K$, $d^{\#}(\mu_k, \mu_{k'}) \leq \varepsilon_1$ (Cauchy assumption). Since $\mu_K(B_{n+1} \setminus B_n) \leq M$, we can find $\underline{r}_n, \overline{r}_n \in (n, n+1)$ such that $\mu_K(B_{\overline{r}_n} \setminus B_{\underline{r}_n}) = 0$ and $\overline{r}_n - \underline{r}_n \geq 1/(M+1)$. Now, by contradiction, assume that $\mu_p(B_{\overline{r}_n - \varepsilon_2} \setminus B_{\underline{r}_n + \varepsilon_2}) \geq 1$ for some $p > K$. Then, using Lemma 8.2.2, we obtain that

$$\varepsilon_1 \geq d^{\#}(\mu_K, \mu_p) = \int_0^{\infty} e^{-r} \frac{d(\mu_K^{(r)}, \mu_p^{(r)})}{1 + d(\mu_K^{(r)}, \mu_p^{(r)})} dr \geq \int_{n+1}^{\infty} e^{-r} \frac{d(\mu_K^{(r)}, \mu_p^{(r)})}{1 + d(\mu_K^{(r)}, \mu_p^{(r)})} dr$$

$$\geq \frac{\varepsilon_2}{1+\varepsilon_2} e^{-n-1},$$

which contradicts the original assumption on ε_1 and ε_2. Consequently, for all $k \geq K$, $\mu_k(B_{\overline{r}_n-\varepsilon_2} \setminus B_{\underline{r}_n+\varepsilon_2}) = 0$, which implies that $r \mapsto d(\mu_p^{(r)}, \mu_q^{(r)})$ is constant on $(\underline{r}_n + \varepsilon_2, \overline{r}_n - \varepsilon_2)$ for all $p, q \geq K$.

Step 3. We finally show that when $r_n =: (\underline{r}_n + \overline{r}_n)/2$, $(\mu_k^{(r_n)})_{k\in\mathbb{N}}$ is a Cauchy sequence for the Prohorov metric d. Let $\varepsilon > 0$ and set

$$\delta := (\overline{r}_n - \underline{r}_n - 2\varepsilon_2)e^{-n-1}\frac{\varepsilon}{1+\varepsilon}.$$

Let $J \in \mathbb{N}$ be such that, for all $p, q \geq J$, $d^{\#}(\mu_k, \mu_{k'}) \leq \delta$ (Cauchy assumption). Then, for all $p, q \geq K \vee J$, we must have

$$\delta \geq \int_{\underline{r}_n+\varepsilon_2}^{\overline{r}_n-\varepsilon_2} e^{-r}\frac{d_{pq}}{1+d_{pq}}dr \geq \frac{d_{pq}}{1+d_{pq}}(\overline{r}_n - \underline{r}_n - 2\varepsilon_2)e^{-n-1},$$

where $d_{pq} := d(\mu_p^{(r_n)}, \mu_q^{(r_n)})$, and which implies

$$d_{pq} \leq \frac{\delta}{(\overline{r}_n - \underline{r}_n - 2\varepsilon_2)e^{-n-1} - \delta} = \frac{1}{\frac{1+\varepsilon}{\varepsilon} - 1} = \varepsilon. \qquad \square$$

Reusing a part of the proof of Theorem A2.6.III in Daley and Vere-Jones [40, p. 404], the above proposition implies that $\mathcal{N}_{\mathcal{U}}^{\#}$ is complete. Still, we would like to mention some points that could deserve a bit more detail. First, one needs to ensure that the limit of each Cauchy sequence $(\mu_k^{(r_n)})_{k\in\mathbb{N}}$ in Proposition 8.5.1 is still integer-valued. This can be done by adapting the proof of Lemma 9.1.V in Daley and Vere-Jones [41, p. 6]. Second, if we denote the limit of $(\mu_k^{(r_n)})_{k\in\mathbb{N}}$ by ν_n, we can show that $\nu_m^{(r_n)} = \nu_n$ when $n < m$ (i.e., the sequence of measures $(\nu_n)_{n\in\mathbb{N}}$ is consistent) by using Theorem A2.3.II.(iv) in Daley and Vere-Jones [40, p. 391] and the fact that $\nu_m(\partial B_{r_n}) = 0$. Third, to show that $\mu(\cdot) := \lim_{n\to\infty} \nu_n(\cdot)$ is continuous from below, one can use the fact that $\lim_{i\to\infty}\lim_{j\to\infty} a_{ij} = \lim_{j\to\infty}\lim_{i\to\infty} a_{ij}$ for any double sequence (a_{ij}) that is non-decreasing in both i and j.

8.5.2 Separability

Next, we prove that the space of boundedly finite integer-valued measures $\mathcal{N}_{\mathcal{U}}^{\#}$ is separable. We wish to show that there exists a countable set in $\mathcal{N}_{\mathcal{U}}^{\#}$ that can approximate well-enough any element of $\mathcal{N}_{\mathcal{U}}^{\#}$. Let $\mathcal{D}_{\mathcal{U}}$ be the separability set of $\mathcal{U}$. It seems natural to expect that the set of totally finite (hence with a finite number of atoms) integer-valued measures with atoms only in $\mathcal{D}_{\mathcal{U}}$ is a good candidate. We denote this set by $\mathcal{D}_{\mathcal{N}}$. Notice that one can define an injection between $\mathcal{D}_{\mathcal{N}}$ and the finite subsets of $\mathbb{N}^2$. For example, the Dirac measure with mass $n \in \mathbb{N}$ at the m^{th} element of $\mathcal{D}_{\mathcal{U}}$ can be represented by the set $\{(m,n)\}$. Since the finite subsets of a countable set form a countable set, we have that $\mathcal{D}_{\mathcal{N}}$ is countable. The following proposition coupled with a part of the proof of Theorem A2.6.III in Daley and Vere-Jones [40, p. 404] allows to conclude.

Proposition 8.5.2. *Let $\mu \in \mathcal{N}_{\mathcal{U}}^{\#}$ and $R, \varepsilon \in \mathbb{R}_{>0}$. Then, there exists $\tilde{\mu} \in \mathcal{D}_{\mathcal{N}}$ such that*

$$\int_0^R e^{-r} \frac{d(\mu^{(r)}, \tilde{\mu}^{(r)})}{1 + d(\mu^{(r)}, \tilde{\mu}^{(r)})} dr \leq \varepsilon.$$

Proof. Let $(u_n)_{n\in\{1,\ldots,N\}}$ be the atoms of μ in B_R where $N \in \mathbb{N}$ is their total number and let $(w_n)_{n\in\{1,\ldots,N\}}$ be their corresponding weights. Let $\varepsilon_1 > 0$ be such that $B_{\varepsilon_1}(u_n) \subset B_R$ for all $n = 1, \ldots, N$. Let $0 \leq r_1 < \ldots < r_{N'} < R$ be the radii at which the atoms are located where $N' \in \mathbb{N}$, $N' \leq N$ ($r_1 = 0$ means that $x_0 \in (u_n)_{n\in\{1,\ldots,N\}}$). Define $\varepsilon_2 := \frac{1}{2}\min_{n<N'}(r_{n+1} - r_n)$ and $\varepsilon_3 := \varepsilon/4N'$. Define $\varepsilon_4 := \varepsilon/(2c - \varepsilon)$, where $c = 1 - e^{-R}$, and assume that $\varepsilon < 2c$ (if this is not the case, then the desired inequality already holds no matter what $\tilde{\mu}$). Finally, set $\delta := \min(\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4)$ and let $(\tilde{u}_n)_{n\in\{1,\ldots,N\}}$ be such that $\tilde{u}_n \in \mathcal{D}_{\mathcal{U}}$, $\tilde{u}_n \in B_\delta(u_n)$, $n = 1, \ldots, N$. We will show that $\tilde{\mu} := \sum_{n=1}^N w_n \delta_{\tilde{u}_n}$ satisfies the desired inequality.

Let $n = 1, \ldots, N' - 1$ and $r \in (r_n + \delta, r_{n+1} - \delta)$. We can check that $d(\mu^{(r)}, \tilde{\mu}^{(r)}) \leq \delta$. Indeed, since $\delta \leq \varepsilon_1$ and $\delta \leq \varepsilon_2$, we have that $u_i \in B_r$ if and only if $\tilde{u}_i \in B_r$. Consequently, for any closed set $A \in \mathcal{B}(\mathcal{U}) \cap B_r$, using the fact that $\tilde{u}_i \in B_\delta(u_i)$,

$$\begin{aligned} \mu^{(r)}(A) = \mu(A) \le \tilde{\mu}(A^\delta \cap B_r) = \tilde{\mu}^{(r)}(A^\delta) \quad \text{and} \\ \tilde{\mu}^{(r)}(A) = \tilde{\mu}(A) \le \mu(A^\delta \cap B_r) = \mu^{(r)}(A^\delta), \end{aligned}$$

which means that $d(\mu^{(r)}, \tilde{\mu}^{(r)}) \le \delta$. Similarly, $d(\mu^{(r)}, \tilde{\mu}^{(r)}) \le \delta$ holds for all $r \in [0, 0 \vee (r_1 - \delta))$ and all $r \in (r_{N'} + \delta, R]$. Using this bound on the Prohorov distance between the restrictions, we obtain that

$$\begin{aligned} &\int_0^R \frac{e^{-r} d(\mu^{(r)}, \tilde{\mu}^{(r)})}{1 + d(\mu^{(r)}, \tilde{\mu}^{(r)})} dr = \\ &\left(\int_0^{0 \vee (r_1 - \delta)} + \sum_{n=1}^{N'} \int_{(r_n - \delta) \vee 0}^{r_n + \delta} + \sum_{n=1}^{N'-1} \int_{r_n + \delta}^{r_{n+1} - \delta} + \int_{r_{N'} + \delta}^{R} \right) \frac{e^{-r} d(\mu^{(r)}, \tilde{\mu}^{(r)})}{1 + d(\mu^{(r)}, \tilde{\mu}^{(r)})} dr \\ &\le \int_0^R e^{-r} \frac{\delta}{1 + \delta} dr + \sum_{n=1}^{N'} \int_{(r_n - \delta) \vee 0}^{r_n + \delta} e^{-r} \frac{d(\mu^{(r)}, \tilde{\mu}^{(r)})}{1 + d(\mu^{(r)}, \tilde{\mu}^{(r)})} dr \\ &\le (1 - e^{-R}) \frac{\delta}{1 + \delta} + 2\delta N' \le (1 - e^{-R}) \frac{\varepsilon_4}{1 + \varepsilon_4} + 2\varepsilon_3 N' = \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \qquad \square \end{aligned}$$

8.6 Characterisation of the σ-algebra $\mathcal{B}(\mathcal{N}_\mathcal{U}^\#)$

This section proves the second part of Theorem 8.1.2. We show that all mappings $\Phi_A : \xi \mapsto \xi(A)$, $\xi \in \mathcal{N}_\mathcal{U}^\#$, $A \in \mathcal{B}(\mathcal{U})$, are measurable with respect to the Borel σ-algebra $\mathcal{B}(\mathcal{N}_\mathcal{U}^\#)$ and that $\mathcal{B}(\mathcal{N}_\mathcal{U}^\#)$ is actually generated by all these mappings. This property is very useful to check the measurability of functionals on $\mathcal{N}_\mathcal{U}^\#$, such as Hawkes functionals, as demonstrated in Subsection 8.7.4. Our proof is different from the original one in Daley and Vere-Jones [40, p. 405] as we identify a convenient basis for the $w^\#$-hash topology (Proposition 8.6.1). Note however that this last result is directly inspired by Proposition A2.5.I in Daley and Vere-Jones [40, p. 398], where three different bases for the weak topology on $\mathcal{M}_\mathcal{U}$ are given. Besides, our proof of Theorem 8.1.2.(ii) shows explicitly why the mapping Φ_A is $\mathcal{B}(\mathcal{N}_\mathcal{U}^\#)$-measurable when A is a bounded closed set.

Proposition 8.6.1. *Consider the family of sets*

$$\begin{aligned}\{\xi \in \mathcal{N}^{\#}_{\mathcal{U}} :\ & \xi(F_i) < \mu(F_i) + \varepsilon \text{ for } i = 1, \dots, m, \\ & |\xi(\overline{B}_{r_j}) - \mu(\overline{B}_{r_j})| < \varepsilon \text{ and } \xi(\partial B_{r_j}) = 0 \text{ for } j = 1, \dots, n\},\end{aligned} \tag{8.2}$$

where $\mu \in \mathcal{N}^{\#}_{\mathcal{U}}$, $\varepsilon \in \mathbb{R}_{>0}$, $m, n \in \mathbb{N}$, F_i, $i = 1, \dots, m$, is a bounded closed set of $\mathcal{U}$ and $r_j \in \mathbb{R}_{>0}$, $j = 1, \dots, n$, is such that $\mu(\partial B_{r_j}) = 0$. This family forms a basis that generates the $w^{\#}$-topology.

Proof. **Step 1.** We check that this family is a basis. Let $\mu, \mu' \in \mathcal{N}^{\#}_{\mathcal{U}}$, $\varepsilon, \varepsilon' \in \mathbb{R}_{>0}$, let $F_1, \dots, F_m$ and $F'_1, \dots, F'_{m'}$ be bounded closed sets and let $r_1, \dots, r_n, r'_1, \dots, r'_{n'} > 0$ be such that $\mu(\partial B_{r_j}) = 0$ and $\mu'(\partial B_{r'_j}) = 0$. Consider the sets A and B of the form (8.2) generated by these two collections, respectively, and let $\mu'' \in A \cap B$. We will now find a set C, again of the form (8.2), such that $\mu'' \in C$ and $C \subset A \cap B$. Set the following parameters:

$$\begin{aligned}\delta &:= \min_{i=1,\dots,m} \mu(F_i) + \varepsilon - \mu''(F_i), \quad & \delta' &:= \min_{i=1,\dots,m'} \mu'(F'_i) + \varepsilon' - \mu''(F'_i), \\ \gamma &:= \min_{j=1,\dots,n} \varepsilon - |\mu''(\overline{B}_{r_j}) - \mu(\overline{B}_{r_j})|, \quad & \gamma' &:= \min_{j=1,\dots,n'} \varepsilon' - |\mu''(\overline{B}_{r'_j}) - \mu'(\overline{B}_{r'_j})|;\end{aligned}$$

and let $\varepsilon'' := \min(\delta, \delta', \gamma, \gamma')$. Now, consider the set

$$\begin{aligned}C := \{\xi \in \mathcal{N}^{\#}_{\mathcal{U}} :\ & \xi(F_i) < \mu''(F_i) + \varepsilon'' \text{ for } i = 1, \dots, m, \\ & \xi(F'_i) < \mu''(F'_i) + \varepsilon'' \text{ for } i = 1, \dots, m', \\ & |\xi(\overline{B}_{r_j}) - \mu''(\overline{B}_{r_j})| < \varepsilon'' \text{ and } \xi(\partial B_{r_j}) = 0 \text{ for } j = 1, \dots, n, \\ & |\xi(\overline{B}_{r'_j}) - \mu''(\overline{B}_{r'_j})| < \varepsilon'' \text{ and } \xi(\partial B_{r'_j}) = 0 \text{ for } j = 1, \dots, n'\}.\end{aligned}$$

Clearly, the set C is of the form (8.2). We now finally check that $C \subset A \cap B$. Let $\xi \in C$. For all $i = 1, \dots, m$,

$$\xi(F_i) < \mu''(F_i) + \varepsilon'' \leq \mu(F_i) + \varepsilon,$$

because $\varepsilon'' \leq \mu(F_i) + \varepsilon - \mu''(F_i)$. For all $j = 1, \dots, n$,

$$|\xi(\overline{B}_{r_j}) - \mu(\overline{B}_{r_j})| \leq |\xi(\overline{B}_{r_j}) - \mu''(\overline{B}_{r_j})| + |\mu''(\overline{B}_{r_j}) - \mu(\overline{B}_{r_j})| < \varepsilon,$$

because $\varepsilon'' \leq \varepsilon - |\mu''(\overline{B}_{r_j}) - \mu(\overline{B}_{r_j})|$. Thus, $\xi \in A$. A similar argument yields $\xi \in B$ and so $C \subset A \cap B$.

Step 2. We check that every element of this basis contains an open ball. Consider first any set A of the form (8.2) but for which $n = 1$ (only one ball). Let $\delta \in (0,1)$ be such that $2\delta < \varepsilon$, $\mu(F_i^\delta) = \mu(F_i)$ for all $i = 1, \dots, m$, and

$$\mu\left(\overline{\overline{B}_{r_1}^\delta \setminus \overline{B}_{r_1}}^\delta\right) = 0,$$

which means that δ is chosen small enough such that there are no atoms within a distance δ of the boundary ∂B_{r_1}. Let $R \in \mathbb{R}_{>0}$ be such that $F_i^\delta \subset B_R$ for all $i = 1, \dots, m$ and such that $r_1 + 2\delta < R$. Consider now the ball $B := \{\xi \in \mathcal{N}_\mathcal{U}^\# \,:\, d^\#(\mu, \xi) < \gamma\}$ where $\gamma := e^{-R}\delta/(1+\delta)$. Take any $\xi \in B$ and, by contradiction, assume that $\xi(F_i) > \mu(F_i^\delta) + \delta$ for some $i = 1, \dots, m$. Then, this implies that $d(\xi^{(r)}, \mu^{(r)}) \geq \delta$ for all $r \geq R$, which in turn implies that

$$d^\#(\xi, \mu) \geq \int_R^\infty e^{-r} \frac{\delta}{1+\delta} dr = \gamma.$$

This contradicts the fact that $\xi \in B$ and, thus, we must have

$$\xi(F_i) \leq \mu(F_i^\delta) + \delta = \mu(F_i) + \delta < \mu(F_i) + \varepsilon, \quad i = 1, \dots, m.$$

The same reasoning holds for the closed sets $\overline{B}_{r_1}$ and ∂B_{r_1}, finally implying that

$$\xi(\partial B_{r_1}) = \mu(\partial B_{r_1}) = 0 \quad \text{and} \quad \xi(\overline{B}_{r_1}) - \mu(\overline{B}_{r_1}) \leq \delta < \varepsilon.$$

To obtain that $\xi \in A$, it remains only to show that $\mu(\overline{B}_{r_1}) - \xi(\overline{B}_{r_1}) < \varepsilon$. Using again the previous reasoning, we also have that

$$\xi(\overline{B}_{r_1}^\delta) - \xi(\overline{B}_{r_1}) = \xi(\overline{B}_{r_1}^\delta \setminus \overline{B}_{r_1}) \leq \xi\left(\overline{\overline{B}_{r_1}^\delta \setminus \overline{B}_{r_1}}\right) \leq \mu\left(\overline{\overline{B}_{r_1}^\delta \setminus \overline{B}_{r_1}}^\delta\right) + \delta = \delta,$$

and also that $\mu(\overline{B}_{r_1}) \leq \xi(\overline{B}_{r_1}^\delta) + \delta$. This implies the desired inequality

$$\mu(\overline{B}_{r_1}) - \xi(\overline{B}_{r_1}) = \mu(\overline{B}_{r_1}) - \xi(\overline{B}_{r_1}^\delta) + \xi(\overline{B}_{r_1}^\delta) - \xi(\overline{B}_{r_1}) \leq \delta + \delta < \varepsilon,$$

and allows us to conclude that the ball B is included in the neighbourhood

A. Regarding the general case when the set A is defined by multiple balls (i.e., $n > 1$), simply view it as an intersection of sets A_j, where each A_j is defined by one ball (i.e., $m = 1$). As shown above, for each A_j, we can find an adequate ball with centre μ and radius γ_j. Then, the ball with radius $\gamma = \min \gamma_i$ must be included in A.

Step 3. We check that every open ball contains an element of this basis. Let $\mu \in \mathcal{N}_{\mathcal{U}}^{\#}$, $\varepsilon \in \mathbb{R}_{>0}$ and consider the ball $B := \{\xi \in \mathcal{N}_{\mathcal{U}}^{\#} : d^{\#}(\mu, \xi) < \varepsilon\}$. Let $R > 0$ be such that $e^{-R} < \frac{1}{2}\varepsilon$. Let $\rho_1 < \ldots < \rho_N$ be all the radii in $(0, R)$ such that $\mu(\partial B_{\rho_j}) > 0$, $j = 1, \ldots, N$. Set also $\rho_0 := 0$ and $\rho_{N+1} := R$. Define

$$\rho := \frac{1}{2} \min_{j=1,\ldots,N+1} \rho_j - \rho_{j-1},$$

let $\gamma < \varepsilon/8(N+2)$ and set $\delta := \min(\rho, \gamma)$. Define the bounded closed sets $G_j := \overline{B}_{\rho_j - \delta} \setminus B_{\rho_{j-1}+\delta}$ for $j = 1, \ldots, N+1$ and notice that $\mu(G_j) = 0$. Also, define the radii $r_j := (\rho_{j-1} + \rho_j)/2$, $j = 1, \ldots, N+1$. For all r_j, reusing the last part of the proof of Proposition A2.5.I in Daley and Vere-Jones [40, p. 399], we know that we can find $\tilde{\varepsilon}_j \in (0, 1)$ and a finite family of closed bounded sets $F_{1,j}, \ldots, F_{m_j,j}$ such that

$$\begin{aligned}
&A_j := \\
&\{\xi \in \mathcal{N}_{\mathcal{U}}^{\#} : \xi(F_{i,j}) < \mu(F_{i,j}) + \tilde{\varepsilon}_j \text{ for } i = 1, \ldots, m_j, |\xi(\overline{B}_{r_j}) - \mu(\overline{B}_{r_j})| < \tilde{\varepsilon}_j\} \\
&\qquad\qquad\qquad\qquad \subset \{\xi \in \mathcal{N}_{\mathcal{U}}^{\#} : d(\mu^{(r_j)}, \xi^{(r_j)}) < c\},
\end{aligned}$$

where here we choose c such that $(1 - e^{-R})c/(1 + c) < \varepsilon/4$. Finally, set $\tilde{\varepsilon} = \min \tilde{\varepsilon}_j$ and consider the set

$$\begin{aligned}
A := \{\xi \in \mathcal{N}_{\mathcal{U}}^{\#} : &\xi(F_{i,j}) < \mu(F_{i,j}) + \tilde{\varepsilon} \text{ for } i = 1, \ldots, m_j, \\
&\xi(G_j) < \mu(G_j) + \tilde{\varepsilon}, \\
&|\xi(\overline{B}_{r_j}) - \mu(\overline{B}_{r_j})| < \tilde{\varepsilon} \text{ and } \xi(\partial B_{r_j}) = 0, \\
&\text{for } j = 1, \ldots, N+1\},
\end{aligned}$$

which is of the form (8.2) and is such that $A \subset A_j$, $j = 1, \ldots, N+1$. For all $\xi \in A$, this implies that $d(\mu^{(r_j)}, \xi^{(r_j)}) < c$, $j = 1, \ldots, N+1$. This also implies that $\xi(G_j) = 0$, and thus $r \mapsto d(\mu^{(r)}, \xi^{(r)})$ is constant on each interval

$(\rho_{j-1}+\delta, \rho_j-\delta)$, $j=1,\dots,N+1$. Noting that $r_j \in (\rho_{j-1}+\delta, \rho_j-\delta)$, it remains to check that

$$\begin{aligned} d^{\#}(\mu,\xi) &< \int_0^R e^{-r}\frac{d(\mu^{(r)},\xi^{(r)})}{1+d(\mu^{(r)},\xi^{(r)})}dr + \frac{1}{2}\varepsilon \\ &< 2\delta(N+2) + (1-e^{-R})\frac{c}{1+c} + \frac{1}{2}\varepsilon < \frac{1}{4}\varepsilon + \frac{1}{4}\varepsilon + \frac{1}{2}\varepsilon = \varepsilon. \end{aligned}$$

Consequently, we have indeed that $A \subset B$, which concludes the proof. □

Proof of Theorem 8.1.2.(ii). **Step 1.** We first show that Φ_A is $\mathcal{B}(\mathcal{N}_{\mathcal{U}}^{\#})$-measurable for all bounded closed set A. Let $n \in \mathbb{N}$. We prove that $I := \{\xi \in \mathcal{N}_{\mathcal{U}}^{\#} : \xi(A) \leq n\}$ is an open set of $\mathcal{N}_{\mathcal{U}}^{\#}$, implying that Φ_A is indeed $\mathcal{B}(\mathcal{N}_{\mathcal{U}}^{\#})$-measurable. If $A=\emptyset$, then $I = \mathcal{N}_{\mathcal{U}}^{\#}$, which is open. From now on, we assume that $A \neq \emptyset$. Let $\mu \in I$ (I is clearly not empty). Let $\delta \in (0,1)$ be such that $\mu(A) = \mu(A^\delta)$ (this is always possible since μ has a finite number of atoms in $A^\gamma \setminus A$, with $\gamma = 1$, say). Let $R>0$ be such that $A^\delta \subset B_R$. Consider the open ball $J := \{\nu \in \mathcal{N}_{\mathcal{U}}^{\#} : d^{\#}(\mu,\nu) < \varepsilon\}$ with $\varepsilon = e^{-R}\delta/(1+\delta)$. We then have that $J \subset I$, which implies that I is open. Indeed, let $\nu \in J$ and, by contradiction, assume that $\nu(A) > \mu(A^\delta)+\delta$. Then, for all $r \geq R$, this implies that

$$\nu^{(r)}(A) = \nu(A) > \mu(A^\delta)+\delta = \mu^{(r)}(A^\delta)+\delta,$$

which means that $d(\mu^{(r)},\nu^{(r)}) \geq \delta$. Hence,

$$d^{\#}(\mu,\nu) \geq \int_R^\infty e^{-r}\frac{\delta}{1+\delta}dr = \varepsilon,$$

which contradicts the assumption that $\nu \in J$. Consequently, we have that $\nu(A) \leq \mu(A^\delta)+\delta = \mu(A)+\delta$. Since, $\nu(A) \in \mathbb{N}$, $\mu(A) \in \mathbb{N}$ and $\delta < 1$, this implies that $\nu(A) \leq \mu(A) \leq n$, and thus $\nu \in I$.

Step 2. Consider the class $\mathcal{C}$ of sets

$$\mathcal{C} := \{A \in \mathcal{B}(\mathcal{U}) : \Phi_A \text{ is } \mathcal{B}(\mathcal{N}_{\mathcal{U}}^{\#})\text{-measurable}\}.$$

By the continuity of measures [74, Lemma 1.14 p. 8], we have that $\Phi_{A_n} \uparrow \Phi_A$

for any sequence $A_n \uparrow A$, and since the limit of measurable functions is measurable [74, Lemma 1.9 p. 6], $\mathcal{C}$ is closed under increasing limits. In other words, $\mathcal{C}$ forms a monotone class. Moreover, consider the class $\mathcal{R}$ of sets of the form $\bigcup_{i=1}^n A_i \setminus B_i$ where $n \in \mathbb{N}$ and $A_i, B_i \in \mathcal{B}(\mathcal{U})$ are bounded closed sets such that $(A_i \setminus B_i) \cap (A_j \setminus B_j) = \emptyset$ as soon as $i \neq j$ (i.e., we consider finite disjoint unions of differences of bounded closed sets). One can check that $\mathcal{R}$ is stable by finite intersections and symmetric differences (perhaps the most difficult is to see that, for any bounded closed sets A_1, A_2, B_1, B_2, the difference $(A_1 \setminus B_1) \setminus (A_2 \setminus B_2)$ can be written as a disjoint union of differences of bounded closed sets). This means that $\mathcal{R}$ forms a ring. Besides, for any bounded closed sets $A, B \in \mathcal{U}$, since $\xi(A) < \infty$ for all $\xi \in \mathcal{N}_{\mathcal{U}}^{\#}$, we have that $\Phi_{A \setminus B} = \Phi_{A \setminus (A \cap B)} = \Phi_A - \Phi_{A \cap B}$. As $A \cap B$ is still a bounded closed set, by applying the first part of the proof, we obtain that $\Phi_{A \setminus B}$ is measurable. By the countable additivity of measures, this implies that Φ_A is measurable for any set $A \in \mathcal{R}$, and thus $\mathcal{R} \subset \mathcal{C}$. By the monotone class theorem [40, p. 369], we then have $\sigma(\mathcal{R}) \subset \mathcal{C}$. But $\mathcal{R}$ contains all the bounded closed balls and any open set in $\mathcal{U}$ is a countable union of those since $\mathcal{U}$ is separable. Consequently, we must have $\mathcal{B}(\mathcal{U}) = \sigma(\mathcal{R}) \subset \mathcal{C}$, meaning that Φ_A is measurable for all $A \in \mathcal{B}(\mathcal{U})$.

Step 3. To show that $\mathcal{B}(\mathcal{N}_{\mathcal{U}}^{\#})$ is actually generated by all mappings Φ_A, $A \in \mathcal{B}(\mathcal{U})$, consider any σ-algebra $\mathcal{R}$ on $\mathcal{N}_{\mathcal{U}}^{\#}$ such that all mappings Φ_A are measurable. Then, all the sets of the form (8.2) should belong to $\mathcal{R}$ and, by Proposition 8.6.1, these sets form a basis for the $w^{\#}$-topology. Since $\mathcal{N}_{\mathcal{U}}^{\#}$ is separable, any open set of the $w^{\#}$-topology can be represented as a countable union of these sets and, thus, $\mathcal{B}(\mathcal{N}_{\mathcal{U}}^{\#}) \subset \mathcal{R}$. $\square$

8.7 Applications

8.7.1 The subspace $\mathcal{N}_{\mathbb{R}\times\mathscr{M}}^{\#g}$ is Borel

The following result is unlikely to be original, but we could not find it in Daley and Vere-Jones [41].

Lemma 8.7.1. *The set $\mathcal{N}_{\mathbb{R}\times\mathscr{M}}^{\#g}$ is a Borel subset of $\mathcal{N}_{\mathbb{R}\times\mathscr{M}}^{\#}$, meaning that*

$\mathcal{N}^{\#g}_{\mathbb{R}\times\mathscr{M}} \in \mathcal{B}(\mathcal{N}^{\#}_{\mathbb{R}\times\mathscr{M}})$.

Proof. For all $n \in \mathbb{N}$, define the sets

$$F_n := \Big\{\xi \in \mathcal{N}^{\#}_{\mathbb{R}\times\mathscr{M}} \,:\, \xi([-n,n]\times\mathscr{M}) < \infty,\ \xi(\{t\}\times\mathscr{M}) \le 1 \text{ for all } t \in [-n,n]\Big\}$$

and notice that $\mathcal{N}^{\#g}_{\mathbb{R}\times\mathscr{M}} = \bigcap_{n\in\mathbb{N}} F_n$. Next, let $n \in \mathbb{N}$. By Proposition A2.1.IV in Daley and Vere-Jones [40, p. 385], the interval $[-n,n]$ contains a dissecting system $((A_{ij})_{j\in\{1,\ldots,j_i\}})_{i\in\mathbb{N}}$ where $A_{ij} \in \mathcal{B}(\mathbb{R})$ for any $j \in \{1,\ldots,j_i\}$ and $i \in \mathbb{N}$, see Definition A1.6.1 in Daley and Vere-Jones [40, p. 382]. We show that

$$F_n = \left\{\xi \in \mathcal{N}^{\#}_{\mathbb{R}\times\mathscr{M}} \,:\, \xi([-n,n]\times\mathscr{M}) < \infty,\ \limsup_{i\to\infty}\ \sup_{j\in\{1,\ldots,j_i\}} \xi(A_{ij}\times\mathscr{M}) \le 1\right\} =: G_n.$$

Let $\xi \in F_n$. Then $\xi(\cdot\times\mathscr{M})$ has finitely many atoms $t_1,\ldots,t_p$ in $[-n,n]$ for some $p \in \mathbb{N}$ and their mass cannot exceed one. A key property of the dissecting system is that, for each pair of distinct atoms t_{q_1} and t_{q_2} with $q_1 \neq q_2$, there exists $n(q_1,q_2) \in \mathbb{N}$ such that, for all $i > n(q_1,q_2)$, $t_{q_1} \in A_{ij}$ implies $t_{q_2} \notin A_{ij}$. Thus, define

$$i^* := \max_{q_1,q_2\in\{1,\ldots,p\},\, q_1\neq q_2} n(q_1,q_2)$$

and then, $\xi(A_{ij}\times\mathscr{M}) \le 1$ for all $j \in \{1,\ldots,j_i\}$ and $i > i^*$, which implies that

$$\limsup_{i\to\infty}\ \sup_{j\in\{1,\ldots,j_i\}} \xi(A_{ij}\times\mathscr{M}) \le 1,$$

which in turn indicates that $\xi \in G_n$. Now, let $\xi \in G_n$ and $t \in [-n,n]$. Another salient property of the dissecting system is that there exists a sequence $(j_i)_{i\in\mathbb{N}}$ such that $\xi(\{t\}\times\mathscr{M}) = \lim_{i\to\infty} \xi(A_{ij_i}\times\mathscr{M})$. But since $\xi \in G_n$, we have that

$$\xi(\{t\}\times\mathscr{M}) = \lim_{i\to\infty} \xi(A_{ij_i}\times\mathscr{M}) \le \limsup_{i\to\infty}\ \sup_{j\in\{1,\ldots,j_i\}} \xi(A_{ij}\times\mathscr{M}) \le 1,$$

which means that $\xi \in F_n$. Now that we have shown that $F_n = G_n$, we invoke Theorem 8.1.2 to deduce that $\xi \mapsto \xi([-n,n] \times \mathscr{M})$ and $\xi \mapsto \xi(A_{ij} \times \mathscr{M})$, for any $j \in \{1,\dots,j_i\}$ and $i \in \mathbb{N}$, are measurable and use Lemma 1.9 in Kallenberg [74, p. 6] to conclude that

$$\xi \mapsto \limsup_{i\to\infty} \sup_{j\in\{1,\dots,j_i\}} \xi(A_{ij} \times \mathscr{M})$$

is measurable. It then follows that $F_n \in \mathcal{B}(\mathcal{N}^{\#}_{\mathbb{R}\times\mathscr{M}})$, whence we have that $\mathcal{N}^{\#g}_{\mathbb{R}\times\mathscr{M}} = \bigcap_{n\in\mathbb{N}} F_n \in \mathcal{B}(\mathcal{N}^{\#}_{\mathbb{R}\times\mathscr{M}})$. $\square$

8.7.2 Measurability and continuity properties of shifts and restrictions

From Daley and Vere-Jones [41, p. 178, Lemma 12.1.I], we know the shift operators are continuous under the $w^{\#}$-topology. We are able to go further and show that $\theta_t\xi$ is actually jointly continuous in ξ and t. We also prove that taking the restriction to the positive or negative real line of a boundedly finite measure is a measurable operation. Moreover, we show that $\theta_t\xi^{<0}$ is left-continuous as a function of $t \in \mathbb{R}$ for any $\xi \in \mathcal{N}^{\#}_{\mathcal{U}}$, which is crucial in our proof that an intensity functional applied to the history of a point process generates a predictable process (Lemma 6.4.4).

Lemma 8.7.2. *When $\mathcal{N}^{\#}_{\mathbb{R}\times\mathcal{U}}$ is equipped with the $w^{\#}$-distance $d^{\#}$ and $\mathcal{N}^{\#}_{\mathbb{R}\times\mathcal{U}} \times \mathbb{R}$ is equipped with the product metric, the mapping*

$$\begin{aligned} \mathcal{N}^{\#}_{\mathbb{R}\times\mathcal{U}} \times \mathbb{R} &\to \mathcal{N}^{\#}_{\mathbb{R}\times\mathcal{U}} \\ (\xi, t) &\mapsto \theta_t\xi \end{aligned}$$

is continuous.

Proof. Let $\xi \in \mathcal{N}^{\#}_{\mathbb{R}\times\mathcal{U}}$, $t \in \mathbb{R}$ and let $(\xi_n, t_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{N}^{\#}_{\mathbb{R}\times\mathcal{U}} \times \mathbb{R}$ such that $d^{\#}(\xi_n, \xi) \to 0$ and $t_n \to t$ as $n \to \infty$. By Theorem 8.1.1, it is enough to show that $\xi_n(A + t_n) \to \xi(A + t)$ as $n \to \infty$ for any bounded $A \in \mathcal{B}(\mathbb{R} \times \mathcal{U})$ such that $\xi(\partial(A + t)) = 0$. For such a set A, which we can assume without loss of generality to be non-empty, there exists $\delta > 0$ such that $B_{2\delta}(a_n) \subset (A+t)$ and $\xi(\partial B_\delta(a_n)) = 0$, $n = 1, \dots, N$, where $a_1, \dots, a_N$ are the atoms of ξ in $A+t$, and such that $\xi((A+t)^\delta) = \xi(A+t)$ with $\xi(\partial((A+t)^\delta)) = 0$. Introduce the two bounded sets

$$S_1 := (A+t)^\delta \setminus (A+t) \quad \text{and} \quad S_2 := (A+t) \setminus \left(\bigcup_{n=1}^N B_\delta(a_n) \right)$$

and notice that $\xi(S_1) = \xi(S_2) = \xi(\partial S_1) = \xi(\partial S_2) = 0$. Since $d^\#(\xi_n, \xi) \to 0$ as $n \to \infty$, we have that $\xi_n(S_1) = \xi_n(S_2) = \xi_n(\partial S_1) = \xi_n(\partial S_2) = 0$ for n large enough. This implies that, for n large enough, all the atoms of ξ_n in $(A+t)^\delta$ actually lie in $(A+t)$ and their distance to the boundary of $(A+t)$ is bigger than δ (all the atoms are in the balls $B_\delta(a_n)$). This means that, for n large enough, $\xi_n((A+t) \setminus (A+s)) = \xi_n((A+s) \setminus (A+t)) = 0$ for all $s \in \mathbb{R}$ such that $|t-s| < \delta$, implying that $\xi_n(A+t) = \xi_n(A+s)$ for all such n and s. But for n large enough, we also have that $|t_n - t| < \delta$ and $\xi_n(A+t) = \xi(A+t)$, which finally gives that, for such large enough n,

$$\xi_n(A+t_n) = \xi_n(A+t+(t_n-t)) = \xi_n(A+t) = \xi(A+t). \qquad \square$$

Lemma 8.7.3. *The restrictions $\xi^{<0}$, $\xi^{\leq 0}$, $\xi^{>0}$ and $\xi^{\geq 0}$ are measurable mappings from $\mathcal{N}^\#_{\mathbb{R}\times\mathcal{U}}$ into itself.*

Proof. We prove the assertion for $\xi^{<0}$, the other three restrictions can be treated similarly. Consider the function $f : \mathcal{N}^\#_{\mathbb{R}\times\mathcal{U}} \ni \xi \mapsto \xi^{<0} \in \mathcal{N}^\#_{\mathbb{R}\times\mathcal{U}}$. Remember that, by Theorem 8.1.2, the Borel σ-algebra $\mathcal{B}(\mathcal{N}^\#_{\mathbb{R}\times\mathcal{U}})$ is generated by the sets

$$F_{A,n} := \{\xi \in \mathcal{N}^\#_{\mathbb{R}\times\mathcal{U}} : \xi(A) \in [n, \infty]\}, \quad A \in \mathcal{B}(\mathcal{N}^\#_{\mathbb{R}\times\mathcal{U}}),\ n \in \mathbb{R}.$$

Since

$$f^{-1}(F_{A,n}) = \{\xi \in \mathcal{N}^\#_{\mathbb{R}\times\mathcal{U}} : \xi(A \cap \mathbb{R}_{<0} \times \mathcal{U}) \in [n, \infty]\} \in \mathcal{B}(\mathcal{N}^\#_{\mathbb{R}\times\mathcal{U}}),$$

we conclude that f is measurable by Lemma 1.4 in Kallenberg [74, p. 4]. $\square$

Lemma 8.7.4. *Let $\xi \in \mathcal{N}^{\#}_{\mathbb{R}\times\mathcal{U}}$. Then the mapping*

$$\begin{aligned} \mathbb{R} &\to \mathcal{N}^{\#}_{\mathbb{R}\times\mathcal{U}} \\ t &\mapsto (\theta_t \xi)^{<0} \end{aligned}$$

is left continuous when $\mathcal{N}^{\#}_{\mathbb{R}\times\mathcal{U}}$ is equipped with the $w^{\#}$-distance $d^{\#}$.

Proof. Fix $t \in \mathbb{R}$ and take any non-decreasing sequence $(t_n)_{n\in\mathbb{N}}$ in $\mathbb{R}$ such that $t_n \uparrow t$ as $n \to \infty$. By Theorem 8.1.1, it is enough to show that

$$(\theta_{t_n}\xi)^{<0}(A) \to (\theta_t\xi)^{<0}(A), \quad n \to \infty,$$

for all bounded $A \in \mathcal{B}(\mathbb{R} \times \mathcal{U})$ such that $(\theta_t\xi)^{<0}(\partial A) = 0$. Clearly, it suffices to consider bounded Borel sets A such that $A \subset \mathbb{R}_{<0} \times \mathcal{U}$. First, consider the case where $\partial A \subset \mathbb{R}_{<0} \times \mathcal{U}$. This implies that

$$(\theta_t\xi)(\partial A) = (\theta_t\xi)^{<0}(\partial A) = 0.$$

By Lemma 8.7.2, and using again the characterisation of Theorem 8.1.1, this implies that

$$(\theta_{t_n}\xi)^{<0}(A) = (\theta_{t_n}\xi)(A) \to (\theta_t\xi)(A) = (\theta_t\xi)^{<0}(A), \quad n \to \infty.$$

Second, consider the remaining case where $\partial A \cap \{0\} \times \mathcal{U} \neq \varnothing$. Then, $(\theta_t\xi)^{<0}(\partial A) = 0$ does not imply anymore that $(\theta_t\xi)(\partial A) = 0$. However, let ξ_- be the measure ξ that omits all atoms with time coordinate t. Then, for this measure ξ_-, we have again that

$$(\theta_t\xi_-)(\partial A) = (\theta_t\xi)^{<0}(\partial A) = 0.$$

Since $t_n \leq t$, and adapting the preceding argument for ξ_-, we finally find that

$$(\theta_{t_n}\xi)^{<0}(A) = (\theta_{t_n}\xi)(A) = (\theta_{t_n}\xi_-)(A) \to (\theta_t\xi_-)(A) = (\theta_t\xi)^{<0}(A), \quad n \to \infty.$$

□

8.7.3 Enumeration representation of marked point processes

The following result confirms that a non-explosive enumeration in $\mathbb{R}_{>0} \times \mathscr{M}$ corresponds indeed to a non-explosive marked point process.

Lemma 8.7.5. *Let $(T_n, M_n)_{n\in\mathbb{N}}$ be an enumeration in $\mathbb{R}_{>0} \times \mathscr{M}$ such that $\lim_{n\to\infty} T_n = \infty$ a.s. Let $F \in \mathcal{F}$ be the almost sure event that $\lim_{n\to\infty} T_n = \infty$ and define*

$$N(\omega) := \begin{cases} \sum_n \delta_{(T_n(\omega), M_n(\omega))} \mathbb{1}_{\{T_n(\omega)<\infty\}}, & \text{if } \omega \in F, \\ 0, & \text{if } \omega \notin F. \end{cases}$$

Then, N defines a non-explosive marked point process on $\mathbb{R}_{\geq 0} \times \mathscr{M}$.

Proof. By Proposition 9.1.X in Daley and Vere-Jones [41, p. 13], N defines a non-explosive point process on $\mathbb{R}_{\geq 0} \times \mathscr{M}$. Moreover, the monotonicity of the sequence $(T_n)_{n\in\mathbb{N}}$ implies that $N(\{t\} \times \mathscr{M}) = 0$ or 1 for all $\omega \in \Omega$. Also, using that $\lim_{n\to\infty} T_n = \infty$ on F, notice that $N(\omega, A \times \mathscr{M}) < \infty$, for every bounded set $A \in \mathcal{B}(\mathbb{R}_{\geq 0})$, for all $\omega \in \Omega$. This means that $N \in \mathcal{N}^{\#g}_{\mathbb{R}_{\geq 0}\times\mathscr{M}}$ and, thus, N defines a non-explosive marked point process. $\square$

Conversely, every non-explosive marked point process generates an enumeration.

Lemma 8.7.6. *Let N be a non-explosive marked point process on $\mathbb{R}_{\geq 0} \times \mathscr{M}$ such that $N(\{0\} \times \mathscr{M}) = 0$ a.s. Define the sequence $(T_n)_{n\in\mathbb{N}}$ by*

$$T_n := \sup\{t > 0 \,:\, N((0,t) \times \mathscr{M}) \leq n\}.$$

Then $(T_n)_{n\in\mathbb{N}}$ is a non-decreasing sequence of random variables in $(0, \infty]$. Moreover, for each $n \in \mathbb{N}$, on $\{T_n < \infty\}$, one can define M_n as the unique element in $\mathscr{M}$ such that $N(\{T_n\} \times \{M_n\}) > 0$. On $\{T_n = \infty\}$, simply set $M_n = m_\infty$ for some fixed $m_\infty \in \mathscr{M}$. Then, $(T_n, M_n)_{n\in\mathbb{N}}$ is an enumeration in $\mathbb{R}_{\geq 0} \times \mathscr{M}$ such that

$$N = \sum_{n\in\mathbb{N}} \delta_{(T_n, M_n)} \mathbb{1}_{\{T_n<\infty\}} \quad a.s. \tag{8.3}$$

and $\lim_{n\to\infty} T_n(\omega) = \infty$ *for all* $\omega \in \Omega$.

Proof. We proceed in several steps.

(i) For each $n \in \mathbb{N}$, the mapping $\omega \mapsto T_n(\omega)$ is measurable. Indeed, notice that $\{T_n < t\} = \{N((0,t) \times \mathscr{M}) > n\}$. Then recall that, by Theorem 8.1.2, $\mathcal{N}^{\#}_{\mathbb{R}_{\geq 0}\times\mathscr{M}} \ni \xi \mapsto \xi((0,t) \times \mathscr{M})$ is measurable and, thus, as a composition [74, Lemma 1.7, p. 5], the mapping $\omega \mapsto N(\omega, (0,t)\times\mathscr{M})$ is measurable. Consequently, $\{T_n < t\} \in \mathcal{F}$. We conclude using Lemma 1.4 in Kallenberg [74, p. 4] that the mapping $\omega \mapsto T_n(\omega)$ is measurable.

(ii) Using the fact that the ground measure is simple, $T_n < \infty$ implies that $T_n < T_{n+1}$. Also, it is easy to check that when $T_n = \infty$, then $T_{n+1} = \infty$. Hence, $(T_n)_{n\in\mathbb{N}}$ is sequence of random variables in $(0, \infty]$ satisfying the monotonicity of an enumeration.

(iii) For each $n \in \mathbb{N}$, the mapping $\omega \mapsto M_n(\omega)$ is well defined (using again the fact that the ground measure is simple). Also, this mapping is measurable. Indeed, let $A \in \mathcal{B}(\mathscr{M})$ and consider the most delicate case where $m_\infty \in A$. Notice that

$$\{M_n \in A\} = (\{T_n < \infty\} \cap \{N(\{T_n\} \times A) > 0\}) \cup \{T_n = \infty\}.$$

Based on what we have seen so far, we know that $\{T_n = \infty\} \in \mathcal{F}$. Therefore, it suffices to show that the set

$$\{T_n < \infty\} \cap \{N(\{T_n\} \times A) > 0\}$$

is measurable. To this end, notice that

$$\{T_n < \infty\} \cap \{N(\{T_n\} \times A) > 0\} = \{T_n < \infty\} \cap \{\theta_{T_n} N(\{0\} \times A) > 0\},$$

where θ_{T_n} is the shift operator defined in Subsection 5.4. Then, by Lemma 8.7.2, we know that the mapping

$$\mathcal{N}^{\#}_{\mathbb{R}_{\geq 0}\times\mathscr{M}} \times \mathbb{R}_{\geq 0} \ni (\xi, t) \mapsto \theta_t \xi \in \mathcal{N}^{\#}_{\mathbb{R}_{\geq 0}\times\mathscr{M}}$$

is continuous and thus, by Lemma 1.5 in Kallenberg [74, p. 4], measurable. Also, by Lemma 1.8 in Kallenberg [74, p. 5], the mapping given by $\omega \mapsto (N(\omega), T_n(\omega))$ is measurable, and thus, as a composition [74, Lemma 1.7, p. 5], the mapping $\omega \mapsto \theta_{T_n(\omega)} N(\omega)$ is measurable. Using again Theorem 8.1.2, we conclude that

$$\{T_n < \infty\} \cap \{\theta_{T_n} N(\{0\} \times A) > 0\} \in \mathcal{F}$$

and, thus, the mapping $\omega \mapsto M_n(\omega)$ is measurable. So far, these first three steps establish that $(T_n, M_n)_{n \in \mathbb{N}}$ is an enumeration. Moreover, (8.3) holds by construction.

(iv) Since $N(\cdot \times \mathscr{M}) \in \mathcal{N}^{\#}_{\mathbb{R}_{\geq 0}}$, we have that $\lim_{n \to \infty} T_n(\omega) = \infty$ for all $\omega \in \Omega$. □

Remark 8.7.7. On the one hand, Lemma 8.7.5 gives us a mapping that generates a non-explosive marked point process out of a non-explosive enumeration. One can see that if two non-explosive enumerations are not almost surely equal, then the corresponding non-explosive marked point processes cannot be almost surely equal either. In other words, the mapping of Lemma 8.7.5 is injective. On the other hand, Lemma 8.7.6 tells us that this mapping is surjective. As a consequence, the above two lemmas tell us that non-explosive enumerations and non-explosive marked point processes are two equivalent ways of looking at the same object.

8.7.4 Hawkes functionals

One can generalise multivariate linear Hawkes processes by defining Hawkes functionals as intensity functionals $\psi : \mathscr{M} \times \mathcal{N}^{\#}_{\mathbb{R} \times \mathscr{M}} \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ of the form

$$\psi(m \,|\, \xi) = \nu(m) + \iint_{(-\infty,0) \times \mathscr{M}} k(-t', m', m) \xi(dt', dm'), \tag{8.4}$$

where $m \in \mathscr{M}$, $\xi \in \mathcal{N}^{\#}_{\mathbb{R} \times \mathscr{M}}$ and $\nu : \mathscr{M} \to \mathbb{R}_{\geq 0}$ and $k : \mathbb{R} \times \mathscr{M} \times \mathscr{M} \to \mathbb{R}_{\geq 0}$ are non-negative measurable functions. We show that such event functionals are measurable, so that they are admissible in our framework.

Proposition 8.7.8. *Hawkes functionals of the form* (8.4) *are jointly measurable in* $m \in \mathscr{M}$ *and* $\xi \in \mathcal{N}^{\#}_{\mathcal{U}}$.

Proof. It will be enough to show that the integral term in (8.4), now denoted by $I(m,\xi)$, is measurable as a function of $m \in \mathcal{E}$ and $\xi \in \mathcal{N}^{\#}_{\mathbb{R}\times\mathscr{M}}$. First, consider the functions k that are of the form $k(t',m',m) = \mathbb{1}_S(t',m',m)$ where $S \in \mathcal{B}(\mathbb{R}\times\mathscr{M}\times\mathscr{M})$ and let $\mathcal{C}$ be the class of sets $S \in \mathcal{B}(\mathbb{R}\times\mathscr{M}\times\mathscr{M})$ such that $(m,\xi) \mapsto I(m,\xi)$ is measurable. By monotone convergence, the class $\mathcal{C}$ is a monotone class (i.e., it is closed under monotonically increasing sequences). Denote by $\mathcal{R}$ the class of sets of the form $\bigcup_{i=1}^{n} A_i \times M'_i \times M_i$ where $A_i \in \mathcal{B}(\mathbb{R})$, $M'_i \in \mathcal{B}(\mathscr{M})$, $M_i \in \mathcal{B}(\mathscr{M}), n \in \mathbb{N}$. This class forms a ring (i.e., it is closed under finite intersections and symmetric differences). Indeed, the difference of unions of Cartesian products is a union of Cartesian product. Moreover, since any union of Cartesian products can be decomposed as a union of disjoint Cartesian products, we have that $\mathcal{R} \subset \mathcal{C}$. Indeed, by Theorem 8.1.2, for any $A \in \mathcal{B}(\mathbb{R})$, $M' \in \mathcal{B}(\mathscr{M})$, $M \in \mathcal{B}(\mathscr{M})$, the function

$$
\begin{aligned}
(m,\xi) \mapsto \iint_{(-\infty,0)\times\mathscr{M}} & \mathbb{1}_A(-t')\mathbb{1}_{M'}(m')\mathbb{1}_M(m)\xi(dt',dm') \\
&= \mathbb{1}_M(m)\xi((-A)\cap(-\infty,0)\times M')
\end{aligned}
$$

is measurable. Then, by the monotone class theorem [40, p. 369], we have that $\mathcal{B}(\mathbb{R}\times\mathscr{M}\times\mathscr{M}) = \sigma(\mathcal{R}) \subset \mathcal{C}$. The linearity of the integral implies that $(m,\xi) \mapsto I(m,\xi)$ is measurable for all simple functions k and, by monotone convergence, for all non-negative measurable functions k [74, Lemma 1.11, p. 7]. □

9 Outlook

The present thesis addressed the HM dichotomy by proposing a general, flexible and unifying modelling framework: hybrid marked point processes. The theoretical foundations of these new processes were laid out with, in particular, the detailed derivation of strong existence and uniqueness results that dispense with a classical Lipschitz condition. From a more applied perspective, the relevance of this new toolbox was illustrated with the estimation of state-dependent Hawkes processes from high-frequency financial data. We hope that both our theoretical and applied contributions will inspire and encourage further progress in this direction. As an ending, we outline below what we think are interesting and challenging research problems to investigate as a follow up.

Stationarity and MLE asymptotics

One challenge is to study the asymptotic properties of the MLE estimator for hybrid marked point processes. Since such endeavour often requires the considered process to be stationary and ergodic (as it is the case for the existing results on Hawkes processes), a complementing problem is to identify reasonable stationarity and ergodicity conditions, as discussed in Section 7.6. Let us also mention that the non-parametric inference techniques for Hawkes processes do not extend naturally to state-dependent Hawkes processes. Such extensions will certainly require innovative approaches and breakthroughs.

Linking multiple scales

We discussed in Subsection 1.6.5 our reserve on the ability of Hawkes processes to inform us on the actual strategic behaviour of traders, arguing that such models can only capture the aggregate market reaction to different situations. What would in fact shed light on this matter is to understand if and how agent-based models can generate LOB data that behaves like a (state-dependent) Hawkes process. In the same spirit, another question is: what are the price and volatility processes implied by a hybrid-marked-point-process model of high-frequency data? Can the non-Markovian feedback loop between events and the state process give rise to low-frequency dynamics that cannot be generated by simple Hawkes processes and continuous-time Markov chains?

Model performance and selection

It might be argued that, intuitively, maximising the likelihood of a point process is equivalent to minimising the distance between the residuals' distribution and the exponential distribution. If this is true (and it is an interesting question in itself), the Q-Q plots in Chapter 4 are not really assessing the goodness of fit but rather the performance of the optimisation algorithm. Besides, Theorem 3.4.3 does not say that the time change that transforms the point process to a simple Poisson process is unique. Consequently, it is not clear that assessing the model performance via the residuals, as it is currently done in the literature, protects against misspecification. More generally, it seems that determining the best techniques to assess and rank different point-process models remains an open question. We also note that, at least in the context of LOB modelling, Hawkes processes are very rarely used to produce forecasts in out-of-sample exercises.

Trading applications

While Chapter 4 has illustrated the statistical significance of the state-dependence in Hawkes processes, the added value of modelling this dependence in actual trading applications remains to be demonstrated.

Given that the effects captured by (state-dependent) Hawkes processes have short effective timescales of typically less than a second (although the excitation kernel can be slowly decreasing), one could expect that these models generate added value mainly when applied to optimal order placement problems. For instance, as these models are sensitive to the recent order flow history, they could maybe output more accurate fill probabilities, or allow for more effective manners of avoiding adverse selection.

Moreover, these point-process models already embed market impact, as an order can trigger further orders in response. Consequently, they can perhaps offer a framework where optimal order placement and optimal execution (how to split a large trade into small orders) are treated as one unique problem. For example, the market and the execution strategy would be represented by two coupled point processes that are self- and cross-exciting, e.g., a market order submitted as part of the execution strategy creates a certain market reaction. We note that in the classical optimal execution literature, *market* impact is only modelled as *price* impact and that the impact function that takes as an input the speed of execution is difficult to estimate. With point-process models, market impact would be essentially modelled by the excitation kernel, which can be naturally estimated from high-frequency data.

Finally, as already pointed out in Subsection 4.9, the two LOB models introduced in this thesis, namely Model_{S} and Model_{QI}, are *reduced-form*, in the sense that the LOB is described by only one variable. These models can of course be naturally extended to incorporate more descriptive quantities of the LOB, but it would interesting to study how hybrid marked point processes can be used as building blocks of *full* LOB models.

References

[1] Abergel, F. and Jedidi, A. (2013). A mathematical approach to order book modeling. *International Journal of Theoretical and Applied Finance*, 16(05):1350025.

[2] Abergel, F. and Jedidi, A. (2015). Long-time behavior of a Hawkes process–based limit order book. *SIAM Journal on Financial Mathematics*, 6(1):1026–1043.

[3] Achab, M., Bacry, E., Muzy, J. F., and Rambaldi, M. (2018). Analysis of order book flows using a non-parametric estimation of the branching ratio matrix. *Quantitative Finance*, 18(2):199–212.

[4] Alfonsi, A. and Blanc, P. (2016). Dynamic optimal execution in a mixed-market-impact Hawkes price model. *Finance and Stochastics*, 20(1):183–218.

[5] Almgren, R. and Chriss, N. (2001). Optimal execution of portfolio transactions. *Journal of Risk*, 3:5–40.

[6] Avellaneda, M., Reed, J., and Stoikov, S. (2011). Forecasting prices from Level-I quotes in the presence of hidden liquidity. *Algorithmic Finance*, 1(1):35–43.

[7] Axioglou, C. and Skouras, S. (2011). Markets change every day: Evidence from the memory of trade direction. *Journal of Empirical Finance*, 18(3):423–446.

[8] Bacry, E., Delattre, S., Hoffmann, M., and Muzy, J.-F. (2013). Modelling microstructure noise with mutually exciting point processes. *Quantitative Finance*, 13(1):65–77.

[9] Bacry, E., Jaisson, T., and Muzy, J.-F. (2016). Estimation of slowly decreasing Hawkes kernels: application to high-frequency order book dynamics. *Quantitative Finance*, 16(8):1179–1201.

[10] Bacry, E., Mastromatteo, I., and Muzy, J.-F. (2015). Hawkes processes in finance. *Market Microstructure and Liquidity*, 1(1):1550005, 59 pages.

[11] Bacry, E. and Muzy, J.-F. (2014). Hawkes model for price and trades high-frequency dynamics. *Quantitative Finance*, 14(7):1147–1166.

[12] Bacry, E. and Muzy, J. F. (2016). First- and second-order statistics characterization of Hawkes processes and non-parametric estimation. *IEEE Transactions on Information Theory*, 62(4):2184–2202.

[13] Bauwens, L. and Hautsch, N. (2009). Modelling financial high frequency data using point processes. In Mikosch, T., Kreiß, J.-P., Davis, R. A., and Andersen, T. G., editors, *Handbook of Financial Time Series*, pages 953–979. Springer, Berlin.

[14] Bonart, J. and Gould, M. D. (2017). Latency and liquidity provision in a limit order book. *Quantitative Finance*, 17(10):1601–1616.

[15] Bouchaud, J.-P., Farmer, J. D., and Lillo, F. (2009). How markets slowly digest changes in supply and demand. In Schenk-Hoppé, T. H. R., editor, *Handbook of Financial Markets: Dynamics and Evolution*, Handbooks in Finance, chapter 2, pages 57–160. North-Holland, San Diego.

[16] Bouchaud, J.-P., Gefen, Y., Potters, M., and Wyart, M. (2004). Fluctuations and response in financial markets: the subtle nature of 'random' price changes. *Quantitative Finance*, 4(2):176–190.

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

[18] Bowsher, C. G. (2007). Modelling security market events in continuous time: Intensity based, multivariate point process models. *Journal of Econometrics*, 141(2):876–912.

[19] Brémaud, P. (1981). *Point Processes and Queues: Martingale Dynamics*. Springer, New York.

[20] Brémaud, P. and Massoulié, L. (1994). Imbedded construction of stationary sequences and point processes with a random memory. *Queueing Systems*, 17(1):213–234.

[21] Brémaud, P. and Massoulié, L. (1996). Stability of nonlinear Hawkes processes. *Annals of Probability*, 24(3):1563–1588.

[22] Brown, T. C. and Nair, M. G. (1988). A simple proof of the multivariate random time change theorem for point processes. *Journal of Applied Probability*, 25(1):210–214.

[23] Cartea, Á., Donnelly, R., and Jaimungal, S. (2018). Enhancing trading strategies with order book signals. *Applied Mathematical Finance*, 25(1):1–35.

[24] Cassandras, C. G. and Lygeros, J. (2007). *Stochastic Hybrid Systems*. CRC/Taylor & Francis, Boca Raton.

[25] Chakraborti, A., Toke, I. M., Patriarca, M., and Abergel, F. (2011a). Econophysics review: I. Empirical facts. *Quantitative Finance*, 11(7):991–1012.

[26] Chakraborti, A., Toke, I. M., Patriarca, M., and Abergel, F. (2011b). Econophysics review: II. Agent-based models. *Quantitative Finance*, 11(7):1013–1041.

[27] Challet, D. and Stinchcombe, R. (2001). Analyzing and modeling 1+1d markets. *Physica A: Statistical Mechanics and its Applications*, 300(1–2):285–299.

[28] Chevallier, J. (2015). Mean-field limit of generalized Hawkes processes. Preprint, available at: http://arxiv.org/abs/1510.05620.

[29] Chevallier, J., Caceres, M., Doumic, M., and Reynaud-Bouret, P. (2015). Microscopic approach of a time elapsed neural model. *Mathematical Models and Methods in Applied Sciences*, 25(14):2669–2719.

[30] Çinlar, E. (2011). *Probability and Stochastics.* Springer, New York.

[31] Clinet, S. and Yoshida, N. (2017). Statistical inference for ergodic point processes and application to limit order book. *Stochastic Processes and their Applications*, 127(6):1800–1839.

[32] Cohen, S. N. and Elliott, R. J. (2013). Filters and smoothers for self-exciting Markov modulated counting processes. Preprint, available at: http://arxiv.org/abs/1311.6257.

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

[34] Cont, R. and De Larrard, A. (2011). Price dynamics in a Markovian limit order market. *SIAM Journal on Financial Mathematics*, 4(1):1–25.

[35] Cont, R. and Kukanov, A. (2017). Optimal order placement in limit order markets. *Quantitative Finance*, 17(1):21–39.

[36] Cont, R., Kukanov, A., and Stoikov, S. (2013). The price impact of order book events. *Journal of Financial Econometrics*, 12(1):47–88.

[37] Cont, R., Stoikov, S., and Talreja, R. (2010). A stochastic model for order book dynamics. *Operations Research*, 58(3):549–563.

[38] Da Fonseca, J. and Zaatour, R. (2014). Hawkes process: Fast calibration, application to trade clustering, and diffusive limit. *Journal of Futures Markets*, 34(6):548–579.

[39] Da Fonseca, J. and Zaatour, R. (2015). Clustering and mean reversion in a Hawkes microstructure model. *Journal of Futures Markets*, 35(9):813–838.

[40] Daley, D. J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes. Vol. I.* Springer, New York, second edition.

[41] Daley, D. J. and Vere-Jones, D. (2008). *An Introduction to the Theory of Point Processes. Vol. II.* Springer, New York, second edition.

[42] Dassios, A. and Dong, X. (2014). Stationarity of bivariate dynamic contagion processes. Preprint, available at: http://arxiv.org/abs/1405.5842.

[43] Delattre, S., Fournier, N., and Hoffmann, M. (2016). Hawkes processes on large networks. *Annals of Applied Probability*, 26(1):216–261.

[44] Donier, J. and Bouchaud, J.-P. (2015). Why do markets crash? Bitcoin data offers unprecedented insights. *PLOS ONE*, 10(10):1–11.

[45] Eichler, M., Dahlhaus, R., and Dueck, J. (2017). Graphical modeling for multivariate Hawkes processes with nonparametric link functions. *Journal of Time Series Analysis*, 38(2):225–242.

[46] Embrechts, P., Liniger, T., and Lin, L. U. (2011). Multivariate Hawkes processes: an application to financial data. *Journal of Applied Probability*, 48 A:367–378.

[47] Engle, R. F. and Russell, J. R. (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, 66(5):1127–1162.

[48] Errais, E., Giesecke, K., and Goldberg, L. R. (2010). Affine point processes and portfolio credit risk. *SIAM Journal on Financial Mathematics*, 1(1):642–665.

[49] Filimonov, V. and Sornette, D. (2012). Quantifying reflexivity in financial markets: toward a prediction of flash crashes. *Physical Review E*, 85(5):056108, 9 pages.

[50] Gatheral, J., Jaisson, T., and Rosenbaum, M. (2018). Volatility is rough. *Quantitative Finance*, 18(6):933–949.

[51] Gonzalez, F. and Schervish, M. (2017). Instantaneous order impact and high-frequency strategy optimization in limit order books. *Market Microstructure and Liquidity*, 03(02):1850001, 33 pages.

[52] Gould, M. D. and Bonart, J. (2016). Queue imbalance as a one-tick-ahead price predictor in a limit order book. *Market Microstructure and Liquidity*, 02(02):1650006.

[53] Gould, M. D., Porter, M. A., and Howison, S. D. (2016). The long memory of order flow in the foreign exchange spot market. *Market Microstructure and Liquidity*, 02(01):1650001.

[54] Gould, M. D., Porter, M. A., and Howison, S. D. (2017). Quasi-centralized limit order books. *Quantitative Finance*, 17(6):831–853.

[55] Gould, M. D., Porter, M. A., Williams, S., McDonald, M., Fenn, D. J., and Howison, S. D. (2013). Limit order books. *Quantitative Finance*, 13(11):1709–1742.

[56] Grigelionis, B. (1971). On the representation of integer valued measures by means of stochastic integrals with respect to Poisson measures. *Litovskij Matematicheskij Sbornik*, 11:93–108.

[57] Gu, A. Y. (2005). A survey of the world's top stock exchanges' trading mechanisms and suggestions to the Shanghai stock exchange. Preprint available at http://www.sse.com.cn/aboutus/research/workstation/c/station20050821.pdf.

[58] Hardiman, S. J., Bercot, N., and Bouchaud, J.-P. (2013). Critical reflexivity in financial markets: a Hawkes process analysis. *The European Physical Journal B*, 86(10):442, 9 pages.

[59] Harris, T. E. (1963). *The Theory of Branching Processes*. Springer, Berlin.

[60] Hasbrouck, J. (1991). Measuring the information content of stock trades. *The Journal of Finance*, 46(1):179–207.

[61] Hautsch, N. (2004). *Modelling Irregularly Spaced Financial Data : Theory and Practice of Dynamic Duration Models*. Springer, Berlin.

[62] Hawkes, A. G. (1971). Spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 33(3):438–443.

[63] Hawkes, A. G. and Oakes, D. (1974). A cluster process representation of a self-exciting process. *Journal of Applied Probability*, 11(3):493–503.

[64] Huang, W., Lehalle, C.-A., and Rosenbaum, M. (2015). How to predict the consequences of a tick value change? Evidence from the Tokyo Stock Exchange pilot program. Preprint available at https://arxiv.org/abs/1507.07052.

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

[66] Huang, W. and Rosenbaum, M. (2017). Ergodicity and diffusivity of markovian order book models: A general framework. *SIAM Journal on Financial Mathematics*, 8(1):874–900.

[67] Jacobsen, M. (2006). *Point Process theory and Applications: Marked Point and Piecewise Deterministic Processes*. Birkhäuser, Boston.

[68] Jacod, J. (1975). Multivariate point processes: predictable projection, Radon-Nikodym derivatives, representation of martingales. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 31(3):235–253.

[69] Jacod, J. (1979). *Calcul stochastique et problèmes de martingales.* Springer, Berlin.

[70] Jain, P. K. (2003). Institutional design and liquidity at stock exchanges around the world. Preprint available at https://ssrn.com/abstract=869253.

[71] Jaisson, T. and Rosenbaum, M. (2015). Limit theorems for nearly unstable Hawkes processes. *Annals of Applied Probability*, 25(2):600–631.

[72] Jaisson, T. and Rosenbaum, M. (2016). Rough fractional diffusions as scaling limits of nearly unstable heavy tailed Hawkes processes. *Annals of Applied Probability*, 26(5):2860–2882.

[73] Jovanović, S., Hertz, J., and Rotter, S. (2015). Cumulants of Hawkes point processes. *Physical Review E*, 91(4):42802.

[74] Kallenberg, O. (2002). *Foundations of Modern Probability*. Springer, New York, second edition.

[75] Kallenberg, O. (2017). *Random Measures, Theory and Applications*. Springer, Cham.

[76] Kerstan, J. (1964). Teilprozesse Poissonscher Prozesse. In *Transactions of the Third Prague Conference on Information Theory, Statistical Decision Functions and Random Processes (Liblice, 1962)*, pages 377–403.

[77] Kirchner, M. (2017). An estimation procedure for the Hawkes process. *Quantitative Finance*, 17(4):571–595.

[78] Large, J. (2007). Measuring the resiliency of an electronic limit order book. *Journal of Financial Markets*, 10(1):1–25.

[79] Last, G. (2004). Ergodicity properties of stress release, repairable system and workload models. *Advances in Applied Probability*, 36(2):471–498.

[80] Last, G. and Brandt, A. (1995). *Marked Point Processes on the Real Line: The Dynamical Approach*. Springer, New York.

[81] Laub, P. J., Taimre, T., and Pollett, P. K. (2015). Hawkes processes. Preprint, available at: http://arxiv.org/abs/1507.02822.

[82] Lehalle, C.-A. and Neuman, E. (2017). Incorporating signals into optimal trading. Preprint, available at http://arxiv.org/abs/1704.00847.

[83] Lewis, P. A. W. and Shedler, G. S. (1976). Simulation of nonhomogeneous Poisson processes with log linear rate function. *Biometrika*, 63(3):501–505.

[84] Lillo, F. and Farmer, J. D. (2004). The long memory of the efficient market. *Studies in nonlinear dynamics & econometrics*, 8(3).

[85] Lillo, F., Farmer, J. D., and Mantegna, R. N. (2003). Econophysics: Master curve for price-impact function. *Nature*, 421(6919):129–130.

[86] Lindvall, T. (1988). Ergodicity and inequalities in a class of point processes. *Stochastic Processes and their Applications*, 30(1):121–131.

[87] Lu, X. and Abergel, F. (2018). High-dimensional Hawkes processes for limit order books: modelling, empirical analysis and numerical calibration. *Quantitative Finance*, 18(2):249–264.

[88] Madhavan, A. (1992). Trading mechanisms in securities markets. *The Journal of Finance*, 47(2):607–641.

[89] Maskawa, J.-i. (2007). Correlation of coming limit price with order book in stock markets. *Physica A: Statistical Mechanics and its Applications*, 383(1):90–95.

[90] Maslov, S. and Mills, M. (2001). Price fluctuations from the order book perspective—empirical facts and a simple model. *Physica A: Statistical Mechanics and its Applications*, 299(1–2):234–246.

[91] Massoulié, L. (1998). Stability results for a general class of interacting point processes dynamics, and applications. *Stochastic Processes and their Applications*, 75(1):1–30.

[92] Matthes, K., Kerstan, J., and Mecke, J. (1974). *Infinitely Divisible Point Processes*. Akademie, Berlin.

[93] Meyer, P. A. (1971). Démonstration simplifiée d'un théorème de Knight. In *Séminaire de Probabilités V (Université de Strasbourg)*, pages 191–195, Berlin, Heidelberg. Springer, Berlin.

[94] Morariu-Patrichi, M. (2018). On the weak-hash metric for boundedly finite integer-valued measures. *Bulletin of the Australian Mathematical Society*, 98(2):265–276.

[95] Morariu-Patrichi, M. and Pakkanen, M. S. (2017). Hybrid marked point processes: characterisation, existence and uniqueness. Preprint, available at: http://arxiv.org/abs/1707.06970.

[96] Morariu-Patrichi, M. and Pakkanen, M. S. (2018). State-dependent Hawkes processes and their application to limit order book modelling. Preprint, available at: http://arxiv.org/abs/1809.08060.

[97] Muni Toke, I. (2011). Market making in an order book model and its impact on the spread. In Abergel, F., Chakrabarti, B., Chakraborti, A., and Mitra, M., editors, *Econophysics of Order-driven Markets*, pages 49–64. Springer, Milan.

[98] Ogata, Y. (1978). The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Annals of the Institute of Statistical Mathematics*, 30(1):243–261.

[99] Ogata, Y. (1981). On Lewis' simulation method for point processes. *IEEE Transactions on Information Theory*, 27(1):23–31.

[100] O'Hara, M. and Ye, M. (2011). Is market fragmentation harming market quality? *Journal of Financial Economics*, 100(3):459–474.

[101] Omi, T., Hirata, Y., and Aihara, K. (2017). Hawkes process model with a time-dependent background rate and its application to high-frequency financial data. *Physical Review E*, 96(1):012303, 10 pages.

[102] Ozaki, T. (1979). Maximum likelihood estimation of Hawkes' self-exciting point processes. *Annals of the Institute of Statistical Mathematics*, 31(1):145–155.

[103] Parlour, C. A. (1998). Price dynamics in limit order markets. *Review of Financial Studies*, 11(4):789–816.

[104] Pirrong, C. (1996). Market liquidity and depth on computerized and open outcry trading systems: a comparison of DTB and LIFFE bund contracts. *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, 16(5):519–543.

[105] Potters, M. and Bouchaud, J.-P. (2003). More statistical properties of order books and price impact. *Physica A: Statistical Mechanics and its Applications*, 324(1–2):133–140.

[106] Rambaldi, M., Bacry, E., and Lillo, F. (2017). The role of volume in order book dynamics: a multivariate Hawkes process analysis. *Quantitative Finance*, 17(7):999–1020.

[107] Rambaldi, M., Pennesi, P., and Lillo, F. (2015). Modeling foreign exchange market activity around macroeconomic news: Hawkes-process approach. *Physical Review E*, 91(1):012819, 15 pages.

[108] Roşu, I. (2009). A dynamic model of the limit order book. *Review of Financial Studies*, 22(11):4601–4641.

[109] Sancetta, A. (2018). Estimation for the prediction of point processes with many covariates. *Econometric Theory*, 34(3):598–627.

[110] Schwartz, R. A. and Weber, B. W. (1997). Next-generation securities market systems: An experimental investigation of quote-driven and order-driven trading. *Journal of Management Information Systems*, 14(2):57–79.

[111] Sigman, K. (1995). *Stationary Marked Point Processes: An Intuitive Approach*. Chapman & Hall, New York.

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

[113] Sokol, A. and Hansen, N. R. (2015). Exponential martingales and changes of measure for counting processes. *Stochastic Analysis and Applications*, 33(5):823–843.

[114] Soros, G. (1989). *The Alchemy of Finance*. Wiley, New-York.

[115] Soros, G. (2008). *The Crash of 2008 and What It Means: The New Paradigm for Financial Markets*. PublicAffairs, New York.

[116] Swishchuk, A. (2017). General compound Hawkes processes in limit order books. Preprint, available at: http://arxiv.org/abs/1706.07459.

[117] Swishchuk, A. and Vadori, N. (2017). A semi-Markovian modeling of limit order markets. *SIAM Journal on Financial Mathematics*, 8(1):240–273.

[118] Taranto, D. E., Bormetti, G., Bouchaud, J.-P., Lillo, F., and Tóth, B. (2018). Linear models for the impact of order flow on prices. II. The mixture transition distribution model. *Quantitative Finance*, 18(6):917–931.

[119] Torrisi, G. (2016). Gaussian approximation of nonlinear Hawkes processes. Preprint, available at: http://arxiv.org/abs/1609.08899.

[120] Tóth, B., Lempérière, Y., Deremble, C., de Lataillade, J., Kockelkoren, J., and Bouchaud, J.-P. (2011). Anomalous price impact and the critical nature of liquidity in financial markets. *Physical Review X*, 1(2):21006.

[121] Vinkovskaya, E. (2014). *A Point Process Model for the Dynamics of Limit Order Books*. PhD thesis, Columbia University.

[122] Wang, T., Bebbington, M., and Harte, D. (2012). Markov-modulated Hawkes process with stepwise decay. *Annals of the Institute of Statistical Mathematics*, 64(3):521–544.

[123] Yao, C. and Cassandras, C. G. (2011). Perturbation analysis of stochastic hybrid systems and applications to resource contention games. *Frontiers of Electrical and Electronic Engineering in China*, 6(3):453–467.