



**HAL**  
open science

## Ergodicity and scaling limit of a constrained multivariate Hawkes process

Ban Zheng, François Roueff, Frédéric Abergel

► **To cite this version:**

Ban Zheng, François Roueff, Frédéric Abergel. Ergodicity and scaling limit of a constrained multivariate Hawkes process. *SIAM Journal on Financial Mathematics*, 2014, 5 (1), pp.P. 99-136. 10.1137/130912980 . hal-00777941v2

**HAL Id: hal-00777941**

**<https://hal.science/hal-00777941v2>**

Submitted on 13 Feb 2014

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

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

# Modelling Bid and Ask prices using constrained Hawkes processes

## Ergodicity and scaling limit

Ban Zheng <sup>\*</sup> <sup>†</sup> <sup>‡</sup>      François Roueff <sup>†</sup>      Frédéric Abergel <sup>‡</sup>

February 13, 2014

### Abstract

We introduce a multivariate point process describing the dynamics of the Bid and Ask price of a financial asset. The point process is similar to a Hawkes process, with additional constraints on its intensity corresponding to the natural ordering of the best Bid and Ask prices. We study this process in the special case where the fertility function is exponential, so that the process is entirely described by an underlying Markov chain including the constraint variable. Natural, explicit conditions on the parameters are established that ensure the ergodicity of the chain. Moreover, scaling limits are derived for the integrated point process.

**Keywords:** Point processes. Hawkes processes. Limit order book. Microstructure noise. Bid-Ask Spread. Ergodicity. Scaling limit. Markov model.

**Mathematics Subject Classification:** 37A25, 60F05, 60G55, 60J05.

**JEL Classification:** C62.

---

<sup>\*</sup>Natixis, Equity Markets. E-mail: ban.zheng@melix.net. The authors would like to thank the members of Natixis quantitative research team for fruitful discussions.

<sup>†</sup>Télécom ParisTech (CNRS LTCI, Institut Mines-Télécom). E-mail: francois.roueff@telecom-paristech.fr

<sup>‡</sup>Corresponding author. Chair of Quantitative Finance, Ecole Centrale Paris, MAS Laboratory. E-mail: frederic.abergel@ecp.fr

# 1 Introduction

Modelling the mechanisms of liquidity taking and providing is key to understanding the finer scale, microscopic dynamics of the price of a financial asset. In the context of electronic, order-driven markets, the modelling effort naturally occurs at the order-book level, and the study of order-driven markets has received much attention over the past two decades. On the one hand, extensive statistical studies of the limit order book dynamics and information content have been performed, see e.g. [9, 23, 33, 19, 25, 10, 22, 14, 36]; on the other hand, stochastic models, including equilibrium models, agent-based models and Markov models, have emerged as a mathematical representation of the limit order book, see [31, 32, 17, 15, 1, 16].

It is quite clear from the references above, or better, from a direct inspection of limit order book high frequency data, that a full-fledged order-book model rapidly becomes cumbersome and may sometimes hide simple yet essential mechanisms. One may therefore question the level of complexity required to understand the essential features of the evolution of a financial asset in the high frequency realm. Obviously, a minimal description should model the Bid/Ask spread, since this quantity reveals essential information on liquidity. Such a description should also endogenously account for the interplay between the spread and the mid-price. It is our goal in this paper to introduce a phenomenological, point-process-based description of the joint dynamics of the mid-price and the spread, or equivalently, of the respective best Bid and Ask prices. This description can be viewed as a simplified form of an order-book model with only two limits and infinite liquidity available at each best bid and ask prices, so that its aim is to represent the joint dynamics of the price and the spread in the absence of liquidity costs. Therefore, it can be viewed as an upgrade of the classical modelling approach that addresses only the mid-price dynamics. Actually, we introduce a more general model that can deal with several constraints, a situation that can happen when studying for instance the joint dynamics of the Bid and Ask prices of several tradable assets.

The theory of point processes provides a natural tool to model the dynamics of the price of a financial asset at the level of individual changes, or “tick” level. Bauwens and Hautsch [8] is a comprehensive introduction to the application of point processes in financial time series. Such processes allow one to directly model the arrival of events affecting the price dynamics. Recently, Hawkes processes have been introduced in market microstructure. Hawkes processes belong to the class of self-exciting point processes, where the intensity is driven by a weighted function of the time distance to previous points of the process. These processes originate from the literature in seismology, where they were introduced to model the arrival of replicas in the aftermath of an earthquake. Later, they have been successfully applied to financial markets in the context of default risk modelling [21], contagion across equity markets [2], or foreign exchange modelling [20]. To the best of our knowledge, the first application of Hawkes processes in financial time series is [11]. Since then, there has been a growing interest in using Hawkes processes to model high frequency financial data, see [7, 24, 12, 26, 5, 29, 6]. There are some very convincing reasons to this growing interest: first, Hawkes processes offer a very natural way to model the dependence structures of the arrival of changes in the price of one or several assets, see e.g. [5, 4], leading to nice representations for high frequency volatility and correlation. Moreover, the mutual excitation mechanism built in multivariate Hawkes processes is consistent with phenomena observed in market microstructure, where the interplay between liquidity providing and taking is crucial, see [29] for empirical evidence. Obviously, other important phenomena also affect the price dynamics and liquidity supply: macro-economic news, idiosyncratic news, private information... but they are exogenous and cannot be built directly into a price model. In this work, we do not address that aspect of the modelling, but rather concentrate on the endogenous interplay between orders of various types.

It is noteworthy that previous studies using point processes focus on the description of

the evolution of the mid-price of one or several assets, without incorporating the spread-price interaction. In this contribution, we wish to model the dynamics of the best Bid and Ask prices in a stylized limit order book; and use Hawkes processes to that aim. There is however a main difficulty in using classical Hawkes processes: there is a natural ordering between the best Bid and Ask prices, and they must evolve according to specific constraints. In particular, the Bid/Ask spread is always nonnegative, and is therefore bounded away from zero by its lower bound, the tick size. This means that, as soon as the minimal bid-ask spread is attained, the best bid price cannot increase until the best ask price goes up, and conversely, the best Ask price cannot go down until the best Bid price does. Hence, we add a constraint variable to the classical multivariate Hawkes process to model the dynamics of the best limit prices: in a situation where one of the prices is constrained, the intensity of the arrival process of events that would violate the constraint becomes zero, and stays there until the constraint becomes inactive.

Below is an example of a concrete problem that motivates our theoretical approach: Figure 1 shows a snapshot of the price evolution of a tradable asset, both in physical time and event time. The simplest description of the dynamics of the best Bid and Ask prices

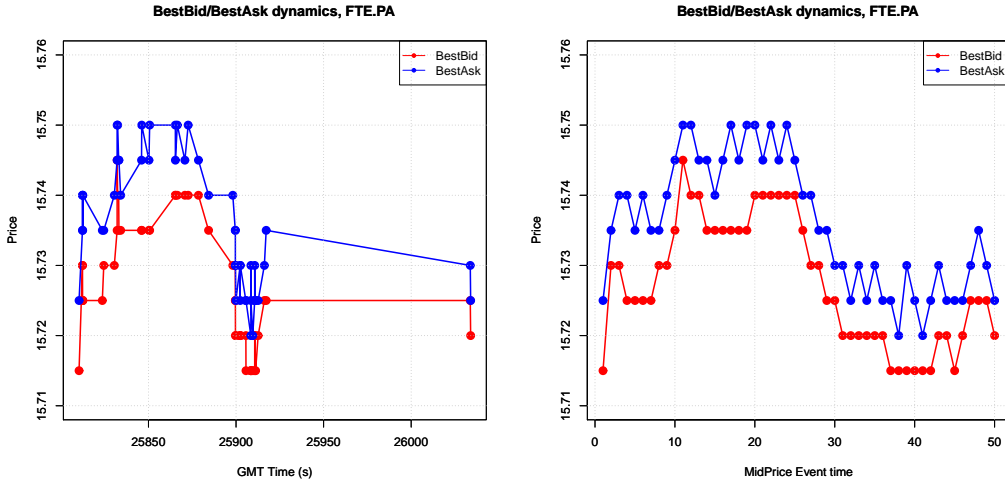


Figure 1: Limit order book of Orange : Evolution of Best Bid and Best Ask prices over a short time window of April 1, 2011. The evolution is displayed in physical time (left) and in event time (right).

is based on four types of events:

- *Event 1*: Best Ask price moves upward one tick,
- *Event 2*: Best Ask price moves downward one tick,
- *Event 3*: Best Bid price moves upward one tick,
- *Event 4*: Best Bid price moves downward one tick.

The arrival of these events is described by a marked point process  $N$ .

As argued above, Hawkes processes offer a sensible framework to model the arrival times of these events. However, since the Best Bid has to remain at least one tick below the Best Ask, we need to introduce a constraint. Namely, defining the spread as  $S = BestAsk - BestBid$  (in ticks), events 2 and 3 cannot happen whenever  $S = 1$ .

Our results show that, under some explicit drift conditions, the spread variable is stationary and the rescaled price behaves like a random walk. More generally, we study

Hawkes processes with multivariate constraints: in a Markovian setting with exponential fertility functions, we provide sufficient conditions that ensure the ergodicity of the embedded chain, and study the scaling limits of the integrated point process. The main methodology is based on establishing geometric drift conditions as detailed in [28, Chapter 15].

The paper is structured as follows: in Section 2, we introduce the constrained multivariate Hawkes processes and present the main assumptions and notation. Section 3 gives the main results on the geometrical ergodicity and the large scale behavior. The application to a limit order book is presented in Section 4. We have gathered the detailed proofs in Section 5. Some useful technical results are postponed to Appendix A.

## 2 Main assumptions and notation

### 2.1 Multivariate Hawkes process with constraints

Let  $p, q$  be two positive integers. We introduce a constrained multivariate Hawkes process defined by a couple  $(\mathbf{S}, \mathbf{N})$ , where  $\{\mathbf{S}(t), t \geq 0\}$  is a multivariate *spread* process taking its values in  $\mathbb{Z}_+^q = \{1, 2, \dots\}^q$  and  $\mathbf{N}$  is a  $p$ -dimensional multivariate point process. We shall describe the dynamics of this model by the conditional intensity of  $\mathbf{N}$  as defined in [13] and an evolution equation for  $\mathbf{S}$ .

We assume that  $\mathbf{S}$  is right-continuous with left limits and denote by  $\mathbf{S}(t-)$  its left limit at time  $t$ . The conditional intensity of the multivariate point process  $\mathbf{N}$  at time  $t$  shall depend on the spread variable  $\mathbf{S}(t-)$ . For notational compactness we will see  $\mathbf{N}$  as a marked point process having arrivals in  $\mathbb{R}$  with marks in  $\{1, \dots, p\}$ . We will denote by  $\mathbf{N}_j, j = 1, \dots, p$  the  $p$  point processes corresponding to each mark  $j$ , namely

$$\mathbf{N} = \sum_{j=1}^p \mathbf{N}_j \otimes \delta_j .$$

In other words, each thinned unmarked process  $\mathbf{N}_j$  is given from  $\mathbf{N}$  by the formula

$$\mathbf{N}_j(g) = \mathbf{N}(g \otimes 1_{\{j\}}) ,$$

for all non-negative measurable function  $g$  defined on  $\mathbb{R}$ . Here and all along the paper, point processes are seen as random point measures and we use the notation  $\xi(g)$  for the integral of a function  $g$  with respect to a measure  $\xi$ . We further use the symbol  $\otimes$  for the product of measures (that is,  $\xi \otimes \zeta(A \times B) = \xi(A)\zeta(B)$ ) and for the tensor product of functions (that is,  $g \otimes h(x, y) = g(x)h(y)$ ). The spread variable  $\mathbf{S}(t-)$  acts on the point process by constraining the set of possible marks for the first event arriving after time  $t$  in  $\mathbf{N}$ . The constraint is added in the usual expression of the conditional intensity of a multivariate Hawkes process. More precisely, for all time  $t \geq 0$  and any mark  $i = 1, \dots, p$ , the conditional intensity of  $\mathbf{N}_i$  given  $\mathbf{N}_{|(-\infty, t)}$  and  $\mathbf{S}(t-)$  is set as

$$\boldsymbol{\mu}(t, i) = \begin{cases} 0 & \text{if } \mathbf{S}(t-) \in \mathbf{A}_i \\ \boldsymbol{\mu}_0(i) + \sum_{j=1}^p \int_{(-\infty, t)} \phi_{i,j}(t-u) \mathbf{N}_j(du) & \text{otherwise ,} \end{cases} \quad (2.1)$$

where  $\boldsymbol{\mu}_0(i)$  is the immigrant rate of mark  $i$ ,  $\phi_{i,j}$  is the fertility rate for producing a mark  $i$  event from a mark  $j$  event, and  $\mathbf{A}_1, \dots, \mathbf{A}_p$  are products of finite subsets of  $\mathbb{Z}_+^q$ ,  $\mathbf{A}_i = A_i(1) \times \dots \times A_i(q)$ .

In turn, an arrival in  $\mathbf{N}$  increases the value of the spread by a value only depending on the mark. More precisely, the spread process  $\mathbf{S}$  satisfies the evolution equation, for all

$t > u \geq 0$ ,

$$\mathbf{S}(t) = \mathbf{S}(u) + \mathbf{N}(1_{(u,t]} \otimes \mathbf{J}), \quad (2.2)$$

where  $\mathbf{J}$  is defined on  $\{1, \dots, p\}$  with values in  $\mathbb{Z}^q$ . In other words, at each arrival in  $\mathbf{N}$  with mark equal to  $i$ ,  $\mathbf{S}$  jump by an increment given by  $\mathbf{J}(i)$ .

Note that equations (2.1) and (2.2) are valid for  $t \geq 0$  and  $t > u \geq 0$  respectively so that the distribution of the process

$$\{(\mathbf{S}(t), \mathbf{N}_j((0, t])), t \geq 0, j = 1, \dots, p\}$$

is defined conditionally to the initial conditions given by  $\mathbf{S}(0-)$  and  $\mathbf{N}$  restricted to  $(-\infty, 0)$ . To summarize, this distribution is parameterized by the constrained sets  $\mathbf{A}_1, \dots, \mathbf{A}_p$ , the immigrants rates  $\boldsymbol{\mu}_0(i)$ ,  $i = 1, \dots, p$ , the cross-fertility rate functions  $\phi_{i,j} : [0, \infty) \rightarrow [0, \infty)$ ,  $i, j = 1, \dots, p$ , and the jump increments  $\mathbf{J}(i) \in \mathbb{Z}^q$  for  $i \in \{1, \dots, p\}$ .

Meanwhile, we define an unconstrained multivariate Hawkes process  $\mathbf{N}'$  with the same fertility rate and immigrant intensity (but without constraints), that is, with conditional intensity on  $[0, \infty)$  satisfying, for  $i = 1, \dots, p$ ,

$$\boldsymbol{\mu}'(t, i) = \boldsymbol{\mu}_0(i) + \sum_{j=1}^p \int_{(-\infty, t)} \phi_{i,j}(t-u) \mathbf{N}'_j(du), \quad t \geq 0,$$

where  $\mathbf{N}'_j = \mathbf{N}'(\cdot \otimes 1_{\{j\}})$  for  $j = 1, \dots, p$ .

The *average fertility* matrix is defined by

$$\aleph = [\alpha_{i,j}]_{i,j=1,\dots,p} \quad \text{with} \quad \alpha_{i,j} = \int_0^\infty \phi_{i,j}(t) dt, \quad 1 \leq i, j \leq p. \quad (2.3)$$

The following assumption ensures the existence of a stationary version of the (unconstrained) multivariate Hawkes process  $\mathbf{N}'$ , see [18, Example 8.3(c)].

**Assumption 1.** *The spectral radius of  $\aleph$  is strictly less than 1.*

Under Assumption 1, we will use the following (well defined) vector

$$\mathbf{u} = (\text{Id}_p - \aleph^T)^{-1} \mathbf{1}_p = \sum_{k \geq 0} \aleph^k \mathbf{1}_p, \quad (2.4)$$

where  $\mathbf{1}_p$  denotes  $p$ -dimensional ones vector,

$$\mathbf{1}_p = \underbrace{[1 \ \cdots \ 1]}_p^T,$$

and  $\text{Id}_p$  denotes the  $p \times p$  identity matrix.

In the following, by convention, we use bold faces symbols only when  $q$  may be larger than one. In the case where  $q$  is set to 1, we shall write  $J$  and  $S_n$  in unbold faces since they are scalar valued and, similarly, we write  $A_i$  in unbold face for the constraint sets.

For a real valued function  $w$  defined on  $\{1, \dots, p\}$ , we shall use the notation

$$\vec{w} = [w(1), \dots, w(p)]^T.$$

If  $w$  is vector-valued, we use the same notation to obtain a matrix. For instance,  $\vec{\mathbf{J}}$  is a  $p \times q$  matrix.

## 2.2 A very special case

When we study the ergodicity of the process, we investigate the joint ergodicity of the spread variable and of the point process. A very special case is obtained by setting

$$\phi_{i,j} \equiv 0, \quad i, j = 1, \dots, p,$$

and  $q = 1$ . In this case, we only need to study the stationarity of the spread variable since the intensities are constant. Suppose moreover that  $J$  only takes values  $+1$  and  $-1$  and the *constraint* sets  $A_i$  are all equal to  $\{1\}$  if  $J(i) = -1$  and to  $\emptyset$  if  $J(i) = 1$ . In this case,  $\{S(t), t \geq 0\}$  is a birth-death process on  $\mathbb{Z}_+$  with constant birth rate  $\mu_+ = \sum_{i \in I_+} \mu_0(i)$  and constant death rate  $\mu_- = \sum_{i \in I_-} \mu_0(i)$ , where  $I_+, I_-$  is the partition of  $\{1, \dots, p\}$  corresponding to the indices where  $J$  takes values  $+1$  and  $-1$ , respectively. By [3, Corollary 2.5],  $(S(t))_{t \geq 0}$  is ergodic if and only if

$$\mu_+ - \mu_- = \vec{J}^T \boldsymbol{\mu}_0 = \sum_{i=1}^p J(i) \mu_0(i) < 0. \quad (2.5)$$

Here and in the following,  $A^T$  denotes the transpose of matrix  $A$ .

## 2.3 A simple order book

In the introduction, we presented a simple order book with only two limit prices, namely the best Bid/Ask prices. We now detail how the constrained multivariate Hawkes process can be used to model this situation. Recall that the four possible events are: Best Ask price moves upward one tick ( $i = 1$ ), Best Ask price moves downward one tick ( $i = 2$ ), Best Bid price moves upward one tick ( $i = 3$ ), Best Bid price moves downward one tick ( $i = 4$ ). Moreover, the spread variable  $S$  is the standard (univariate) Bid-Ask spread. In the formalism of Section 2.1, this case corresponds to  $p = 4$ ,  $q = 1$  and the constrained sets  $A_i$  are defined by

$$A_i = \begin{cases} \emptyset, & i = 1 \text{ or } 4 \\ \{1\}, & i = 2 \text{ or } 3 \end{cases} \quad (2.6)$$

The spread variable  $S$  evolves according to (2.2) with a scalar  $J$  defined by

$$J(i) = \begin{cases} 1, & i = 1 \text{ or } 4 \\ -1, & i = 2 \text{ or } 3 \end{cases} \quad (2.7)$$

Hence the constrained multivariate Hawkes process provides a general framework for modelling this simple order book. If  $\aleph = 0$  as in Section 2.2, Condition (2.5) for the ergodicity of the model becomes

$$\boldsymbol{\mu}_0(2) + \boldsymbol{\mu}_0(3) < \boldsymbol{\mu}_0(1) + \boldsymbol{\mu}_0(4). \quad (2.8)$$

This condition can be interpreted as a drift that forces the mean intensity of decreases of the spread variable  $S$  to be larger than the mean intensity of increases of the spread variable  $S$ . However, in such a special case, given  $\mathbf{N}$  and  $S$  up to time  $t$ , the conditional distribution of  $\mathbf{N}$  restricted to  $(t, \infty)$  only depends on  $S(t)$  and not on the past events. Hence this model has very limited interest for modelling the dynamics of the limit order book. Therefore we wish to investigate the case where  $\aleph$  is non-zero and, in particular, to determine how the ergodicity condition (2.8) should be adapted to this case. This will be answered in Section 4.

## 2.4 The Markov assumption

We shall consider a particular shape of the fertility functions  $\phi_{i,j}$  which implies a Markov property for the model. This allows us to obtain very precise results on the ergodicity of  $(\mathbf{S}, \mathbf{N})$  and is also of interest in applications since the model is parameterized by a restricted set of well understood parameters.

From now on, we suppose that there exists  $\beta > 0$  such that

$$\phi_{i,j}(u) = \alpha_{i,j}\beta e^{-\beta u}, \quad 1 \leq i, j \leq p.$$

This parameter is the reciprocal of a time : the larger  $\beta$  is, the shorter the dependence persists along the time between successive events. We denote a new process defined in the state space  $\mathbf{X} = \mathbb{Z}_+^q \times \mathbb{R}_+^p$  by  $\mathbf{X}(t) = (\mathbf{S}(t), \boldsymbol{\lambda}(t))$  where  $\boldsymbol{\lambda}(t) = [\boldsymbol{\lambda}(t, 1), \dots, \boldsymbol{\lambda}(t, p)]^T$  with, for all  $i \in \{1, \dots, p\}$ ,

$$\boldsymbol{\lambda}(t, i) = \sum_{j=1}^p \int_{(-\infty, t)} \phi_{i,j}(t-u) \mathbf{N}_j(du).$$

**Proposition 2.1.** *The process  $\{\mathbf{X}(t), t \geq 0\}$  is a Markov process.*

This property directly follows from the exponential form of the fertility functions  $\phi_{i,j}$  (see [30, 21, 27] in the unconstrained case). We omit the proof of Proposition 2.1 which is similar to the unconstrained case.

In the following, we shall essentially rely on a Markov chain  $(\mathbf{Z}_n)_{n \geq 0}$  for describing the dynamics of the process  $\{\mathbf{X}(t), t \geq 0\}$ . We now detail the construction of this Markov chain. Let  $\mu_0^0$  denote an arbitrary positive constant (say  $\mu_0^0 = 1$ ). An embedded Markov chain  $(\mathbf{X}_n)_{n \geq 0}$  shall be defined by sampling the continuous time process  $\{\mathbf{X}(t), t \geq 0\}$  at increasing discrete times  $(T_n)_{n \geq 0}$  gathering the arrivals of  $\mathbf{N}'$  (the unconstrained Hawkes process defined in Section 2.1) and the arrivals of an independent homogeneous Poisson point process (PPP)  $n_0$  with intensity  $\mu_0^0$ . Observe that  $\mathbf{S}(t)$  is constant between two consecutive arrivals of the point process  $\mathbf{N}$ . Other sampling points are those generated by  $n_0$  and those present in  $\mathbf{N}'$  but not in  $\mathbf{N}$  (which correspond to the arrivals of  $\mathbf{N}'$  that violate the constraint). To these points we shall assign a mark of new type  $i = 0$ , see the second line in (2.9) below). Hence, almost surely,  $\mathbf{S}$  does not change at arrivals corresponding to marks of this type. The sample time instants with marks of type  $i = 0$  are artificially added to avoid the periodicity of the embedded chain, see the end of the proof of Proposition 5.3.

Each jump of  $\boldsymbol{\lambda}(t)$  corresponds to an arrival in the point process  $\mathbf{N}$ . We denote its mark by  $I_n$ , which takes values in  $\{1, \dots, p\}$ , where  $n$  is the positive integer such that the corresponding jump instant  $t$  equals  $T_n$ . For other sample times  $T_n$ , which thus correspond to the arrivals of  $n_0$ , we set  $I_n = 0$ . Hence the marks of the embedded chain will take values in the set  $\{0, 1, \dots, p\}$ .

Finally, we define  $\Delta_1 = T_1$  and, for each  $n \geq 2$ ,  $\Delta_n = T_n - T_{n-1}$ ,  $\mathbf{S}_n = \mathbf{S}(T_n)$ ,  $\boldsymbol{\lambda}_n = \boldsymbol{\lambda}(T_n)$  and  $\mathbf{X}_n = \mathbf{X}(T_n) = (\mathbf{S}_n, \boldsymbol{\lambda}_n)$ . Then, given  $(\mathbf{S}_0, \boldsymbol{\lambda}_0)$  and  $T_0 = 0$ , we can generate the sequence  $(\Delta_n, \mathbf{S}_n, \boldsymbol{\lambda}_n)$ ,  $n = 0, 1, \dots$  iteratively as follows.

$$\Delta_{n+1} = \min(\Delta_{n+1}^0, \Delta_{n+1}^1, \dots, \Delta_{n+1}^p),$$

where, given  $\mathbf{X}_0, \dots, \mathbf{X}_n$ ,  $\Delta_1, \dots, \Delta_n$  and  $I_1, \dots, I_n$ , the conditional distribution of  $\Delta_{n+1}^0$ ,  $\Delta_{n+1}^1, \dots$  and  $\Delta_{n+1}^p$  is that of positive independent variables whose marginal distributions are determined by hazard rates  $\text{Hr}_0, \text{Hr}_1, \dots, \text{Hr}_p$  defined for each  $i$  and  $t \geq 0$  by

$$\text{Hr}_i(t) = \begin{cases} (\mu_0(i) + \boldsymbol{\lambda}_n(i)e^{-\beta t}) 1_{\mathbf{A}_i^c}(\mathbf{S}_n), & i > 0, \\ \mu_0^0 + \sum_{j=1}^p 1_{\mathbf{A}_j}(\mathbf{S}_n) [\mu_0(j) + \boldsymbol{\lambda}_n(j)e^{-\beta t}], & i = 0. \end{cases} \quad (2.9)$$



Then, the event type at time  $T_{n+1}$  is

$$I_{n+1} = \arg \min_{i \in \{0, \dots, p\}} (\Delta_{n+1}^i),$$

the spread variable at time  $T_{n+1}$  is given by

$$\mathbf{S}_{n+1} = \mathbf{S}(T_{n+1}) = \mathbf{S}_n + \mathbf{J}_o(I_{n+1}), \quad (2.10)$$

where  $\mathbf{J}_o$  is the extension of  $\mathbf{J}$  to  $\{0, 1, \dots, p\}$  defined by

$$\mathbf{J}_o(i) = \begin{cases} \mathbf{J}(i) & \text{if } i \geq 1 \\ \mathbf{0}_q & \text{if } i = 0. \end{cases} \quad (2.11)$$

and, for all  $i \in \{1, \dots, p\}$ , the self-excitation intensities at time  $T_{n+1}$  are given by

$$\lambda_{n+1}(i) = \lambda(T_{n+1}, i) = \lambda_n(i)e^{-\beta\Delta_{n+1}} + \beta\alpha_{i, I_{n+1}}. \quad (2.12)$$

Observe that, for all  $T_n \leq t < T_{n+1}$ , the continuous time process is interpolated as

$$\begin{aligned} \mathbf{S}(t) &= \mathbf{S}_n \\ \lambda(t, i) &= \lambda_n(i)e^{-\beta t}. \end{aligned}$$

For any probability distribution  $\nu$  on  $\mathbf{X} = \mathbb{Z}_+^q \times \mathbb{R}_+^p$ , we denote by  $\mathbb{P}^\nu$  the probability corresponding to the initial distribution  $\nu$  at time  $t = 0$ . Notation  $\mathbb{E}^\nu$  corresponds to the corresponding expectation. If  $\nu$  is a Dirac distribution at point  $\mathbf{x} \in \mathbf{X}$ , we will simply write  $\mathbb{P}^\mathbf{x}$  and  $\mathbb{E}^\mathbf{x}$ .

The above facts thus imply that  $\mathbf{X}_n = (\mathbf{S}_n, \lambda_n)$ ,  $\mathbf{Y}_n = (I_n, \mathbf{S}_n, \lambda_n)$  and  $\mathbf{Z}_n = (\Delta_n, I_n, \mathbf{S}_n, \lambda_n)$ , defined for  $n = 0, 1, \dots$  are Markov chains respectively valued in  $\mathbf{X} = \mathbb{Z}_+^q \times \mathbb{R}_+^p$ ,  $\mathbf{Y} = \{0, 1, \dots, p\} \times \mathbb{Z}_+^q \times \mathbb{R}_+^p$  and  $\mathbf{Z} = \mathbb{R}_+ \times \{0, \dots, p\} \times \mathbb{Z}_+^q \times \mathbb{R}_+^p$ . We shall denote by  $Q$ ,  $\tilde{Q}$  and  $\bar{Q}$  the transition kernels of these Markov processes, respectively. We shall use these three transition kernels, depending on the context. As we shall see, the ergodicity of  $\tilde{Q}$  and  $\bar{Q}$  essentially follow from that of  $Q$ . However, the kernel  $\tilde{Q}$  turned out to be easier to handle for proving the ergodicity when  $q \geq 2$ , see Theorem 3.7 below. This is the reason why we introduced the chain  $(\mathbf{Y}_n)_{n \geq 0}$ . The chain  $(\mathbf{Z}_n)_{n \geq 0}$  was introduced because the processes  $\mathbf{N}([0, t])$  and  $\mathbf{S}(t)$  can be deterministically expressed for all  $t \in \mathbb{R}_+$  using this chain. This fact is used in Theorem 3.9.

**Remark 1.** *Observe that, from the description above, the conditional distribution of  $(\Delta_1, I_1, \mathbf{S}_1, \lambda_1)$  given  $(\Delta_0, I_0, \mathbf{S}_0, \lambda_0)$  does not depend on  $(\Delta_0, I_0)$ . In particular, the whole path  $\{\mathbf{Z}_n, n \geq 1\}$  only depends on the initial condition set on  $\mathbf{X}_0$  and we may write, for all  $\mathbf{z} = (\delta, i, \mathbf{x}) \in \mathbf{Z}$  and all  $A \in \mathcal{B}(\mathbf{Z})$ ,*

$$\bar{Q}(\mathbf{z}, A) = \mathbb{P}^\mathbf{x}((\Delta_1, I_1, \mathbf{X}_1) \in A). \quad (2.13)$$

*Clearly the same remark holds for  $\tilde{Q}$ , namely,  $\tilde{Q}(\mathbf{y}, A) = \mathbb{P}^\mathbf{x}((I_1, \mathbf{X}_1) \in A)$  for all  $\mathbf{y} = (i, \mathbf{x}) \in \mathbf{Y}$  and  $A \in \mathcal{B}(\mathbf{Y})$ .*

Because the ergodicity of  $Q$  will be proved using an induction on the number of constraints  $q$ , we need to introduce further Markov chains and transition kernels. To initiate the induction, we define the Markov chain  $\{(\tilde{I}_n, \tilde{\lambda}_n), n \geq 0\}$  as the chain valued in  $\{0, 1, \dots, p\} \times \mathbb{R}_+^q$  that starts at the same state as  $(I_0, \lambda_0)$  but with transition kernel  $\tilde{Q}$ , which is defined as  $\tilde{Q}$  but with all the  $\mathbf{A}_i$ 's replaced by the empty set and without the spread variable. We will call this chain the *unconstrained Hawkes embedded chain*, since it is associated to a classical (unconstrained) multivariate Hawkes point process. This chain corresponds to the case  $q = 0$ .

**Remark 2.** As in the case  $q \geq 1$ , the conditional distribution of  $\{(\check{I}_n, \check{\lambda}_n), n \geq 1\}$  given the initial condition  $(\check{I}_0, \check{\lambda}_0)$  does not depend on  $\check{I}_0$ . Hence we will use the notation  $\mathbb{E}^\ell$  or  $\mathbb{P}^\ell$  to underline this fact, for instance, for any  $\mathbf{y} = (i, \ell) \in \{0, 1, \dots, p\} \times \mathbb{R}_+^p$ ,  $j \in \{0, 1, \dots, p\}$  and Borel subset  $A \subset \mathbb{R}_+^p$ ,

$$\check{Q}(\mathbf{y}, \{j\} \times A) = \mathbb{P}^\ell(\check{I}_1 = j, \check{\lambda}_1 \in A).$$

Furthermore, for any  $\mathcal{J} \subseteq \{1, \dots, q\}$ , we denote by  $\check{Q}^{(-\mathcal{J})}$  the transition kernel defined on  $\{0, 1, \dots, p\} \times \mathbb{Z}_+^{q-\#\mathcal{J}} \times \mathbb{R}_+^p$  defined as the transition kernel  $Q$  but without the spread variable  $\mathbf{S}_j$ ,  $j \in \mathcal{J}$  and their corresponding constraint sets  $\mathbf{A}_1(j), \dots, \mathbf{A}_p(j)$ . In particular, we have  $\check{Q}^{(-\emptyset)} = \check{Q}$  and  $\check{Q}^{(-\{1, \dots, q\})} = \check{Q}$ . Meanwhile, for any  $\mathcal{J} \subseteq \{1, \dots, q\}$ , we denote by  $\check{Q}^{(+\mathcal{J})}$  the transition kernel defined on  $\{0, 1, \dots, p\} \times \mathbb{Z}_+^{\#\mathcal{J}} \times \mathbb{R}_+^p$  defined as the transition kernel  $Q$  with the spread variable  $\mathbf{S}_j$ ,  $j \in \mathcal{J}$  and their corresponding constraints  $\mathbf{A}_1(j), \dots, \mathbf{A}_p(j)$ . In particular  $\check{Q}^{(+\{1, \dots, q\})} = \check{Q}$  and  $\check{Q}^{(+\emptyset)} = \check{Q}$ .

### 3 Main results

We now present the main results of this work. All the results will rely on the Markov assumption introduced in Section 2.4. Under this Markov framework, in Section 3.1, we show that the kernel  $Q$  and  $\check{Q}$  defined in Section 2.4 are  $\psi$ -irreducible and aperiodic. We provide a partial drift condition in Section 3.2 and then prove that  $\check{Q}$  is  $V$ -geometrically ergodic in Section 3.3, where  $V$  is unbounded off petite sets. Then we study the ergodicity in the case  $q = 1$  in Section 3.4. The general case is presented in Section 3.5. Finally, we determine the scaling limit of the point process in *physical time* in Section 3.6.

#### 3.1 Irreducibility

Because of the constraints sets  $\mathbf{A}_i$  and the function  $\mathbf{J}$ , the path of the process  $\mathbf{S}$  cannot evolve arbitrarily. The following definitions will be useful.

**Definition 1.** Let  $m$  be a positive integer and  $\mathbf{s} \in \mathbb{Z}_+^q$ , the set of admissible paths  $\mathcal{A}_m(\mathbf{s})$  is defined as the set of  $(j_1, \dots, j_m) \in \{0, \dots, p\}^m$  such that

$$\mathbf{s} \in \mathbf{A}_{j_1}^c, \mathbf{s} + \mathbf{J}_o(j_1) \in \mathbf{A}_{j_2}^c, \dots, \mathbf{s} + \sum_{n=1}^{m-1} \mathbf{J}_o(j_n) \in \mathbf{A}_{j_m}^c. \quad (3.1)$$

An admissible path  $(j_1, \dots, j_m) \in \mathcal{A}_m(\mathbf{s})$  implies that, for any  $i = 0, 1, \dots, p$  and  $\ell \in (0, \infty)^p$ , given  $\mathbf{Y}_0 = (i, \mathbf{s}, \ell)$  the conditional probability to have  $I_1 = j_1, \dots, I_m = j_m$  (and thus, by (2.10)  $\mathbf{S}_k = \mathbf{s} + \sum_{n=1}^k \mathbf{J}_o(j_n)$  for all  $k = 1, 2, \dots, m$ ) is positive.

The main assumption of this section is the following one.

**Assumption 2.** The fertility matrix  $\aleph$  defined in (2.3) is invertible. Moreover, if  $q \geq 1$ , there exists  $\mathbf{s}_o \in \mathbb{Z}_+^q$  such that the following assertion holds. For all integer  $K \geq 1$ , there exists an integer  $m \geq p + 1$  such that, for all  $\mathbf{s} \in \{1, \dots, K\}^q$ , there exists an admissible path  $(j_1, \dots, j_m) \in \mathcal{A}_m(\mathbf{s})$  such that  $\mathbf{s} + \sum_{n=1}^m \mathbf{J}_o(j_n) = \mathbf{s}_o$  and  $\{j_{m-p+1}, \dots, j_m\} = \{1, \dots, p\}$  (possibly in a different order).

We say that  $C$  is an  $(m, \epsilon, \nu)$ -small set for the kernel  $Q$  if, for all  $\mathbf{x} \in C$  and  $A \in \mathcal{B}(X)$ ,

$$Q^m(\mathbf{x}, A) \geq \epsilon \nu(A),$$

where  $m$  is a positive integer,  $\epsilon > 0$  and  $\nu$  a probability measure. Following [28], the existence of small sets is related to  $\psi$ -irreducibility. The assumption on the invertibility of  $\aleph$  will be useful for proving the  $\psi$ -irreducibility of the embedded Markov chain. The second assumption says that there exists an admissible path (see Definition 1) with the last  $p$  steps containing the  $p$  different marks. We will check this condition for the application of the constrained Hawkes process to a limit order book in Section 4. It will be used to establish the existence of small sets for  $Q$ , see Proposition 5.3. Most of the time, we will rather refer to the notion of petite sets, which is a slight extension of small sets. Following [28], we say that  $A$  is a petite set for the kernel  $Q$  if, for some probability measure  $a$  on  $\{0, 1, 2, \dots\}$ ,  $\epsilon > 0$  and some probability measure  $\nu$  on  $\mathbf{X}$ , it is a  $(1, \epsilon, \nu)$ -small set for the  $a$ -sampled chain which has kernel  $\sum_{m \geq 0} a(m)Q^m$ .

We can now state the main result of this section.

**Theorem 3.1.** *Let  $Q$  and  $\tilde{Q}$  be the transition kernels defined in Section 2.4 on the spaces  $\mathbb{Z}_+^q \times \mathbb{R}_+^p$  and  $\{0, 1, \dots, p\} \times \mathbb{Z}_+^q \times \mathbb{R}_+^p$ , respectively, with  $p \geq 1$  and  $q \geq 0$ . Suppose moreover that Assumption 2 holds. Then the kernels  $Q$  and  $\tilde{Q}$  are aperiodic and  $\psi$ -irreducible. Moreover for all  $K \geq 1$  and  $M > 0$ ,  $\{1, \dots, K\}^q \times (0, M]^p$  and  $\{0, \dots, p\} \times \{1, \dots, K\}^q \times (0, M]^p$  are petite sets for  $Q$  and  $\tilde{Q}$ , respectively.*

The proof is postponed to Section 5.1.

### 3.2 Partial drift condition

We now derive a partial drift condition that will be useful to obtain a “complete” drift condition on the process  $\{\mathbf{X}(t), t \geq 0\}$ . In the following result, the drift condition is partial in the sense that it does not control  $\mathbf{S}_n$ . It only says that, under Assumption 1, independently of the process  $\mathbf{S}_n$ ,  $\boldsymbol{\lambda}_n$  should not have large excursions away of a compact set.

**Proposition 3.2.** *Let  $Q$  be the kernel defined in Section 2.4 on the space  $\mathbb{Z}_+^q \times \mathbb{R}_+^p$  with  $p \geq 1$  and  $q \geq 0$ . Suppose that Assumption 1 holds. Then there exists  $\gamma > 0$ ,  $\theta \in (0, 1)$ ,  $M > 0$  and  $b > 0$  such that, for all  $\mathbf{s} \in \mathbb{Z}_+^q$  and  $\boldsymbol{\ell} \in \mathbb{R}_+^p$ , we have*

$$[Q(1_{\mathbb{Z}_+^q} \otimes V_{1,\gamma})](\mathbf{s}, \boldsymbol{\ell}) \leq \theta V_{1,\gamma}(\boldsymbol{\ell}) + b1_{(0,M]^p}(\boldsymbol{\ell}), \quad (3.2)$$

where

$$V_{1,\gamma}(\boldsymbol{\ell}) = e^{\gamma \mathbf{u}^T \boldsymbol{\ell}}, \quad (3.3)$$

with  $\mathbf{u}$  defined as in (2.4).

The proof is postponed to Section 5.2.

### 3.3 The case $q = 0$

If  $q = 0$ , the process  $\mathbf{N}$  is a standard (unconstrained) multivariate Hawkes process. In this case the process  $\{\boldsymbol{\lambda}_n, n \geq 1\}$  is a Markov chain (the spread variable  $\mathbf{S}_n$  vanishes) and the “partial” drift condition of Proposition 3.2 becomes a “complete” drift condition for this chain. Moreover, if  $q = 0$  in Theorem 3.1, Assumption 2 boils down to assuming that the fertility matrix  $\aleph$  defined in (2.3) is invertible. We also note that the sublevel sets of  $V_{1,\gamma}$  are petite sets by Theorem 3.1. Hence we obtain that  $\{\tilde{\boldsymbol{\lambda}}_n, n \geq 0\}$  is  $V_{1,\gamma}$ -geometrically ergodic, see [28, Chapter 15] (and also [28, Theorem 9.1.8] to get the Harris recurrence). In fact the same result holds on the extended chain  $\{(\tilde{I}_n, \tilde{\boldsymbol{\lambda}}_n), n \geq 0\}$ , whose transition kernel has been denoted by  $\tilde{Q}$  in Section 2.4.

**Proposition 3.3.** *Let  $\check{Q}$  be the transition kernel defined in Section 2.4 on the space  $\{0, 1, \dots, p\} \times \mathbb{R}_+^p$ . Suppose that  $\aleph$  is invertible and that Assumption 1 holds. Then there exists  $\gamma > 0$  such that  $\check{Q}$  is  $(1_{\{0, \dots, p\}} \otimes V_{1, \gamma})$ -geometrically ergodic, where  $V_{1, \gamma}$  is defined in (3.3).*

*Proof.* As explained above, Theorem 3.1 and Proposition 3.2 apply to the case  $q = 0$  and, as a consequence,  $\{\check{\lambda}_n, n \geq 0\}$  is a  $V_{1, \gamma}$ -geometrically ergodic. This conclusion also applies for the kernel  $\check{Q}$  of the chain  $\{(\check{I}_n, \check{\lambda}_n), n \geq 0\}$  with  $V_{1, \gamma}$  replaced by  $(1_{\{0, \dots, p\}} \otimes V_{1, \gamma})$  because the conditional probability of  $(\check{I}_1, \check{\ell}_1)$  given  $(\check{I}_0, \check{\ell}_0)$  does not depend on  $\check{I}_0$ .  $\square$

Having proved the ergodicity of  $\check{Q}$ , we can establish the following result on its stationary distribution. It will be used to obtain the case  $q = 1$ .

**Corollary 3.4.** *Under the same assumptions as Proposition 3.3, denote by  $\tilde{\pi}$  the stationary distribution of  $\check{Q}$ . Let  $\mathbf{w} : \{1, \dots, p\} \rightarrow \mathbb{R}$ . Let  $\mathbf{w}_o$  be the extension of  $\mathbf{w}$  to  $\{0, 1, \dots, p\}$  obtained by setting  $\mathbf{w}_o(0) = 0$ . Then we have*

$$\tilde{\pi}(\mathbf{w}_o \otimes 1_{\mathbb{R}_+^p}) = \frac{\overrightarrow{\mathbf{w}}^T (\text{Id}_p - \aleph)^{-1} \boldsymbol{\mu}_0}{\boldsymbol{\mu}_0^0 + \mathbf{1}_p^T (\text{Id}_p - \aleph)^{-1} \boldsymbol{\mu}_0}. \quad (3.4)$$

The proof is postponed to Section 5.3.

### 3.4 The case $q = 1$

We shall further denote

$$s^* = 1 + \max \left( \bigcup_{j=1, \dots, p} A_j \right). \quad (3.5)$$

In the case  $q = 1$ , we need a drift function that applies to  $\mathbf{X}_n = (S_n, \boldsymbol{\lambda}_n)$  which is unbounded off petite sets, that is, the drift function must diverge as at least one component of  $\mathbf{X}_n$  goes to infinity, while, in Proposition 3.2 the drift function goes to infinity only when one of the components of  $\boldsymbol{\lambda}_n$  goes to infinity. To this end we define

$$V_{0, \gamma}(s) = e^{\gamma s}. \quad (3.6)$$

We introduce the following assumption.

**Assumption 3.** *The following inequality holds*

$$\overrightarrow{\mathbf{J}}^T (\text{Id}_p - \aleph)^{-1} \boldsymbol{\mu}_0 < 0. \quad (3.7)$$

Note that Condition (2.5) corresponds to Assumption 3 in the special case  $\aleph = 0$ .

**Theorem 3.5.** *Let  $q = 1$  and  $p \geq 1$  and suppose that Assumption 1, 2 and 3 hold. Let  $\check{Q}$  be the kernel defined in Section 2.4 on the space  $\{0, 1, \dots, p\} \times \mathbb{Z}_+ \times \mathbb{R}_+^p$  with  $p \geq 1$ . Then, for all  $\gamma_1 > 0$  small enough, there exists  $\gamma_0^* > 0$  such that, for all  $\gamma_0 \in (0, \gamma_0^*]$ ,  $\check{Q}$  is  $(1_{\{0, \dots, p\}} \otimes V_{0, \gamma_0} \otimes V_{1, \gamma_1})$ -geometrically ergodic, where  $V_{0, \gamma_0}$  and  $V_{1, \gamma_1}$  are defined in (3.6) and (3.3).*

The proof of Theorem 3.5 is omitted as it is a particular case of Theorem 3.7, which is proved in Section 5.5.

Assumptions 1 and 2 appear to be very mild. Assumption 1 is related to the stability of the underlying unconstrained Hawkes process. Assumption 2 is used to obtain small sets for the chains  $Q$  and  $\check{Q}$ . A natural question is to ask whether Assumption 3  $\overrightarrow{\mathbf{J}}^T (\text{Id}_p - \aleph)^{-1} \boldsymbol{\mu}_0 < 0$  is sharp. The following theorem partially answers to this question.

**Theorem 3.6.** Let  $\{(S_n, \lambda_n), n \geq 0\}$  be the Markov chain on the space  $\mathbb{Z}_+ \times \mathbb{R}_+^p$  with transition kernel  $Q$  defined in Section 2.4. Suppose that Assumptions 1 and 2 hold. Then  $Q$  is  $\psi$ -irreducible. Moreover the two following assertions holds.

- (i) If  $\psi(\{s^*, s^* + 1, \dots\} \times (0, \infty)^p) = 0$ , then  $Q$  is  $(1_{\{1, \dots, s^*-1\}} \otimes V_{1, \gamma_1})$ -geometrically ergodic, where  $s^*$  is defined by (3.5).
- (ii) Otherwise, if

$$\vec{J}^T (\text{Id}_p - \aleph)^{-1} \mu_0 > 0, \quad (3.8)$$

then  $Q$  is transient.

The proof is postponed to Section 5.4.

**Remark 3.** The fact that  $Q$  is  $\psi$ -irreducible follows from Theorem 3.1. Then, by definition of  $\nu$  in the proof of Proposition 5.3, Assumption 2 implies that  $\psi(\{s_0\} \times (0, \infty)^p) > 0$ . Thus if  $s_0$  in Assumption 2 satisfies  $s_0 \geq s^*$ , then the case (i) does not happen.

### 3.5 General case

To obtain a drift function in the general case  $q \geq 1$ , we shall use the function  $V_{0, \gamma}$  defined by (3.6) applied multiplicatively to each component of  $\mathbf{S}_n$  and the function  $V_{1, \gamma_1}$  defined by (3.3) applied to intensity  $\lambda_n$ .

We shall see that the proof of the ergodicity for  $q \geq 1$  relies on an induction on  $q$ , see the details in Section 5.5. For  $q = 1$  the induction applies under the simple Assumption 3 because, using Corollary 3.4, it allows us to compute some moment under the stationary distribution of the unconstrained chain (with kernel  $\tilde{Q}$ ). In general the induction is more involved. We shall need some additional notation. For any set  $\mathcal{J} \subseteq \{1, \dots, q\}$ , we define  $J^{\mathcal{J}} : \{1, \dots, p\} \rightarrow \mathbb{Z}$  by

$$J^{\mathcal{J}}(i) = \sum_{j \in \mathcal{J}} \mathbf{J}_j(i). \quad (3.9)$$

We further denote by  $J_o^{\mathcal{J}}$  the extension of  $J^{\mathcal{J}}$  on  $\{0, \dots, p\}$  defined by  $J_o^{\mathcal{J}}(0) = 0$ .

**Theorem 3.7.** Let  $q \geq 1$  and  $p \geq 1$ , and suppose that Assumption 1 and 2 hold. Let  $\tilde{Q}$  be the kernel defined in Section 2.4 on the space  $\{0, 1, \dots, p\} \times \mathbb{Z}_+^q \times \mathbb{R}_+^p$ . Define, for all  $\mathcal{J} \subset \{1, \dots, q\}$ , the kernel  $\tilde{Q}^{(+\mathcal{J})}$  as in Section 2.4. Check the following conditions in this order :

**for**  $k = 1, 2, 3, \dots, q$  **do**

**for**  $\mathcal{J} \subset \{1, \dots, q\}$  such that  $\#\mathcal{J} = k$  **do**  
    Check that

$$\left( \vec{J}^{\mathcal{J}} \right)^T (\text{Id}_p - \aleph)^{-1} \mu_0 < 0. \quad (3.10)$$

**for**  $\mathcal{J}' \subset \mathcal{J}$  with  $\mathcal{J}' \notin \{\emptyset, \mathcal{J}\}$  **do**  
    Check that

$$\tilde{\pi}^{(+\mathcal{J}')} \left[ J_o^{\mathcal{J}} \otimes 1_{\mathbb{Z}_+^q \times \mathbb{R}_+^p} \right] < 0. \quad (3.11)$$

**end**

Then, for all  $\gamma_1 > 0$  small enough, there exists  $\gamma_0^* > 0$  such that for all  $\gamma_0 \in (0, \gamma_0^*]$ ,  $\tilde{Q}^{(+\mathcal{J})}$  is  $(1_{\{0, \dots, p\}} \otimes V_{0, \gamma_0}^{\otimes (\#\mathcal{J})} \otimes V_{1, \gamma_1})$ -geometrically ergodic. We denote by  $\tilde{\pi}^{(+\mathcal{J})}$  its stationary distribution.

**end**

**end**

Then, for all  $\gamma_1 > 0$  small enough, there exists  $\gamma_0^* > 0$  such that for all  $\gamma_0 \in (0, \gamma_0^*]$ ,  $\tilde{Q}$  is  $(1_{\{0, \dots, p\}} \otimes V_{0, \gamma_0}^q \otimes V_{1, \gamma_1})$ -geometrically ergodic.

The proof is postponed to the end of Section 5.5.

**Remark 4.** By Corollary 3.4, Condition (3.10) is equivalent to (3.11) with  $\mathcal{J}' = \emptyset$ .

**Remark 5.** Observe that Condition (3.11) makes sense at this step. Indeed, since  $\#\mathcal{J} = k$ , and  $\#\mathcal{J}' < k$ , we have already checked that  $\tilde{Q}^{(+\mathcal{J}'})$  is geometrically ergodic. Thus the stationary distribution  $\tilde{\pi}^{(+\mathcal{J}'})$  is well defined. In practice, one may check this assumption using Monte Carlo simulations, since the law of large number holds in this case (see [28, Theorem 17.1.7]).

### 3.6 Scaling limit

In the application of this model to the limit order book, an increment of the best bid price or the best ask price at time  $T_k$  is specified by the mark  $I_k \in \{0, \dots, p\}$ , see the details in Section 4. It is interesting to investigate the microscopic behavior of the mid-price (mean of the best bid price and the best ask price) at large scales, see [5] for the non-constrained case.

Suppose that the Assumptions of Theorem 3.7 hold, then  $\tilde{Q}$  is  $(1_{\{0, \dots, p\}} \otimes V_{0, \gamma_0}^q \otimes V_{1, \gamma_1})$ -geometrically ergodic for some  $\gamma_0, \gamma_1 > 0$  and  $V_{0, \gamma_0}$  and  $V_{1, \gamma_1}$  defined in (3.6) and (3.3). Applying the results of [28, Theorem 17.4.2], we obtain a functional central limit theorem (FCLT) in *discrete time* for the chain  $\tilde{Q}$ . To obtain a result in (physical) *continuous time*, we actually need a functional CLT for the kernel  $\tilde{Q}$  defined in Section 2.4. Hence we must first check that  $\tilde{Q}$  is geometrically ergodic.

Let us define  $V_{2, \gamma} : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  by

$$V_{2, \gamma}(\delta) = e^{\gamma \delta}. \quad (3.12)$$

We provide the ergodicity property of  $\tilde{Q}$  in the following corollary.

**Corollary 3.8.** *Let  $q, p \geq 1$  and  $\tilde{Q}$  be the kernel defined in Section 2.4 on the space  $\mathbb{R}_+ \times \{0, 1, \dots, p\} \times \mathbb{Z}_+^q \times \mathbb{R}_+^p$ . Suppose that the Assumptions of Theorem 3.7 hold. Then, for all  $\gamma_1 > 0$  small enough, there exists  $\gamma_0^* > 0$  such that for all  $\gamma_0 \in (0, \gamma_0^*)$  and all  $\gamma_2 > 0$  small enough,  $\tilde{Q}$  is  $(V_{2, \gamma_2} \otimes 1_{\{0, \dots, p\}} \otimes V_{0, \gamma_0}^q \otimes V_{1, \gamma_1})$ -geometrically ergodic.*

The proof is postponed to Section 5.6.

By Corollary 3.8,  $\tilde{Q}$  admits a stationary distribution  $\tilde{\pi}$ . We shall use the notation  $\bar{\mathbb{E}}$  for the expectation under the stationary distribution.

Let now  $g : \mathbf{Z} \rightarrow \mathbb{R}$  be such that for all  $\gamma_0, \gamma_1, \gamma_2 > 0$ ,

$$\sup_{\mathbf{z} \in \mathbf{Z}} \frac{g(\mathbf{z})}{[V_{2, \gamma_2} \otimes 1_{\{0, \dots, p\}} \otimes V_{0, \gamma_0}^q \otimes V_{1, \gamma_1}](\mathbf{z})} < \infty. \quad (3.13)$$

By Corollary 3.8, under this condition, we have  $\tilde{\pi}(|g|) < \infty$ . Let  $\bar{g} = g - \tilde{\pi}(g)$ . Using Corollary 3.8 and [28, Theorem 17.4.4], for all  $\gamma_0, \gamma_1, \gamma_2 > 0$ , there exists  $R > 0$  such that the Poisson equation  $\hat{g} - \tilde{Q}(\hat{g}) = \bar{g}$  (see [28, Chapter 17]) admits a solution  $\hat{g}$  satisfying the bound  $|\hat{g}| \leq R ((V_{2, \gamma_2} \otimes 1_{\{0, \dots, p\}} \otimes V_{0, \gamma_0}^q \otimes V_{1, \gamma_1}) + 1)$ . Moreover, we have, for all  $\mathbf{z} \in \mathbf{Z}$ ,

$$\hat{g}(\mathbf{z}) = \sum_{k=0}^{\infty} \tilde{Q}^k(\mathbf{z}, \bar{g}).$$

It also follows that  $\tilde{\pi}(|\hat{g}|^2) < \infty$ . Let us denote the linearly interpolated partial sums of  $g(\mathbf{Z}_n)$  by

$$s_n(t, g) = \Sigma_{[nt]}(\bar{g}) + (nt - [nt]) [\Sigma_{[nt]+1}(\bar{g}) - \Sigma_{[nt]}(\bar{g})],$$

where  $\Sigma_k(h) = \sum_{j=1}^k h(\mathbf{Z}_j)$ . Applying [28, Theorem 17.4.4], if the (nonnegative) constant

$$\sigma_g^2 := \bar{\pi} (\hat{g}^2 - \{\overline{Q\hat{g}}\}^2) \quad (3.14)$$

is strictly positive, we have, for any initial distribution, as  $n \rightarrow \infty$ ,

$$(n\sigma_g^2)^{-1/2} s_n(t, g) \xrightarrow{d} B_t, \quad (3.15)$$

where  $\xrightarrow{d}$  denotes the weak convergence. Here, the weak convergence holds in  $C([0, 1])$  (the space of continuous functions defined on  $[0, 1]$ ) and  $B$  denotes a standard Brownian motion on  $[0, 1]$ .

**Remark 6.** Note that  $\sigma_g^2$  may not be strictly positive. An interesting example is given by  $g = 1_{\mathbb{R}_+} \otimes \mathbf{J}_o \otimes 1_{\mathbb{Z}_+^q} \otimes 1_{\mathbb{R}_+^p}$ . In this case, we have  $\Sigma_n(g) = \mathbf{S}_n - \mathbf{S}_0$  and thus

$$\Sigma_n(\bar{g}) = \mathbf{S}_n - \mathbf{S}_0 - \bar{\mathbb{E}}(\mathbf{S}_n) + \bar{\mathbb{E}}(\mathbf{S}_0) + \sum_{k=1}^n \bar{\mathbb{E}}(\mathbf{J}_o(I_k)) - \bar{\pi}(g),$$

Hence, we get that  $\Sigma_n(\bar{g}) = 0_{P_*}(1)$ , which implies  $\sigma_g^2 = 0$ .

We now derive the main result of this section, which determines the scaling limit of the constrained Hawkes process. We need some additional notation before stating the result. Let  $\mathbf{w} : \{1, \dots, p\} \rightarrow \mathbb{R}$ . Define

$$\mathcal{E}(\mathbf{w}) = \frac{\bar{\mathbb{E}}[\mathbf{w}(I_1)]}{\bar{\mathbb{E}}[\Delta_1]}. \quad (3.16)$$

Further define

$$v(\mathbf{w}) = \sigma_g^2, \quad (3.17)$$

where  $\sigma_g^2$  is defined as in (3.14) with  $g : \mathbf{Z} \rightarrow \mathbb{R}$  defined by  $g(\mathbf{z}) = \mathbf{w}(i) - \mathcal{E}(\mathbf{w})\delta$  for all  $\mathbf{z} = (\delta, i, \mathbf{s}, \ell)$ .

**Theorem 3.9.** Under the assumptions of Theorem 3.7, if  $v(\mathbf{w}) > 0$ , we have, for any initial condition, as  $T \rightarrow \infty$ ,

$$T^{-1/2} (\mathbf{N}(1_{[0, tT]} \otimes \mathbf{w}) - t T \mathcal{E}(\mathbf{w})) \xrightarrow{d} (v(\mathbf{w})/\bar{\mathbb{E}}[\Delta_1])^{1/2} B_t, \quad (3.18)$$

where the weak convergence holds in  $D([0, 1])$  (the space of càdlàg functions defined on  $[0, 1]$ ).

The proof is postponed to Section 5.7.

## 4 Application to a limit order book

Let us apply our results to the simple limit order book presented in the introduction. The function  $J$  and the sets  $A_i$  of the corresponding constrained multivariate Hawkes process are detailed in Section 2.3. The Markov assumption of Section 2.4 allows us to use a finite set of parameters for this model, namely  $\beta, \aleph$  and  $\boldsymbol{\mu}_0$ .

Assumptions 1 and 3 then amount to assumptions on the parameters  $\boldsymbol{\mu}_0$  and  $\aleph$ . Assumption 2 says that  $\aleph$  has to be invertible and requires an additional property which only depends on the constrained sets  $A_i$  and  $J$ . Hence, to apply Theorem 3.5, we only need to check that this additional property holds. We set  $s_o = 2$ . For any  $K \geq 1$ , we set  $m = |K - 2| + p$ . Then given  $s \in \{1, \dots, K\}$  we need to show that there is an admissible

path  $(j_1, \dots, j_m) \in \mathcal{A}_m(s)$  such that  $s + \sum_{k=1}^m J_o(j_k) = 2 (= s_o)$ . To see why, let us set, for  $1 \leq i \leq |s - 2|$ ,

$$j_i = \begin{cases} 2 & \text{if } s \geq 2, \\ 1 & \text{if } s = 1. \end{cases}$$

Then, by (2.7), we have  $s + \sum_{k=1}^{|s-2|} J(j_k) = 2$ . To conclude, we set  $j_i = 0$  for  $|s - 2| < i \leq m - 4$  and  $j_{m-3} = 1$ ,  $j_{m-2} = 4$ ,  $j_{m-1} = 2$  and  $j_m = 3$ , so that, by (2.7) and (2.11),  $s + \sum_{k=1}^m J_o(j_k) = 2$ . Moreover, we easily check that (3.1) holds for this choice of  $(j_1, \dots, j_m)$  and that  $\{j_{m-3}, j_{m-2}, j_{m-1}, j_m\} = \{1, \dots, 4\}$ . Hence Assumption 2 holds and Theorem 3.5 applies provided that  $\aleph$  is invertible and has spectral radius smaller than 1, and  $\mu_0$  satisfies (3.7) with  $J$  defined by (2.7). The ergodicity of the underlying chain allows one to perform meaningful statistical analysis of the data. This will be done in a forthcoming paper.

For the moment, we focus on the scaling limit of the mid-price, defined as the middle price between the Best Bid and Best Ask prices. This value is often considered as the continuous time price of the asset. In the framework of our model, the mid price satisfies the following equation

$$P(t) = P(u) + \mathbf{N}(1_{(u,t]} \otimes w) , \quad (4.1)$$

where  $w(i)$  takes values  $1/2, -1/2, 1/2$  and  $-1/2$  for  $i = 1, \dots, 4$ , respectively. (Note that similar equations hold for the Best Bid and Best Ask prices with different functions  $w$ .)

Applying Theorem 3.9, we get the scaling limit of the mid-price in physical time is given by

$$T^{-1/2} (P(tT) - P(0) - t T \mathcal{E}(w)) \xrightarrow{d} (v(w) / \overline{\mathbb{E}}[\Delta_1])^{1/2} B_t ,$$

where  $\mathcal{E}(w)$  and  $v(w)$  are respectively defined in (3.16) and (3.17), provided that  $v(w) > 0$ . Although it does not seem easy to check that  $v(w) > 0$  (except perhaps by numerical means), we do expect this to be true. Indeed, while the spread behaves as a stationary variable, yielding a vanishing asymptotic variance in the large scale behavior (see Remark 6), the best-bid, best-ask and mid prices behave as co-integrated random walks.

In practice, we can apply our results in several ways. We refer to the forthcoming thesis [35] for details about the estimation methods related to constrained Hawkes processes and their applications to limit order books data. Let us here briefly illustrate an application on a real data example. Using all intraday data from 1<sup>st</sup> April 2011 to 10<sup>th</sup> April 2011, parameters estimations for Eni SpA and Total yield the following results. The decaying rate  $\beta$  is estimated to 1.65 for Eni SpA and to 1.79 for Total, which correspond to half-times respectively equal to 0.6 and 0.55 seconds. The spectral radii of the estimated fertility matrices  $\aleph$  are equal to 0.5723 and 0.6221 for Eni SpA and Total, respectively, so Assumption 1 holds for the estimated parameters. We also computed the corresponding constants appearing in the left-hand side of the drift condition (3.7) and obtained  $-0.0115$  and  $-0.0682$  respectively for Eni SpA and Total. We note that for these estimated values, Assumption 3 holds and thus Theorem 3.5 applies. Following [5], to illustrate the relevance of our model, we shall briefly present empirical signature plots and compare them with the signature plots of the fitted models. For a given asset price (best bid, best ask or mid-price) at time  $t$   $P(t)$ , the empirical signature plot is defined as an empirical estimate of the quadratic variation of  $P(t)$  over  $[0, T]$  expressed as a function of the scale  $\tau > 0$ ,



namely

$$C(\tau) = \frac{1}{T} \sum_{n=1}^{T/\tau} |P(n\tau) - P((n-1)\tau)|^2 .$$

This can be compared with a theoretical signature plot based on the fitted model, namely

$$\tilde{C}(\tau) = \frac{1}{\tau} \mathbb{E}_{\hat{\theta}} \left[ (P(t+\tau) - P(t))^2 \right] ,$$

where  $\mathbb{E}_{\theta}$  here denotes the expectation of the model with parameter  $\theta$  and  $\hat{\theta}$  is the parameter estimated from the data. In practice this expectation is computed via Monte Carlo simulations of the fitted model.

In contrast with [5] where only the mid-price is considered and the signature plot is always decreasing, we observe in Figure 2 that the empirical signature plot differs depending whether the considered price is the best bid, best ask or mid-price. In Figure 3, we present the corresponding theoretical signature plots using the fitted models. We observe that, with a limited number of parameters, the model is able to recover shapes somewhat similar to the empirical ones.

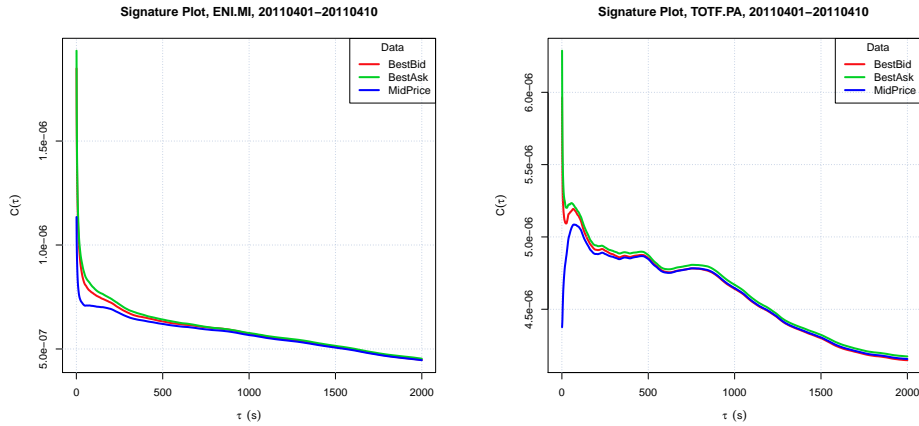


Figure 2: Signature Plot of Eni SpA (left) and Total (right) calculated on the data of best bid, best ask and mid-price over the period from 1<sup>st</sup> April 2011 to 10<sup>th</sup> April 2011.

## 5 Postponed proofs

### 5.1 Proof of Theorem 3.1

In the following we shall denote

$$\bar{\mu}_0 = \max_j \mu_0(j), \quad \underline{\mu}_0 = \min_j \mu_0(j), \quad (5.1)$$

$$\bar{\alpha} = \max_{i,j} \alpha_{i,j}, \quad \underline{\alpha} = \min_{i,j} \alpha_{i,j}, \quad (5.2)$$

$$\bar{J} = \max_{i=1,\dots,p} \sum_{j=1}^q |\vec{J}_{i,j}|, \quad (5.3)$$

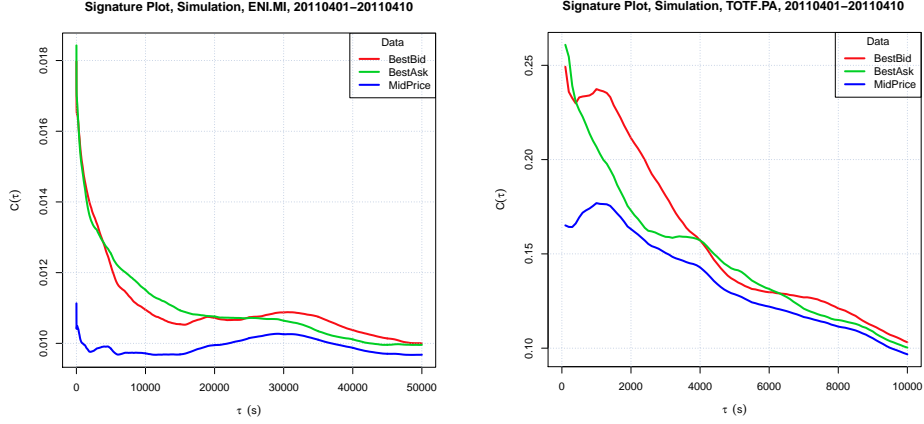


Figure 3: Signature plots calculated with simulated data where the parameters are calibrated from the mid-quotes data of Eni SpA (left) and Total (right) with  $p = 4$  and  $q = 1$ .

and the  $j$ -th column vector of  $\aleph$  by  $\aleph_j$ , so that  $\aleph = (\aleph_1 \dots \aleph_p)$ . To establish the existence of small sets, we need two preliminary results.

**Lemma 5.1.** *Let  $\mathbf{x} = (\mathbf{s}, \ell) \in \mathcal{X}$ ,  $j \in \mathcal{A}_1(\mathbf{s})$  with  $j \neq 0$  and  $g : \mathcal{X} \rightarrow \mathbb{R}_+$ . Then*

$$\mathbb{E}^{\mathbf{x}} [g(\lambda_1) 1_{\{I_1=j\}}] \geq \underline{\mu}_0 \int_0^{+\infty} g(e^{-t\beta} \ell + \beta \aleph_j) e^{-2(\overline{\mu}_0 + \bar{\ell})t} dt, \quad (5.4)$$

where  $\bar{\ell} = \max_i \ell(i)$ .

*Proof.* Using (2.12), we get that

$$\mathbb{E}^{\mathbf{x}} [g(\lambda_1) 1_{\{I_1=j\}}] = \mathbb{E}^{\mathbf{x}} [g(e^{-\Delta_1^j \beta} \ell + \beta \aleph_j) 1_{\{\Delta_1^j < U_1^j\}}],$$

where  $U_1^j = \min_{i \neq j} \Delta_1^i$ .

Observe that  $U_1^j \geq \text{Exp}(\overline{\mu}_0 + \bar{\ell})$  where  $\text{Exp}(z)$  denotes the exponential distribution with mean  $\frac{1}{z}$ , thus, since  $\Delta_1^j$  and  $U_1^j$  are independent, we have

$$\mathbb{E}^{\mathbf{x}} [g(\lambda_1) 1_{\{I_1=j\}}] \geq \mathbb{E}^{\mathbf{x}} [g(e^{-\Delta_1^j \beta} \ell + \beta \aleph_j) e^{-(\overline{\mu}_0 + \bar{\ell})\Delta_1^j}].$$

Since  $j \in \mathcal{A}_1(\mathbf{s})$ , using that the hazard rate of  $\Delta_1^j$  is bounded between  $\underline{\mu}_0$  and  $\overline{\mu}_0 + \bar{\ell}$ , we see that the density of  $\Delta_1^j$  is bounded from below by  $\underline{\mu}_0 e^{-(\overline{\mu}_0 + \bar{\ell})t}$  on  $t \in \mathbb{R}_+$ . Hence we get (5.4).  $\square$

**Corollary 5.2.** *Let  $m$  be an integer  $\geq 1$ ,  $M \in \mathbb{R}_+$  and  $g : \mathcal{X} \rightarrow \mathbb{R}_+$ . Let  $\mathbf{x} = (\mathbf{s}, \ell) \in \mathcal{X}$  with  $\ell \in (0, M]^p$  and  $(j_1, \dots, j_m) \in \mathcal{A}_m(\mathbf{s})$  with  $j_i \neq 0$  for  $i = 1, \dots, m$ . Then, we have*

$$\begin{aligned} & \mathbb{E}^{\mathbf{x}} [g(\lambda_m) 1_{\{I_1=j_1, \dots, I_m=j_m\}}] \\ & \geq \underline{\mu}_0^m \int_{\mathbb{R}_+^m} g[e^{-\sum_{i=1}^m t_i \beta} \ell + e^{-\sum_{i=2}^m t_i \beta} \beta \aleph_{j_1} + \dots + \beta \aleph_{j_m}] \\ & \quad \times e^{-C_{m,M} \sum_{i=1}^m t_i} dt_m \cdots dt_1, \end{aligned}$$

where  $C_{m,M} = 2(\bar{\mu}_0 + M + (m-1)\beta\bar{\alpha})$ .

*Proof.* The case  $m = 1$  is given by Lemma 5.1.

To get the result for any  $m \geq 2$ , we proceed by induction. Let  $\mathbf{x} = (\mathbf{s}, \ell) \in \mathsf{X}$  with  $\ell \in (0, M]^p$  and  $(j_1, \dots, j_m) \in \{1, \dots, p\}^m$ , if  $(j_1, \dots, j_m) \in \mathcal{A}_m(\mathbf{s})$ . We have

$$\begin{aligned} \mathbb{E}^{\mathbf{x}} [g(\boldsymbol{\lambda}_m) 1_{\{I_1=j_1, \dots, I_m=j_m\}}] \\ = \mathbb{E}^{\mathbf{x}} [1_{\{I_1=j_1\}} \mathbb{E} [g(\boldsymbol{\lambda}_m) 1_{\{I_2=j_2, \dots, I_m=j_m\}} | (I_1, \mathbf{X}_1)]] . \end{aligned} \quad (5.5)$$

Recall that given  $(I_1, \mathbf{X}_1)$ , the distribution of  $\{(I_k, \mathbf{X}_k), k \geq 2\}$  does not depend on  $I_1$ , hence, with the Markov property, we obtain that

$$\mathbb{E} [g(\boldsymbol{\lambda}_m) 1_{\{I_2=j_2, \dots, I_m=j_m\}} | (I_1, \mathbf{X}_1)] = \mathbb{E}^{\mathbf{X}_1} [g(\boldsymbol{\lambda}_{m-1}) 1_{\{I_1=j_2, \dots, I_{m-1}=j_m\}}] .$$

Note that, using (2.12) and (5.2),  $\{\mathbf{X}_0 = \mathbf{x}\}$  and  $\ell \in (0, M]^p$  implies  $\boldsymbol{\lambda}_1 \in (0, M + \beta\bar{\alpha}]^p$ . Thus, in the event  $\{\mathbf{X}_0 = \mathbf{x}\}$ , the right-hand side of the last display can be bounded from below by applying the induction hypothesis with  $(j_1, \dots, j_{m-1})$  replaced by  $(j_2, \dots, j_m)$ ,  $\ell = \boldsymbol{\lambda}_1$  and the constant

$$C_{m-1, M+\beta\bar{\alpha}} = 2(\bar{\mu}_0 + M + \beta\bar{\alpha} + (m-2)\beta\bar{\alpha}) = C_{m,M} .$$

Plugging this bound in (5.5), we get that

$$\begin{aligned} \mathbb{E}^{\mathbf{x}} [g(\boldsymbol{\lambda}_m) 1_{\{I_1=j_1, \dots, I_m=j_m\}}] \\ \geq \underline{\mu}_0^{m-1} \mathbb{E}^{\mathbf{x}} \left[ 1_{\{I_1=j_1\}} \int_{\mathbb{R}_+^{m-1}} g[\dots] \times e^{-C_{m,M} \sum_{i=1}^{m-1} t_i} dt_{m-1} \dots dt_1 \right] , \end{aligned} \quad (5.6)$$

where the omitted argument of  $g$  reads

$$[\dots] = e^{-\sum_{i=1}^{m-1} t_i \beta} \boldsymbol{\lambda}_1 + e^{-\sum_{i=2}^{m-1} t_i \beta} \beta \aleph_{j_2} + \dots + \beta \aleph_{j_m} .$$

Using again Lemma 5.1, we have

$$\begin{aligned} \mathbb{E}^{\mathbf{x}} [1_{\{I_1=j_1\}} g[\dots]] &\geq \underline{\mu}_0 \int_0^{+\infty} g[\dots] e^{-2(\bar{\mu}_0 + \bar{\ell})t} dt \\ &\geq \underline{\mu}_0 \int_0^{+\infty} g[\dots] e^{-C_{m,M} t} dt , \end{aligned} \quad (5.7)$$

where the omitted argument in the right-hand side reads

$$\begin{aligned} [\dots] &= e^{-\sum_{i=1}^{m-1} t_i \beta} (e^{-t\beta} \boldsymbol{\ell} + \beta \aleph_{j_1}) + e^{-\sum_{i=2}^{m-1} t_i \beta} \beta \aleph_{j_2} + \dots + \beta \aleph_{j_m} \\ &= e^{-\sum_{i=1}^{m-1} t_i \beta - t\beta} \boldsymbol{\ell} + e^{-\sum_{i=1}^{m-1} t_i \beta} \beta \aleph_{j_1} + \dots + \beta \aleph_{j_m} . \end{aligned}$$

Hence we get the result by applying the Tonelli theorem in (5.6) and then using (5.7).  $\square$

**Proposition 5.3.** *Let  $Q$  be the transition kernel defined in Section 2.4 on the space  $\mathbb{Z}_+^q \times \mathbb{R}_+^p$  with  $p, q \geq 1$ . Suppose that Assumption 2 holds. Then there exists a probability measure  $\nu$  on  $\mathsf{X}$  such that, for all  $K \geq 1$  and  $M > 0$ , there exists  $\epsilon > 0$  and a positive integer  $m$  such that  $\{1, \dots, K\}^q \times (0, M]^p$  is an  $(m, \epsilon, \nu)$  and  $(m+1, \epsilon, \nu)$ -small set for  $Q$ .*

Before providing the proof of this result, let us state the following corollary which follows by observing that the conditional distribution of  $(I_1, \mathbf{S}_1, \boldsymbol{\lambda}_1)$  given  $(I_0, \mathbf{S}_0, \boldsymbol{\lambda}_0)$  does not depend on  $I_0$ .

**Corollary 5.4.** Let  $\tilde{Q}$  be the transition kernel defined in Section 2.4 on the space  $\{0, 1, \dots, p\} \times \mathbb{Z}_+^q \times \mathbb{R}_+^p$  with  $p, q \geq 1$ . Suppose that Assumption 2 holds. Then there exists a probability measure  $\tilde{\nu}$  on  $\mathsf{X}$  such that, for all  $K \geq 1$  and  $M > 0$ , there exists  $\epsilon > 0$  and a positive integer  $m$  such that  $\{1, \dots, p\} \times \{1, \dots, K\}^q \times (0, M]^p$  is an  $(m, \epsilon, \tilde{\nu})$  and  $(m+1, \epsilon, \tilde{\nu})$ -small set for  $\tilde{Q}$ .

*Proof of Proposition 5.3.* Let  $K \geq 1$  and  $M > 0$  and denote  $C = \{1, \dots, K\}^q \times (0, M]^p$ . Let  $\mathbf{x} = (\mathbf{s}, \boldsymbol{\ell}) \in C$ . By Assumption 2, there exists an integer  $m > 0$ , and an admissible path  $(j_1, \dots, j_m)$  such that  $\mathbf{s} + \sum_{n=1}^m \mathbf{J}(j_n) = \mathbf{s}_o$  and  $\{j_{m-p+1}, \dots, j_m\} = \{1, \dots, p\}$ , where  $\mathbf{s}_o \in \mathbb{Z}_+^q$  is independent of  $m, M$  and  $K$ .

Now, for all subsets  $A \in \mathcal{B}(\mathbb{Z}_+^q)$  and  $B \in \mathcal{B}(\mathbb{R}_+^p)$ , we may write

$$\begin{aligned} Q^m(\mathbf{x}, A \times B) &= \mathbb{E}^{\mathbf{x}} [1_A(\mathbf{S}_m) 1_B(\boldsymbol{\lambda}_m)] \\ &\geq \mathbb{E}^{\mathbf{x}} \left[ 1_A(\mathbf{S}_m) 1_B(\boldsymbol{\lambda}_m) \prod_{i=1}^m 1_{\{I_i=j_i\}} \right] \\ &= 1_A(\mathbf{s}_o) \mathbb{E}^{\mathbf{x}} \left[ \prod_{i=1}^{m-p} 1_{\{I_i=j_i\}} \mathbb{E}^{\mathbf{X}_{m-p}} \left[ 1_B(\boldsymbol{\lambda}_p) \prod_{i=1}^p 1_{\{I_i=j_{m-p+i}\}} \right] \right], \quad (5.8) \end{aligned}$$

where we used that, given  $\mathbf{Y}_i$ ,  $i = 0, 1, \dots, m-p$ , the conditional distribution of  $\mathbf{Y}_j$ ,  $j > m-p$  only depends on  $\mathbf{X}_{m-p}$  and the Markov property.

Take now  $K \geq 1$  and  $M > 0$  and let  $\mathbf{x} \in \{1, \dots, K\}^q \times (0, M]^p$ . By (2.12) and (5.2), under  $\{\mathbf{X}_0 = \mathbf{x}\}$ , we have  $\boldsymbol{\lambda}_{m-p} \in (0, M']^p$  where  $M' = M + \bar{\alpha}\beta(m-p)$ . Moreover, under  $\{\mathbf{X}_0 = \mathbf{x}, I_i = j_i, i = 1, \dots, m-p\}$ , the last  $p$ -steps path  $(j_{m-p+1}, \dots, j_m)$  is admissible, that is,  $(j_{m-p+1}, \dots, j_m) \in \mathcal{A}_p(\mathbf{S}_{m-p})$ . Applying Corollary 5.2, we get that, under  $\{\mathbf{X}_0 = \mathbf{x}, I_i = j_i, i = 1, \dots, m-p\}$ ,

$$\begin{aligned} \mathbb{E}^{\mathbf{X}_{m-p}} \left[ 1_B(\boldsymbol{\lambda}_m) \prod_{i=m-p+1}^m 1_{\{I_i=j_i\}} \right] \\ \geq \underline{\mu}_0^p \int_{\mathbb{R}_+^p} 1_B \left( e^{-\sum_{i=1}^p t_i \beta} \boldsymbol{\ell} + e^{-\sum_{i=2}^p t_i \beta} \beta \aleph_{j_1} + \dots + \beta \aleph_{j_p} \right) \\ \times e^{-C_{m,M'} \sum_{i=1}^p t_i} dt_p \dots dt_1, \quad (5.9) \end{aligned}$$

where  $C_{m,M'} = 2(\bar{\mu}_0 + M' + (m-1)\bar{\alpha})$ . Setting  $u = e^{-\beta \sum_{i=1}^p t_i}$ ,  $\boldsymbol{\vartheta} = [\boldsymbol{\vartheta}(1), \dots, \boldsymbol{\vartheta}(p-1)]^T$  where  $\boldsymbol{\vartheta}(j) = e^{-\sum_{i=j+1}^p t_i \beta}$  for  $j = 1, \dots, p-1$ , Inequality (5.9) yields, under  $\{\mathbf{X}_0 = \mathbf{x}, I_i = j_i, i = 1, \dots, m-p\}$ ,

$$\begin{aligned} \mathbb{E}^{\mathbf{X}_{m-p}} \left[ 1_B(\boldsymbol{\lambda}_m) \prod_{i=m-p+1}^m 1_{\{I_i=j_i\}} \right] \\ \geq \underline{\mu}_0^p \frac{1}{\beta^{p-1}} \int_D 1_B(u\boldsymbol{\ell} + \beta \aleph \boldsymbol{\vartheta}) u^{\frac{C_{m,M'}}{\beta}} d\boldsymbol{\vartheta} du, \quad (5.10) \end{aligned}$$

where  $D = \{(u, \boldsymbol{\vartheta}) : 0 < u < \boldsymbol{\vartheta}(1) < \dots < \boldsymbol{\vartheta}(p-1) < 1\}$ . Applying Lemma A.1 with  $\Gamma = \beta \aleph$  and  $\gamma = \frac{C_{m,M'}}{\beta}$  in (5.10), we get that there exists  $\epsilon_1 > 0$  and a probability measure  $\nu_1$  such that, under  $\{\mathbf{X}_0 = \mathbf{x}, I_i = j_i, i = 1, \dots, m-p\}$ ,

$$\mathbb{E}^{\mathbf{X}_{m-p}} \left[ 1_B(\boldsymbol{\lambda}_m) \prod_{i=m-p+1}^m 1_{\{I_i=j_i\}} \right] \geq \epsilon_1 \nu_1(B). \quad (5.11)$$

Moreover, it is important to note that  $\nu_1$  only depends on  $\beta$  and  $\aleph$ .

Using Inequalities (5.8) and (5.11), we get that

$$Q^m(\mathbf{x}, A \times B) \geq \epsilon_1 1_A(\mathbf{s}_o) \nu_1(B) \mathbb{E}^{\mathbf{x}} \left[ \prod_{i=1}^{m-p} 1_{\{I_i=j_i\}} \right]. \quad (5.12)$$

Using that  $\mathbf{x} \in \{1, \dots, K\}^q \times (0, M]^p$  and Corollary 5.2 with  $g = 1$ , we further have that

$$\mathbb{E}^{\mathbf{x}} \left[ \prod_{i=1}^{m-p} 1_{\{I_i=j_i\}} \right] \geq \underline{\mu}_0^m \int_{\mathbb{R}_+^m} e^{-C_{m,M} \sum_{i=1}^m t_i} dt_m \cdots dt_1 = \left( \frac{\underline{\mu}_0}{C_{m,M}} \right)^m.$$

From Inequality (5.12), we get that  $Q^m(\mathbf{x}, A \times B) \geq \epsilon \nu(A \times B)$  for some positive constant  $\epsilon$  not depending on  $\mathbf{x}$  and  $\nu$  is the measure on  $\mathbf{X}$  defined by  $\nu(A \times B) = 1_A(\mathbf{s}_o) \nu_1(B)$ , which does neither depend on  $M$  nor  $K$ . In other words,  $C$  is a  $(m, \epsilon, \nu)$ -small set (see [28]).

To obtain that  $C$  is also an  $(m+1, \epsilon, \nu)$ -small set (by possibly decreasing  $\epsilon$ ), we simply observe that we can carry out the same proof as above with  $m$  replaced by  $m+1$  and the sequence of marks  $(j_1, \dots, j_m)$  replaced by  $(0, j_1, \dots, j_m)$ .  $\square$

*Proof of Theorem 3.1.* In Proposition 5.3, the measure  $\nu$  does not depend on the set  $C$ , which can be chosen to contain any arbitrary point  $\mathbf{x}$  in the state space  $\mathbf{X}$ . Hence the chain is  $\phi$ -irreducible with  $\phi = \nu$ . Moreover the set  $C$  is  $(m, \epsilon, \nu)$ -small and  $(m+1, \epsilon, \nu)$ -small; hence, the chain is aperiodic, sse [28, Section 5.4.3]. These arguments hold for the Kernel  $\tilde{Q}$  and thus we obtain Theorem 3.1.  $\square$

## 5.2 Proof of Proposition 3.2

Recall the definition of  $\mathbf{u}$  in (2.4). We shall further denote

$$1 \leq \underline{u} = \min_{i=1, \dots, p} \mathbf{u}(i) \leq \bar{u} = \max_{i=1, \dots, p} \mathbf{u}(i) < \infty. \quad (5.13)$$

The inequality  $\underline{u} \geq 1$  follows from the fact that  $\aleph^k$  has non-negative entries for all  $k \geq 1$ .

Observe that, given the initial condition  $\mathbf{x} = (\mathbf{s}, \boldsymbol{\ell}) \in \mathbb{Z}_+^q \times \mathbb{R}_+^q$ , the hazard rate of  $\Delta_1 = \min(\Delta_1^i, i = 0, \dots, p)$  reads

$$\text{Hr}(u) = \sum_{i=0}^p \text{Hr}_i(u) = \mu_0^0 + \mathbf{1}_p^T \boldsymbol{\mu}_0 + \mathbf{1}_p^T \boldsymbol{\ell} e^{-\beta u},$$

where we used (2.9). Moreover, for any  $i \in \{0, 1, \dots, p\}$ , we get that, given  $\Delta_1$ , the conditional probability that the first arrival's mark is  $i$  is given by

$$\mathbb{P}^{\mathbf{x}}(I_1 = i | \Delta_1) = \text{Hr}_i(\Delta_1) / \text{Hr}(\Delta_1). \quad (5.14)$$

Using (2.12), we have  $\boldsymbol{\lambda}_1 = \boldsymbol{\ell} e^{-\beta \Delta_1} + \beta \aleph_{I_1} 1_{\{I_1 \neq 0\}}$ , and thus

$$\begin{aligned} & [Q(1_{\mathbb{Z}_+^q} \otimes V_{1,\gamma})](\mathbf{x}) \\ &= \mathbb{E}^{\mathbf{x}} \left[ \exp(\gamma \mathbf{u}^T \boldsymbol{\ell} e^{-\beta \Delta_1} + \gamma \beta \mathbf{u}^T \aleph_{I_1} 1_{\{I_1 \neq 0\}}) \right] \\ &= \mathbb{E}^{\mathbf{x}} \left[ \exp(\gamma \mathbf{u}^T \boldsymbol{\ell} e^{-\beta \Delta_1}) \sum_{i=0}^p \mathbb{P}^{\mathbf{x}}(I_1 = i | \Delta_1) e^{\gamma \beta (\mathbf{u}(i) - 1)} \right], \end{aligned}$$

by conditioning on  $\Delta_1$  and using that  $\aleph^T \mathbf{u} = \mathbf{u} - \mathbf{1}_p$  (see (2.4)) and setting  $\mathbf{u}(0) = 1$ . Inserting Formula (5.14), we get

$$[Q(1_{\mathbb{Z}_+^q} \otimes V_{1,\gamma})](\mathbf{x}) = \mathbb{E}^{\mathbf{x}} \left[ \sum_{i=0}^p \exp(\gamma \mathbf{u}^T \boldsymbol{\ell} e^{-\beta \Delta_1} + \gamma \beta (\mathbf{u}(i) - 1)) \text{Hr}_i(\Delta_1) / \text{Hr}(\Delta_1) \right].$$

Using the hazard rate of  $\Delta_1$  under  $\mathbb{P}^{\mathbf{x}}$ , the corresponding probability density function is given by  $u \mapsto \text{Hr}(u)e^{-\text{Ir}(u)}$  on  $\mathbb{R}_+$ , where

$$\begin{aligned} \text{Ir}_i(t) &= \int_0^t \text{Hr}_i(u) \, du, \quad t \geq 0, \\ \text{Ir}(t) &= \sum_{i=1}^p \text{Ir}^i(t) = \mu_0^0 t + \mathbf{1}_p^T \boldsymbol{\mu}_0 t + \frac{\mathbf{1}_p^T \boldsymbol{\ell}}{\beta} (1 - e^{-\beta t}). \end{aligned} \quad (5.15)$$

Hence we get

$$[Q(1_{\mathbb{Z}_+^q} \otimes V_{1,\gamma})](\mathbf{x}) = \int_0^\infty e^{\gamma \mathbf{u}^T \boldsymbol{\ell} e^{-\beta u} - \text{Ir}(u)} \left[ \sum_{i=0}^p e^{\gamma \beta (\mathbf{u}^{(i)} - 1)} \text{Hr}_i(u) \right] du.$$

Using that  $e^{\gamma \beta (\mathbf{u}^{(0)} - 1)} \leq e^{\gamma \beta (\mathbf{u}^{(i)} - 1)}$  for all  $i = 1, \dots, p$  and the particular form of  $\text{Hr}(0)$  and  $\text{Hr}(i)$  in (2.9), we easily obtain that

$$\sum_{i=0}^p e^{\gamma \beta (\mathbf{u}^{(i)} - 1)} \text{Hr}_i(u) \leq \left[ \mu_0^0 + \sum_{i=1}^p e^{\gamma \beta (\mathbf{u}^{(i)} - 1)} (\boldsymbol{\mu}_0(i) + \boldsymbol{\ell}(i) e^{-u\beta}) \right].$$

Consequently, we have

$$\begin{aligned} & \frac{1}{V_{1,\gamma}(\boldsymbol{\ell})} [Q(1_{\mathbb{Z}_+^q} \otimes V_{1,\gamma})](\mathbf{x}) \\ & \leq \left( \mu_0^0 + \sum_{i=1}^p e^{\gamma \beta (\mathbf{u}^{(i)} - 1)} \boldsymbol{\mu}_0(i) \right) \frac{1}{V_{1,\gamma}(\boldsymbol{\ell})} \int_0^\infty e^{\gamma \mathbf{u}^T \boldsymbol{\ell} e^{-\beta u} - \text{Ir}(u)} du \\ & \quad + \left( \sum_{i=1}^p e^{\gamma \beta (\mathbf{u}^{(i)} - 1)} \boldsymbol{\ell}(i) \right) \frac{1}{V_{1,\gamma}(\boldsymbol{\ell})} \int_0^\infty e^{\gamma \mathbf{u}^T \boldsymbol{\ell} e^{-\beta u} - \text{Ir}(u) - u\beta} du. \end{aligned} \quad (5.16)$$

Using (5.15), we have

$$\begin{aligned} \frac{1}{V_{1,\gamma}(\boldsymbol{\ell})} \int_0^\infty e^{\gamma \mathbf{u}^T \boldsymbol{\ell} e^{-\beta u} - \text{Ir}(u)} du &= \int_0^\infty e^{(\gamma \mathbf{u}^T \boldsymbol{\ell} + \frac{\mathbf{1}_p^T \boldsymbol{\ell}}{\beta})(e^{-\beta u} - 1) - (\mu_0^0 + \mathbf{1}_p^T \boldsymbol{\mu}_0)u} du \\ &\rightarrow 0 \quad \text{as } \mathbf{1}_p^T \boldsymbol{\ell} \rightarrow \infty, \end{aligned} \quad (5.17)$$

where we used Lemma A.2 with  $a = \gamma \mathbf{u}^T \boldsymbol{\ell} + \frac{\mathbf{1}_p^T \boldsymbol{\ell}}{\beta}$  and  $\beta' = \mu_0^0 + \mathbf{1}_p^T \boldsymbol{\mu}_0$ .

Similarly, applying Lemma A.2 with  $\beta' = \beta + \mu_0^0 + \mathbf{1}_p^T \boldsymbol{\mu}_0$ , we obtain

$$\begin{aligned} & \frac{1}{V_{1,\gamma}(\boldsymbol{\ell})} \int_0^\infty e^{\gamma \mathbf{u}^T \boldsymbol{\ell} e^{-\beta u} - \text{Ir}(u) - u\beta} du \\ &= \int_0^\infty \exp \left( (\gamma \mathbf{u}^T \boldsymbol{\ell} + \frac{\mathbf{1}_p^T \boldsymbol{\ell}}{\beta})(e^{-\beta u} - 1) - (\beta + \mu_0^0 + \mathbf{1}_p^T \boldsymbol{\mu}_0)u \right) du \\ &\sim \frac{1}{\gamma \beta \mathbf{u}^T \boldsymbol{\ell} + \mathbf{1}_p^T \boldsymbol{\ell}} \quad \text{as } \mathbf{1}_p^T \boldsymbol{\ell} \rightarrow \infty. \end{aligned} \quad (5.18)$$

Therefore, inserting (5.17) and (5.18) in (5.16), we get

$$\limsup_{\mathbf{1}_p^T \boldsymbol{\ell} \rightarrow \infty} \frac{1}{V_{1,\gamma}(\boldsymbol{\ell})} [Q(1_{\mathbb{Z}_+^q} \otimes V_{1,\gamma})](\mathbf{x}) \leq \sup_{\boldsymbol{\ell} \in (0, \infty)^p} \frac{\sum_{i=1}^p e^{\gamma \beta (\mathbf{u}^{(i)} - 1)} \boldsymbol{\ell}(i)}{\gamma \beta \mathbf{u}^T \boldsymbol{\ell} + \mathbf{1}_p^T \boldsymbol{\ell}}.$$

Now observe that, using (5.13), we have

$$0 < \underline{u} \leq \mathbf{u}^T \boldsymbol{\ell} / \mathbf{1}_p^T \boldsymbol{\ell} \leq \bar{u} < \infty ,$$

and thus, for  $\gamma > 0$  small enough, using a Taylor expansion of the exponential function at the origin, we have

$$\begin{aligned} \frac{\sum_{i=1}^p e^{\gamma \beta (\mathbf{u}^{(i)} - 1) \boldsymbol{\ell}(i)}}{\beta \gamma \mathbf{u}^T \boldsymbol{\ell} + \mathbf{1}_p^T \boldsymbol{\ell}} &\leq \frac{\mathbf{1}_p^T \boldsymbol{\ell} + \gamma \beta (\mathbf{u}^T \boldsymbol{\ell} - \mathbf{1}_p^T \boldsymbol{\ell}) + C \gamma^2 \mathbf{1}_p^T \boldsymbol{\ell}}{\beta \gamma \mathbf{u}^T \boldsymbol{\ell} + \mathbf{1}_p^T \boldsymbol{\ell}} \\ &\leq 1 - \beta \gamma \frac{\mathbf{1}_p^T \boldsymbol{\ell}}{\beta \gamma \mathbf{u}^T \boldsymbol{\ell} + \mathbf{1}_p^T \boldsymbol{\ell}} + C' \gamma^2 , \\ &\leq 1 - \beta \gamma \frac{1}{1 + \beta \gamma \bar{u}} + C' \gamma^2 = 1 - \beta \gamma + O(\gamma^2) , \end{aligned}$$

where  $C, C'$  do not depend on  $\boldsymbol{\ell}$ . Hence taking  $\gamma_1 > 0$  small enough, we have for all  $\gamma \in (0, \gamma_1]$ ,

$$\theta(\gamma) = \sup_{\boldsymbol{\ell} \in (0, \infty)^p} \frac{\sum_{i=1}^p e^{\gamma \beta (\mathbf{u}^{(i)} - 1) \boldsymbol{\ell}(i)}}{\beta \gamma \mathbf{u}^T \boldsymbol{\ell} + \mathbf{1}_p^T \boldsymbol{\ell}} < 1 .$$

Hence choosing such a  $\gamma$  and setting  $\theta = \theta(\gamma)$ , we get (3.2) for  $\boldsymbol{\ell}$  out of  $(0, M]^p$  provided that  $M$  large enough. Since  $V_{1, \gamma}$  is bounded on  $(0, M]^p$ , this achieves the proof.

### 5.3 Proof of Corollary 3.4

By [5, Theorem (1)], using the definitions and notation of Section 2.1, we have

$$\lim_{T \rightarrow \infty} \frac{1}{T} \mathbf{N}'(1_{[0, T]} \otimes \mathbf{w}) = \overline{\mathbf{w}}^T (\text{Id}_p - \aleph)^{-1} \boldsymbol{\mu}_0 \quad \text{a.s.} \quad (5.19)$$

On the other hand by the law of large numbers for Harris recurrent chains (see [28, Theorem 17.1.7]), we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbf{w}_o(\check{I}_k) = \check{\pi}(\mathbf{w}_o \otimes \mathbf{1}_{\mathbb{R}_+^p}) \quad \text{a.s.}[P_*] . \quad (5.20)$$

Here  $[P_*]$  means that the result holds for any initial distribution on  $(\check{I}_0, \check{\boldsymbol{\lambda}}_0)$ . In particular taking  $\check{\boldsymbol{\lambda}}_0 = 0$ , we get that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{N}'(1_{[0, \check{T}_n]} \otimes \mathbf{w}) = \check{\pi}(\mathbf{w}_o \otimes \mathbf{1}_{\mathbb{R}_+^p}) , \quad (5.21)$$

where  $\check{T}_n$  is defined as  $T_n$  in Section 2.4 but for the unconstrained chain. Note that we used that  $\mathbf{w}_o(0) = 0$ , so that only the marks in  $\{1, \dots, p\}$  are counted in the sum of the left-hand side of (5.20). Observe now that under the event  $\{\check{\boldsymbol{\lambda}}_0 = 0\}$ , we have

$$n_0([0, \check{T}_n]) + \mathbf{N}'(1_{[0, \check{T}_n]} \otimes \mathbf{1}_{\{1, \dots, p\}}) = n .$$

(Recall that  $n_0$  is an independent homogeneous Poisson point process which corresponds to the events with marks equal to 0). Since  $(\check{T}_n)$  is an increasing sequence, we get that it diverges a.s. and thus, applying (5.19) with  $\mathbf{w} = \mathbf{1}_{\{1, \dots, p\}}$ , we have, under the event  $\{\check{\boldsymbol{\lambda}}_0 = 0\}$ ,

$$\lim_{n \rightarrow \infty} \frac{1}{\check{T}_n} \mathbf{N}'(1_{[0, \check{T}_n]} \otimes \mathbf{1}_{\{1, \dots, p\}}) = \mathbf{1}_p^T (\text{Id}_p - \aleph)^{-1} \boldsymbol{\mu}_0 \quad \text{a.s.}$$

The last two displays yield, under the event  $\{\check{\boldsymbol{\lambda}}_0 = 0\}$ ,

$$\lim_{n \rightarrow \infty} \frac{\check{T}_n}{n} = \frac{1}{\boldsymbol{\mu}_0^0 + \mathbf{1}_p^T (\text{Id}_p - \aleph)^{-1} \boldsymbol{\mu}_0} .$$

Applying (5.19), (5.21) and the last display, we get (3.4).

## 5.4 Proof of Theorem 3.6

The case (i) follows from Proposition 3.2.

We next consider the case (ii) so that  $\psi(\{s_2\} \times (0, \infty)^p) > 0$  for some  $s_2 \geq s^*$ . By (3.8) there exists  $i \in \{1, \dots, p\}$  such that  $J(i) > 0$ . Thus, for any  $m \geq 1$ , the constant sequence  $(i, \dots, i)$  of length  $m$  belongs to  $\mathcal{A}_m(s_2)$  (see Definition 1). It follows that  $\psi(\{s_2 + mJ(i)\} \times (0, \infty)^p) > 0$  for all  $m \geq 1$ . We shall prove that for all  $s' \geq 1$  and  $M > 0$ ,

$$\liminf_{s \rightarrow \infty} \inf_{\ell \in (0, M]^p} \mathbb{P}^{(s, \ell)} \left( \inf_{n \geq 1} \mathbf{S}_n > s' \right) > 0. \quad (5.22)$$

Take  $s'$  such that  $\psi(\{s'\} \times (0, \infty)^p) > 0$ . Define  $s = s_2 + mJ(i)$  and take  $m$  and  $M > 0$  large enough so that  $\psi(\{s\} \times (0, M)^p) > 0$  and  $\mathbb{P}^{(s, \ell)} \left( \inf_{n \geq 1} \mathbf{S}_n > s' \right) > 0$  for all  $\ell \in (0, M]^p$ . We thus have that the probability that the chain  $\{\mathbf{X}_k, k \geq 1\}$  avoids  $\{s'\} \times (0, \infty)^p$  starting from anywhere in  $\{s\} \times (0, M]^p$  is positive. Since these two sets have positive  $\psi$ -measure, we get by [28, Theorem 8.3.6] that  $Q$  is transient.

Hence it only remains to show that (5.22) holds. Applying Corollary 3.4 and Condition (3.8), we have

$$\lim_{n \rightarrow \infty} M_n = \infty \quad \text{a.s.}[P_*], \quad \text{with} \quad M_n = \sum_{k=1}^n J_o(\check{I}_k).$$

Hence  $U = \inf_{n \geq 1} M_n$  is valued in  $\mathbb{R}$  a.s. $[P_*]$  (it equals  $-\infty$  with probability 0). Let  $\epsilon \in (0, 1)$ .

Now under the event  $\{S_0 = s\}$ , we have that  $S_n = s + M_n$  as long as  $S_k$  does not hit the constraint set  $\cup_{i=1}^p A_i$  for  $k = 1, \dots, n$ .

For all  $s \geq s^*$ , the event  $\{\inf_{n \geq 0} S_n \geq s^*\}$  coincides with the event  $\{U + s \geq s^*\}$ . Hence, under this latter event we have  $\inf_{n \geq 1} S_n = s + U$ . Finally, we get that, for all  $s \geq s^*$  and all  $\ell \in \mathbb{R}_+^p$ ,

$$\mathbb{P}^{(s, \ell)} \left( \inf_{n \geq 1} S_n > s' \right) \geq \mathbb{P}^\ell (U > (s' \vee s^*) - s). \quad (5.23)$$

(Recall that  $U$  only depends on the unconstrained chain  $(\check{I}_n, \check{\lambda}_n)$ ,  $n \geq 0$ .) Let us denote, for all  $k \geq 1$ ,  $U_k = \inf_{n \geq k} M_n$  so that  $U = M_1 \wedge \dots \wedge M_k \wedge U_{k+1}$ . Since  $M_1, \dots, M_k \geq -k\bar{J}$ , we have, for all  $s_1 > k\bar{J}$ ,

$$w_{s_1}(\ell) := \mathbb{P}^\ell (U > -s_1) = \mathbb{P}^\ell (U_{k+1} > -s_1) = [\check{Q}^k(w_{s_1})](\ell).$$

Observe that  $w_{s_1}$  is a function valued in  $[0, 1]$ . By Proposition 3.3 and [28, Section 15]),  $\check{Q}$  admits a stationary distribution  $\tilde{\pi}$  and there are constants  $\gamma, C > 0$  and  $\theta \in (0, 1)$  not depending on  $s_1$  such that, for all  $\ell, \ell_o \in \mathbb{R}_+^p$  and  $k \geq 1$ ,

$$[\check{Q}^k(w_{s_1})](\ell) \geq [\check{Q}^k(w_{s_1})](\ell_o) - C\theta^k (V_{1, \gamma}(\ell) + V_{1, \gamma}(\ell_o)).$$

Now, fix some  $\ell_o \in \mathbb{R}_+^p$  and  $M > 0$ . Choose  $k$  large enough so that

$$c = C\theta^k (V_{1, \gamma}(M\mathbf{1}_p) + V_{1, \gamma}(\ell_o)) < 1/2.$$

Then we may choose  $s_1$  large enough so that  $s_1 > k\bar{J}$  and

$$[\check{Q}^k(w_{s_1})](\ell_o) = \mathbb{P}^{\ell_o} (U_{k+1} > -s_1) \geq \mathbb{P}^{\ell_o} (U > -s_1) \geq 1/2.$$

The last four displays yield that, for all  $\ell \in (0, M]^p$ , and all  $s_1$  large enough,

$$\mathbb{P}^\ell (U > -s_1) \geq 1/2 - c > 0.$$

This and (5.23) yields (5.22), which concludes the proof.



## 5.5 Proof of Theorem 3.7 : induction on $q$

We have already shown that the chains  $Q$  and  $\tilde{Q}$  are  $\psi$ -irreducible and aperiodic and exhibited petite sets, see Theorem 3.1. To obtain a drift condition on the constrained case  $q \geq 1$ , we shall reason by induction on the number of constraints  $q$ . We shall rely on the transition kernels  $\tilde{Q}^{(-\mathcal{J})}$  introduced in Section 2.4 for subsets  $\mathcal{J}$  in  $\{1, \dots, q\}$ . For such a given set  $\mathcal{J}$ , we shall further denote by  $\{(\tilde{I}_n, \tilde{\mathbf{S}}_n, \tilde{\boldsymbol{\lambda}}_n), n \geq 0\}$  the chain with transition kernel  $\tilde{Q}^{(-\mathcal{J})}$  which starts at the same state as  $(I_0, \mathbf{S}_0, \boldsymbol{\lambda}_0)$  and, for all  $\mathbf{x} \in \mathbb{Z}_+^{q-\#\mathcal{J}} \times \mathbb{R}_+^p$ , by  $\mathbb{E}^{(-\mathcal{J}), \mathbf{x}}[\dots]$  the expectation under  $\{(\tilde{\mathbf{S}}_0, \tilde{\boldsymbol{\lambda}}_0) = \mathbf{x}\}$ . (The fact that this expectation only depends on  $\mathbf{x}$  is explained in Remark 1.)

**Proposition 5.5.** *Let  $p, q \geq 1$  and define the transition kernels  $Q$  on the space  $\mathbb{Z}_+^q \times \mathbb{R}_+^p$  as in Section 2.4. Suppose that  $\aleph$  is invertible and that Assumption 1 holds. For any non-empty  $\mathcal{J} \subseteq \{1, \dots, q\}$ , let  $\{(\tilde{I}_n, \tilde{\mathbf{S}}_n, \tilde{\boldsymbol{\lambda}}_n), n \geq 0\}$  be a Markov chain on the space  $\{0, \dots, p\} \times \mathbb{Z}_+^{q-\#\mathcal{J}} \times \mathbb{R}_+^p$  with transition kernel  $\tilde{Q}^{(-\mathcal{J})}$  defined in Section 2.4, and suppose that it satisfies the following condition.*

(C) *For all  $\gamma_1 > 0$  small enough, there exists  $\gamma_0^* > 0$  such that for all  $\gamma_0 \in (0, \gamma_0^*]$ ,*

$$\lim_{m \rightarrow \infty} \sup_{(\mathbf{s}, \boldsymbol{\ell}) \in \mathbb{Z}_+^{q-\#\mathcal{J}} \times \mathbb{R}_+^p} \mathbb{E}^{(-\mathcal{J}), (\mathbf{s}, \boldsymbol{\ell})} \left[ e^{\gamma_0 \sum_{k=1}^m \mathbf{w}(\tilde{I}_k)} V_{1, \gamma_1}(\tilde{\boldsymbol{\lambda}}_m - \boldsymbol{\ell}) \right] = 0, \quad (5.24)$$

where we denoted  $\mathbf{w}$  is defined in (5.42).

Then, for all  $\gamma_1 > 0$  small enough, there exists  $\gamma_0^* > 0$  such that for all  $\gamma_0 \in (0, \gamma_0^*]$ , there exist  $\theta \in (0, 1)$ ,  $b > 0$  and  $m \geq 1$ , such that for any initial condition  $\mathbf{x} \in \mathbf{X}$ ,

$$Q^m(V_{0, \gamma_0}^{\otimes q} \otimes V_{1, \gamma_1})(\mathbf{x}) \leq \theta (V_{0, \gamma_0}^{\otimes q} \otimes V_{1, \gamma_1})(\mathbf{x}) + b, \quad (5.25)$$

where  $V_{1, \gamma}$  and  $V_{0, \gamma}$  are defined in (3.3) and (3.6) respectively.

*Proof.* Take  $\mathbf{x} = (\mathbf{s}, \boldsymbol{\ell}) \in \mathbf{X}$ . We already have the drift condition (3.2) of Proposition 3.2 (which holds for some  $\gamma > 0$  and thus, by Jensen inequality for any  $\gamma_1 \in (0, \gamma)$ ). By induction, for all small enough  $\gamma_1 > 0$ , there exists  $b_1 > 0$  and  $\theta_1 > 0$ , such that, for all  $m \geq 1$ ,

$$\begin{aligned} Q^m(1_{\mathbb{Z}_+^q} \otimes V_{1, \gamma_1})(\mathbf{x}) &\leq \theta_1^m V_{1, \gamma_1}(\boldsymbol{\ell}) + b_1 (1 + \theta_1 + \dots + \theta_1^m) \\ &\leq \theta_1^m V_{1, \gamma_1}(\boldsymbol{\ell}) + b_1 / (1 - \theta_1). \end{aligned}$$

By (2.10) and (5.3) we have, in the event  $\{\mathbf{S}_0 = \mathbf{s}\}$ ,  $V_{0, \gamma_0}^{\otimes q}(\mathbf{S}_1) \leq e^{\gamma_0 \bar{J}} V_{0, \gamma_0}^{\otimes q}(\mathbf{s})$ , then we get  $V_{0, \gamma_0}^{\otimes q}(\mathbf{S}_m) \leq e^{\gamma_0 \bar{J} m} V_{0, \gamma_0}^{\otimes q}(\mathbf{S}_0)$ . Hence, we obtain, for any  $\gamma_0 > 0$ ,

$$\begin{aligned} Q^m(V_{0, \gamma_0}^{\otimes q} \otimes V_{1, \gamma_1})(\mathbf{x}) &\leq e^{\gamma_0 \bar{J} m} V_{0, \gamma_0}^{\otimes q}(\mathbf{s}) [\theta_1^m V_{1, \gamma_1}(\boldsymbol{\ell}) + b_1 / (1 - \theta_1)] \\ &\leq \left( e^{\gamma_0 \bar{J}} \theta_1 \right)^m (V_{0, \gamma_0}^{\otimes q} \otimes V_{1, \gamma_1})(\mathbf{x}) + \frac{b_1 e^{\gamma_0 \bar{J} m}}{1 - \theta_1} V_{0, \gamma_0}^{\otimes q}(\mathbf{s}). \end{aligned} \quad (5.26)$$

Observe that, for a given  $\theta_1 < 1$ , we can choose  $\gamma_0 > 0$  small enough so that  $e^{\gamma_0 \bar{J}} \theta_1 < 1$ . It follows that, choosing such a  $\gamma_0$ , for all  $m \geq 1$  and  $K \geq 1$ , the bound (5.26) implies (5.25) for all  $\mathbf{x} = (\mathbf{s}, \boldsymbol{\ell}) \in \mathbf{X}$  such that  $V_{0, \gamma_0}^{\otimes q}(\mathbf{s}) \leq K$ . To address the case  $V_{0, \gamma_0}^{\otimes q}(\mathbf{s}) > K$ , we shall prove the following assertion.

(D) *For all  $\gamma_1 > 0$  small enough, there exists  $\gamma_0^* > 0$  such that for all  $\gamma_0 \in (0, \gamma_0^*]$ , there exist  $m \geq 1$ ,  $K > 0$ ,  $b \geq 0$  and  $\theta \in (0, 1)$  such that (5.25) holds for all  $\mathbf{x} = (\mathbf{s}, \boldsymbol{\ell}) \in \mathbf{X}$  satisfying  $V_{0, \gamma_0}^{\otimes q}(\mathbf{s}) > K$ .*

To conclude the proof, we now show that Assertion (D) holds. Let  $m \geq 1$  and  $\mathbf{x} = (\mathbf{s}, \ell) \in \mathsf{X}$ . Observe that, by (2.10), for all  $\gamma_0 > 0$ , we have, in the event  $\{\mathbf{X}_0 = \mathbf{x}\}$ ,

$$V_{0,\gamma_0}^{\otimes q}(\mathbf{S}_m) = V_{0,\gamma_0}^{\otimes q}(\mathbf{s}) e^{\gamma_0 \sum_{k=1}^m \sum_{j=1}^q \mathbf{J}_j(I_k)} = V_{0,\gamma_0}^{\otimes q}(\mathbf{s}) e^{\gamma_0 \sum_{k=1}^m \mathbf{w}(I_k)}.$$

Hence we have, for all  $\mathbf{x} = (\mathbf{s}, \ell) \in \mathsf{X}$ ,

$$\begin{aligned} Q^m(V_{0,\gamma_0}^{\otimes q} \otimes V_{1,\gamma_1})(\mathbf{x}) &= V_{0,\gamma_0}^{\otimes q}(\mathbf{s}) \mathbb{E}^{\mathbf{x}} \left[ e^{\gamma_0 \sum_{k=1}^m \mathbf{w}(I_k)} V_{1,\gamma_1}(\boldsymbol{\lambda}_m) \right] \\ &= (V_{0,\gamma_0}^{\otimes q} \otimes V_{1,\gamma_1})(\mathbf{x}) \mathbb{E}^{\mathbf{x}} \left[ e^{\gamma_0 \sum_{k=1}^m \mathbf{w}(I_k)} V_{1,\gamma_1}(\boldsymbol{\lambda}_m - \ell) \right]. \end{aligned} \quad (5.27)$$

Now the idea of the proof relies on the fact that if  $V_{0,\gamma_0}^{\otimes q}(\mathbf{S}_0)$  is large enough, then it will take some time for all the components  $j$  in a set  $\mathcal{J}$  (with cardinal  $\#\mathcal{J}$  at least 1) of the multivariate spread to reach their constraint set

$$\mathbf{A}(j) = \bigcup_{i=1,\dots,p} \mathbf{A}_i(j). \quad (5.28)$$

Up to this time the process defined by  $\{\mathbf{Y}_n^{(-\mathcal{J})} = (I_n, \mathbf{S}_n^{(-\mathcal{J})}, \boldsymbol{\lambda}_n), n \geq 0\}$ , where  $\mathbf{S}_n^{(-\mathcal{J})}$  is obtained by removing the  $j$ -th component of  $\mathbf{S}_n$  for all  $j \in \mathcal{J}$ , behaves as a Markov chain with kernel  $\tilde{Q}^{(-\mathcal{J})}$  (defined as  $\tilde{Q}$  but without the  $j$ -th constraints with  $j \in \mathcal{J}$ , see above). More precisely, we use the following coupling argument. For a positive integer  $m$ , we denote

$$s_m^*(j) = 1 + m\bar{\mathcal{J}} + \max(\mathbf{A}(j)). \quad (5.29)$$

We can then define a Markov chain  $\{(\tilde{I}_n, \tilde{\mathbf{S}}_n, \tilde{\boldsymbol{\lambda}}_n), n \geq 0\}$  with transition kernel  $\tilde{Q}^{(-\mathcal{J})}$  and such that in the event  $\{\mathbf{S}_0(j) \geq s_m^*(j), j \in \mathcal{J}\}$ , the sequence  $\{(\tilde{I}_n, \tilde{\mathbf{S}}_n, \tilde{\boldsymbol{\lambda}}_n), n = 0, 1, \dots, m\}$  coincides with  $\{(I_n, \mathbf{S}_n^{(-\mathcal{J})}, \boldsymbol{\lambda}_n), n = 0, 1, \dots, m\}$ . Then, by construction, for all  $\mathbf{x} = (\mathbf{s}, \ell)$  such that  $\mathbf{s}(j) \geq s_m^*(j)$  for all  $j \in \mathcal{J}$ , denoting  $\mathbf{x}' = (\mathbf{s}^{(-\mathcal{J})}, \ell)$ , we have

$$\mathbb{E}^{\mathbf{x}} \left[ e^{\gamma_0 \sum_{k=1}^m \mathbf{w}(I_k)} V_{1,\gamma_1}(\boldsymbol{\lambda}_m - \ell) \right] = \mathbb{E}^{(-\mathcal{J}), \mathbf{x}'} \left[ e^{\gamma_0 \sum_{k=1}^m \mathbf{w}(\tilde{I}_k)} V_{1,\gamma_1}(\boldsymbol{\lambda}_m - \ell) \right]. \quad (5.30)$$

Now, we take  $K > 0$  such that  $\log K \geq p\gamma_0 \max_{j=1,\dots,q} s_m^*(j)$ . We let  $\mathcal{J}$  denote the set of all indices  $j \in \{1, \dots, q\}$  such that  $\mathbf{s}(j) \geq s_m^*(j)$ . The set  $\mathcal{J}$  is not empty since  $V_{0,\gamma_0}^{\otimes q}(\mathbf{s}) > K$  implies  $\max(\mathbf{s}) \geq \max(s_m^*)$ . Hence, Equality (5.30) applies for all  $\mathbf{x} = (\mathbf{s}, \ell) \in \mathsf{X}$  satisfying  $V_{0,\gamma_0}^{\otimes q}(\mathbf{s}) > K$ . For  $\gamma_0^*$  as in Condition (C) (we can choose it independently of  $\mathcal{J}$  since there is a finite number of such subsets), we obtain Assertion (D).  $\square$

Condition (C) of Proposition 5.5 needs to be simplified. A first step in this direction is given by the following lemma.

**Lemma 5.6.** *Let  $\{(I_n, \mathbf{S}_n, \boldsymbol{\lambda}_n), n \geq 0\}$  be the Markov chain on the space  $\{0, 1, \dots, p\} \times \mathbb{Z}_+^q \times \mathbb{R}_+^p$  with transition kernel  $\tilde{Q}$  defined in Section 2.4. Let  $\mathbf{w} : \{0, 1, \dots, p\} \rightarrow \mathbb{R}$ . Suppose that the following assertions hold.*

(E-1) *For all  $\gamma' > 0$  small enough, there exists  $\gamma^* > 0$  such that, for all  $\gamma \in (0, \gamma^*]$ ,  $\tilde{Q}$  is  $(1_{\{0,1,\dots,p\}} \otimes V_{0,\gamma}^{\otimes q} \otimes V_{1,\gamma'})$ -geometrically ergodic, where  $V_{0,\gamma}$  and  $V_{1,\gamma'}$  are defined in (3.6) and (3.3) respectively.*

(E-2) *The stationary distribution  $\tilde{\pi}$  of  $\tilde{Q}$  satisfies  $\tilde{\pi} \left[ \mathbf{w} \otimes 1_{\mathbb{Z}_+^q \times \mathbb{R}_+^p} \right] < 0$ .*

*Then, for all  $\gamma_1 > 0$  small enough, there are positive constants  $\gamma, \rho$  and  $\gamma_0^*$  such that, for all  $\gamma_0 \in (0, \gamma_0^*]$ ,*

$$\sup_{\mathbf{x} \in \mathsf{X}} \sup_{m \geq 1} \frac{\mathbb{E}^{\mathbf{x}} \left[ e^{\gamma_0 \sum_{k=1}^m \{\mathbf{w}(I_k) + \rho\}} (V_{0,\gamma}^{\otimes q} \otimes V_{1,\gamma_1})(\mathbf{X}_m) \right]}{(V_{0,\gamma}^{\otimes q} \otimes V_{1,\gamma_1})(\mathbf{x})} < \infty. \quad (5.31)$$

*Proof.* Using Assumption (E-1), for any  $\gamma' > 0$  small enough, we may choose  $\gamma, \gamma_1 \in (0, \gamma^*]$  such that  $\tilde{Q}$  is  $1_{\{0, \dots, p\}} \otimes V_{0, \gamma}^{\otimes q} \otimes V_{1, \gamma_1}$ -geometrically ergodic. Let  $\bar{w} = \max_{0 \leq i \leq p} \mathbf{w}(i)$ . Then, for any  $\gamma^* > 0$ , there  $\gamma, \gamma_1 > 0$  small enough, we have for all  $j \geq 1$  and  $\mathbf{x} = (\mathbf{s}, \ell) \in \mathbf{X}$ ,

$$\begin{aligned} \mathbb{E}^{\mathbf{x}} \left[ e^{\gamma_0 \sum_{k=1}^j \mathbf{w}(I_k)} (V_{0, \gamma}^{\otimes q} \otimes V_{1, \gamma_1})(\mathbf{X}_j) \right] &\leq e^{\gamma_0 \bar{w} j} \mathbb{E}^{\mathbf{x}} [(V_{0, \gamma}^{\otimes q} \otimes V_{1, \gamma_1})(\mathbf{X}_j)] \\ &\leq C e^{\gamma_0 \bar{w} j} (V_{0, \gamma}^{\otimes q} \otimes V_{1, \gamma_1})(\mathbf{x}), \end{aligned} \quad (5.32)$$

where, in the last inequality, we used (E-1) and thus the constant  $C$  only depends on  $\gamma, \gamma_1$ . Let now  $m, l \geq 1$  and write  $m = nl + j$  with  $j \in \{0, \dots, l\}$ . Then, using (5.32), the Hölder inequality and (5.32) again, we get

$$\begin{aligned} &\mathbb{E}^{\mathbf{x}} \left[ e^{\gamma_0 \sum_{k=1}^m \mathbf{w}(I_k)} (V_{0, \gamma}^{\otimes q} \otimes V_{1, \gamma_1})(\mathbf{X}_m) \right] \\ &\leq C e^{\gamma_0 \bar{w} j} \mathbb{E}^{\mathbf{x}} \left[ e^{\gamma_0 \sum_{k=1}^{nl} \mathbf{w}(I_k)} (V_{0, \gamma}^{\otimes q} \otimes V_{1, \gamma_1})(\mathbf{X}_{nl}) \right] \\ &= C e^{\gamma_0 \bar{w} j} \mathbb{E}^{\mathbf{x}} \left[ \left\{ \prod_{i=1}^l e^{\gamma_0 \sum_{k=0}^{n-1} \mathbf{w}(I_{kl+i})} \right\} (V_{0, \gamma}^{\otimes q} \otimes V_{1, \gamma_1})(\mathbf{X}_{nl}) \right] \\ &\leq C e^{\gamma_0 \bar{w} j} \left( \prod_{i=1}^l \mathbb{E}^{\mathbf{x}} \left[ e^{l \gamma_0 \sum_{k=0}^{n-1} \mathbf{w}(I_{kl+i})} (V_{0, \gamma}^{\otimes q} \otimes V_{1, \gamma_1})(\mathbf{X}_{nl}) \right] \right)^{1/l} \\ &\leq C e^{\gamma_0 \bar{w} j} \left( \prod_{i=1}^l C e^{\gamma_0 \bar{w} (l-i)} \mathbb{E}^{\mathbf{x}} \left[ e^{l \gamma_0 \sum_{k=0}^{n-1} \mathbf{w}(I_{kl+i})} (V_{0, \gamma}^{\otimes q} \otimes V_{1, \gamma_1})(\mathbf{X}_{(n-1)l+i}) \right] \right)^{1/l} \\ &\leq C^2 e^{2\gamma_0 \bar{w} l} \left( \prod_{i=1}^l \mathbb{E}^{\mathbf{x}} \left[ e^{l \gamma_0 \sum_{k=0}^{n-1} \mathbf{w}(I_{kl+i})} (V_{0, \gamma}^{\otimes q} \otimes V_{1, \gamma_1})(\mathbf{X}_{(n-1)l+i}) \right] \right)^{1/l}. \end{aligned} \quad (5.34)$$

Applying [28, Theorem 15.0.1],  $\tilde{Q}$  has a unique invariant probability measure  $\tilde{\pi}$  and there exists  $R > 0$  and  $\kappa > 0$  such that, for all  $i \in \{0, \dots, p\}$ ,  $\mathbf{s} \in \mathbb{Z}_+^q$  and  $\ell \in \mathbb{R}_+^p$  and all positive integer  $m$ ,

$$\|\tilde{Q}^m((i, \mathbf{s}, \ell), \cdot) - \tilde{\pi}\|_{(\gamma, \gamma_1)} \leq R e^{-\kappa m} V_{0, \gamma}^{\otimes q}(\mathbf{s}) V_{1, \gamma_1}(\ell), \quad (5.35)$$

where for any signed measure  $\xi$  on  $\{0, \dots, p\} \times \mathbb{Z}_+^q \times \mathbb{R}_+^p$ , we set

$$\|\xi\|_{\gamma, \gamma_1} = \sup \{ \xi(g) : \|g / (1_{\{0, \dots, p\}} \otimes V_{0, \gamma}^{\otimes q} \otimes V_{1, \gamma_1})\|_{\infty} \leq 1 \}.$$

Define

$$w^*(i, \mathbf{s}, \ell) = e^{l \gamma_0 \mathbf{w}(i)} V_{0, \gamma}(\mathbf{s}) V_{1, \gamma_1}(\ell),$$

for all  $(i, \mathbf{s}, \ell) \in \{0, \dots, p\} \times \mathbb{Z}_+^q \times \mathbb{R}_+^p$ . Then

$$\|w^* / (1_{\{0, \dots, p\}} \otimes V_{0, \gamma}^{\otimes q} \otimes V_{1, \gamma_1})\|_{\infty} \leq e^{l \gamma_0 \bar{w}}.$$

Thus, (5.35) yields that, for all  $(i, \mathbf{s}, \ell) \in \{0, \dots, p\} \times \mathbb{Z}_+^q \times \mathbb{R}_+^p$ ,

$$\tilde{Q}^l((i, \mathbf{s}, \ell), w^*) \leq \tilde{\pi}(w^*) + R e^{-\kappa l} e^{l \gamma_0 \bar{w}} V_{0, \gamma}^{\otimes q}(\mathbf{s}) V_{1, \gamma_1}(\ell) \quad (5.36)$$

Now, by dominated convergence, we have, as  $a \rightarrow 0$ ,

$$\tilde{\pi} \left( a^{-1} \left[ \exp(a \mathbf{w} \otimes 1_{\mathbb{Z}_+^q \times \mathbb{R}_+^p}) - 1 \right] \right) \rightarrow \tilde{\pi} \left( \mathbf{w} \otimes 1_{\mathbb{Z}_+^q \times \mathbb{R}_+^p} \right),$$

which is negative by Condition (E-2). Hence there exists  $a > 0$  such that

$$b := \tilde{\pi} \left( \exp(a \mathbf{w} \otimes 1_{\mathbb{Z}_+^q \times \mathbb{R}_+^p}) \right) < 1. \quad (5.37)$$

Now, by dominated convergence, we may decrease  $\gamma^*$  so that  $\gamma, \gamma_1 \in (0, \gamma^*]$  implies  $\tilde{\pi}(w^*) \leq b^{1/2}$  for all  $l$  and  $\gamma_0$  satisfying  $l\gamma_0 = a$ . Then we have for these  $a$  and  $\gamma, \gamma_1$ , which set the values of  $R$  and  $\kappa$ , and for all  $l$  and  $\gamma_0$  satisfying  $l\gamma_0 = a$ ,

$$\tilde{Q}^l((i, \mathbf{s}, \boldsymbol{\ell}), w^*) \leq \left( b^{1/2} + R e^{-\kappa l} e^{a\bar{w}} \right) V_{0,\gamma}^{\otimes q}(\mathbf{s}) V_{1,\gamma_1}(\boldsymbol{\ell}).$$

We now denote by  $L$  some integer such that  $b^{1/2} + R e^{-\kappa l} e^{a\bar{w}} \leq b^{1/4}$  for all  $l \geq L$ . Then, provided that  $l\gamma_0 = a$  and  $l \geq L$ , we get

$$\tilde{Q}^l((i, \mathbf{s}, \boldsymbol{\ell}), w^*) \leq b^{1/4} V_{0,\gamma}^{\otimes q}(\mathbf{s}) V_{1,\gamma_1}(\boldsymbol{\ell}).$$

Using this bound, by successive conditioning, we get that, for all  $n \geq 1$  and all  $1 \leq j \leq l$ , and  $\mathbf{x} = (\mathbf{s}, \boldsymbol{\ell})$ , if  $l\gamma_0 = a$ ,

$$\begin{aligned} & \mathbb{E}^{\mathbf{x}} \left[ e^{l\gamma_0 \sum_{k=0}^{n-1} \mathbf{w}(I_{kl+i})} (V_{0,\gamma}^{\otimes q} \otimes V_{1,\gamma_1})(\mathbf{X}_{(n-1)l+i}) \right] \\ & \leq b^{(n-1)/4} \mathbb{E}^{\mathbf{x}} \left[ e^{l\gamma_0 \mathbf{w}(I_i)} (V_{0,\gamma}^{\otimes q} \otimes V_{1,\gamma_1})(\mathbf{X}_i) \right] \\ & \leq C e^{a\bar{w}i} b^{(n-1)/4} (V_{0,\gamma}^{\otimes q} \otimes V_{1,\gamma_1})(\mathbf{x}), \end{aligned}$$

where, in the last inequality, we used  $e^{l\gamma_0 \mathbf{w}(I_j)} \leq e^{a\bar{w}}$  and Condition (E-1) and thus  $C$  is some constant not depending on  $n$ . Inserting this in (5.34), we get, for  $m = nl + j$  with  $j \in \{0, \dots, l\}$  and if  $l\gamma_0 = a$  with  $l \geq L$ ,

$$\begin{aligned} & \mathbb{E}^{\mathbf{x}} \left[ e^{\gamma_0 \sum_{k=1}^m \mathbf{w}(I_k)} (V_{0,\gamma}^{\otimes q} \otimes V_{1,\gamma_1})(\mathbf{X}_m) \right] \\ & \leq C^3 e^{3a\bar{w}} b^{(n-1)/4} (V_{0,\gamma}^{\otimes q} \otimes V_{1,\gamma_1})(\mathbf{x}) \\ & \leq C^3 e^{3a\bar{w}} b^{-1/2} b^{m/(4l)} (V_{0,\gamma}^{\otimes q} \otimes V_{1,\gamma_1})(\mathbf{x}) \\ & \leq C^3 e^{3a\bar{w}} b^{-1/2} b^{\gamma_0 m/(4a)} (V_{0,\gamma}^{\otimes q} \otimes V_{1,\gamma_1})(\mathbf{x}) \end{aligned} \quad (5.38)$$

Now, this bound holds for any  $\gamma_0 = a/l$  with  $l \geq L$ , where the constants  $a$  and  $b$  are chosen so that (5.37) holds and  $L$  only depends on  $a, b, \gamma$  and  $\gamma_1$ . This yields (5.31) by taking  $\rho > 0$  such that  $e^\rho < b^{-1/(4a)}$ . Note that we have shown that (5.31) holds for all  $\gamma_0 \in (0, a/L]$  but only if  $a/\gamma_0$  is an integer. Nevertheless, it is easy to show that this can be extended to all  $\gamma_0 \in (0, \gamma_0^*]$  by choosing  $\rho$  and  $\gamma_0^*$  adequately.  $\square$

If  $q = 1$  in Proposition 5.5, then we only need to prove Condition (C) for the chain  $\tilde{Q}^{-\{1\}}$ , which then corresponds to  $q = 0$ , that is  $V_{0,\gamma_0}^{\otimes q} \equiv 1$ . Hence an immediate consequence of Lemma 5.6 is that if  $q = 1$  in Proposition 5.5, Condition (C) is implied by the geometric ergodicity of the unconstrained chain  $\tilde{Q}$  (see Section 2.4) and a simple moment condition on the stationary distribution  $\tilde{\pi}$  of this chain, see (E-2).

However, for  $q \geq 2$ , Lemma 5.6 is no longer sufficient to prove Condition (C) of Proposition 5.5 because of the presence of  $V_{0,\gamma_0}^{\otimes q}$  in the denominator of (5.31). Again we shall rely on an inductive reasoning to obtain the following result.

**Lemma 5.7.** *Let  $p, q \geq 1$ . Let  $\{(I_n, \mathbf{S}_n, \boldsymbol{\lambda}_n), n \geq 0\}$  be the Markov chain on the space  $\{0, 1, \dots, p\} \times \mathbb{Z}_+^q \times \mathbb{R}_+^p$  with transition kernel  $\tilde{Q}$  defined in Section 2.4. For any  $\mathcal{J} \subseteq \{1, \dots, q\}$ , further define the transition kernel  $\tilde{Q}^{(-\mathcal{J})}$  on the space  $\{0, \dots, p\} \times \mathbb{Z}_+^{q-\#\mathcal{J}} \times \mathbb{R}_+^p$ , as in Section 2.4. Let  $\mathbf{w} : \{0, 1, \dots, p\} \rightarrow \mathbb{R}$ . Suppose that for any subset  $\mathcal{J}$  in  $\{1, \dots, q\}$ , the transition kernel  $\tilde{Q}^{(-\mathcal{J})}$  satisfies Assumptions (E-1) and (E-2). Then for all  $\gamma_1 > 0$  small enough, there are positive constants  $\rho_0$  and  $\gamma_0^* > 0$  such that, for all  $\gamma_0 \in (0, \gamma_0^*]$ ,*

$$\sup_{(\mathbf{s}, \boldsymbol{\ell}) \in \mathcal{X}} \sup_{m \geq 1} \mathbb{E}^{\mathbf{x}} \left[ e^{\gamma_0 \sum_{k=1}^m \{\mathbf{w}(I_k) + \rho_0\}} V_{1,\gamma_1}(\boldsymbol{\lambda}_m - \boldsymbol{\ell}) \right] < \infty. \quad (5.39)$$

*Proof.* We prove this lemma by induction on  $q$ .

For  $q = 0$ , Lemma 5.6 holds with  $V_{0,\gamma}^{\otimes q} \equiv 1$ , which implies (5.39) for all  $\gamma_1 > 0$  small enough with some  $\rho_0, \gamma_0^*$  only depending on  $\gamma_1$ . This provides the case  $q = 0$  which initiates the induction.

Next we show the result in the case  $q \geq 1$  using the induction hypothesis that the result holds for lower values of  $q$ .

Define for any  $j = 1, \dots, q$ ,  $\mathbf{A}(j)$  as in (5.28) and

$$s^*(j) = 1 + \max(\mathbf{A}(j)) \quad \text{and} \quad \bar{s}^* = \max_{j=1, \dots, q} s^*(j).$$

If  $\mathbf{x} = (\mathbf{s}, \ell) \in \mathbf{X}$  is such that

$$\mathbf{s} \in \bar{\mathbf{A}} = \{1, \dots, s^*(1)\} \times \dots \times \{1, \dots, s^*(q)\},$$

then, applying Lemma 5.6, for all  $\gamma_1 > 0$  small enough, we can choose  $\gamma, \rho, \gamma^* > 0$  such that, for all  $\gamma_0 \in (0, \gamma^*]$ , there is some constant  $C_0 > 0$  such that, for all  $m \geq 1$ ,

$$\begin{aligned} \mathbb{E}^{\mathbf{x}} \left[ e^{\gamma_0 \sum_{k=1}^m \mathbf{w}(I_k)} V_{1,\gamma_1}(\boldsymbol{\lambda}_m) \right] &\leq \mathbb{E}^{\mathbf{x}} \left[ e^{\gamma_0 \sum_{k=1}^m \mathbf{w}(I_k)} (V_{0,\gamma}^{\otimes q} \otimes V_{1,\gamma_1})(\mathbf{X}_m) \right] \\ &\leq C_0 e^{-\rho\gamma_0 m} (V_{0,\gamma}^{\otimes q} \otimes V_{1,\gamma_1})(\mathbf{x}) \\ &\leq C_0 e^{q\bar{s}^*\gamma} e^{-\rho\gamma_0 m} V_{1,\gamma_1}(\ell). \end{aligned} \quad (5.40)$$

We denote by  $\tau$  the first hitting time of the set  $\bar{\mathbf{A}}$ ,

$$\tau = \inf \{k \geq 0, \mathbf{S}_k \in \bar{\mathbf{A}}\}$$

and we set  $\tau_m = \tau \wedge m$ , which is a stopping time with respect to the filtration  $\mathcal{F}_k = \sigma(\mathbf{Y}_n, 0 \leq n \leq k)$ ,  $k \geq 0$ . We define by  $\mathcal{J}_m$  the set of indices  $j$  such that  $\mathbf{S}_k(j)$  stay away of  $\mathbf{A}(j)$  for all  $k = 0, 1, \dots, \tau_m$ , which is a random set measurable with respect to  $\mathcal{F}_{\tau_m}$  and is non-empty by definition of  $\tau_m$ . By definition of  $\tau_m$ , we have, in the event  $\{\tau_m < m\}$ ,  $\mathbf{S}_{\tau_m} \in \bar{\mathbf{A}}$  and thus, by (5.40), setting  $C_1 = C_0 e^{q\bar{s}^*\gamma}$ ,

$$\mathbb{E} \left[ e^{\gamma_0 \sum_{k=\tau_m+1}^m \mathbf{w}(I_k)} V_{1,\gamma_1}(\boldsymbol{\lambda}_m) \mid \mathcal{F}_{\tau_m} \right] \leq C_1 e^{-\rho\gamma_0(m-\tau_m)} V_{1,\gamma_1}(\boldsymbol{\lambda}_{\tau_m}) \quad \text{a.s.} \quad (5.41)$$

On the other hand, up to  $n = \tau_m$  the chain  $\mathbf{X}_n$  (with transition kernel  $\tilde{Q}$ ) coincides with the chain with transition kernel  $\tilde{Q}^{(-\mathcal{J}_m)}$  that starts at the same state, where  $\mathcal{J}_m$  is defined as above. Hence we may write, for all integer  $l = 1, 2, \dots, m$ , all non-empty subsets  $\mathcal{J}$  of  $\{1, \dots, q\}$  and  $\mathbf{x} = (\mathbf{s}, \ell) \in \mathbf{X}$  satisfying  $\mathbf{s} \notin \bar{\mathbf{A}}$ ,

$$\begin{aligned} &\mathbb{E}^{\mathbf{x}} \left[ e^{\gamma_0 \sum_{k=1}^{\tau_m} \mathbf{w}(I_k)} e^{-\rho\gamma_0(m-\tau_m)} V_{1,\gamma_1}(\boldsymbol{\lambda}_{\tau_m}) \mathbf{1}_{\{\tau_m=l\} \cap \{\mathcal{J}_m=\mathcal{J}\}} \right] \\ &= \mathbb{E}^{(-\mathcal{J}), \mathbf{x}} \left[ e^{\gamma_0 \sum_{k=1}^l \mathbf{w}(\tilde{I}_k)} e^{-\rho\gamma_0(m-l)} V_{1,\gamma_1}(\tilde{\boldsymbol{\lambda}}_l) \mathbf{1}_{\{\tau_m=l\} \cap \{\mathcal{J}_m=\mathcal{J}\}} \right] \\ &\leq e^{-\rho\gamma_0(m-l)} \mathbb{E}^{(-\mathcal{J}), \mathbf{x}} \left[ e^{\gamma_0 \sum_{k=1}^l \mathbf{w}(\tilde{I}_k)} V_{1,\gamma_1}(\tilde{\boldsymbol{\lambda}}_l) \right], \end{aligned}$$

where  $\{(\tilde{I}_n, \tilde{\mathbf{S}}_n, \tilde{\boldsymbol{\lambda}}_n), n \geq 0\}$  here denotes the chain with transition kernel  $\tilde{Q}^{(-\mathcal{J})}$  which starts at the same state as  $(I_0, \mathbf{S}_0, \boldsymbol{\lambda}_0)$ . Using the induction hypothesis, if  $\gamma_1 > 0$  is small enough, there are positive constants  $\gamma'$  and  $\rho'$  such that for all  $\gamma_0 \in (0, \gamma']$ ,  $\mathbf{x} \in \mathbf{X}$  and  $l \geq 1$ ,

$$\mathbb{E}^{(-\mathcal{J}), \mathbf{x}} \left[ e^{\gamma_0 \sum_{k=1}^l \mathbf{w}(\tilde{I}_k)} V_{1,\gamma_1}(\tilde{\boldsymbol{\lambda}}_l) \right] \leq C_2 e^{-\rho'\gamma_0 l} V_{1,\gamma_1}(\ell),$$

for some constant  $C_2 > 0$  independent of  $l$  and  $\mathbf{x}$ . Inserting this bound in the previous display and summing over all  $l = 0, 1, \dots, m$  and  $\mathcal{J}$ , we get, for all  $\gamma_0 \in (0, \gamma']$  and

$\mathbf{x} = (\mathbf{s}, \ell) \in \mathbf{X}$  such that  $\mathbf{s} \notin \overline{\mathbf{A}}$ ,

$$\begin{aligned} \mathbb{E}^{\mathbf{x}} \left[ e^{\gamma_0 \sum_{k=1}^{\tau_m} \mathbf{w}(I_k)} e^{-\rho \gamma_0 (m - \tau_m)} V_{1, \gamma_1}(\boldsymbol{\lambda}_{\tau_m}) \right] &\leq C_3 \sum_{l=0}^m e^{-\rho \gamma_0 (m-l) - \rho' \gamma_0 l} V_{1, \gamma_1}(\ell) \\ &\leq C_3 m e^{-(\rho \wedge \rho') \gamma_0 m} V_{1, \gamma_1}(\ell), \end{aligned}$$

where  $C_3 > 0$  is independent of  $m$  and  $\mathbf{x}$ . This last bound, with (5.41) yields, for all  $\gamma_0 \in (0, \gamma' \wedge \gamma^*]$  and  $\mathbf{x} = (\mathbf{s}, \ell) \in \mathbf{X}$ ,

$$\begin{aligned} \mathbb{E}^{\mathbf{x}} \left[ e^{\gamma_0 \sum_{k=1}^m \mathbf{w}(I_k)} V_{1, \gamma_1}(\boldsymbol{\lambda}_m) \right] &\leq \mathbb{E}^{\mathbf{x}} \left[ e^{\gamma_0 \sum_{k=1}^{\tau_m} \mathbf{w}(I_k)} \mathbb{E} \left[ e^{\gamma_0 \sum_{k=\tau_m+1}^m \mathbf{w}(I_k)} V_{1, \gamma_1}(\boldsymbol{\lambda}_m) \mid \mathcal{F}_{\tau_m} \right] \right] \\ &\leq C_1 \mathbb{E}^{\mathbf{x}} \left[ e^{\gamma_0 \sum_{k=1}^{\tau_m} \mathbf{w}(I_k)} e^{-\rho(m - \tau_m)} V_{1, \gamma_1}(\boldsymbol{\lambda}_{\tau_m}) \right] \\ &\leq C_4 m e^{-(\rho \wedge \rho') m} V_{1, \gamma_1}(\ell), \end{aligned}$$

where  $C_4$  is some positive constant independent of  $m$  and  $\mathbf{x}$ . Now, taking  $\gamma_0^* = \gamma' \wedge \gamma^*$   $\rho_0 \in (0, \rho \wedge \rho')$ , we get (5.39).  $\square$

**Theorem 5.8.** *Let  $p, q \geq 1$  and suppose that Assumptions 1 and 2 hold. Let  $\{(I_n, \mathbf{S}_n, \boldsymbol{\lambda}_n), n \geq 0\}$  be the Markov chain on the space  $\{0, 1, \dots, p\} \times \mathbb{Z}_+^q \times \mathbb{R}_+^p$  with transition kernel  $\tilde{Q}$  defined in Section 2.4. For any  $\mathcal{J} \subseteq \{1, \dots, q\}$ , further define the transition kernel  $\tilde{Q}^{(-\mathcal{J})}$  on the space  $\{0, \dots, p\} \times \mathbb{Z}_+^{q-\#\mathcal{J}} \times \mathbb{R}_+^p$ , as in Section 2.4. Define*

$$\mathbf{w}(i) = \begin{cases} \sum_{j=1}^q \vec{\mathbf{J}}_{i,j} & \text{for all } i = 1, \dots, p \\ 0 & \text{for } i = 0. \end{cases} \quad (5.42)$$

Suppose that for any subset  $\mathcal{J}$  in  $\{1, \dots, q\}$ , the transition kernel  $\tilde{Q}^{(-\mathcal{J})}$  satisfies

(B-1) For all  $\gamma' > 0$  small enough, there exists  $\gamma^* > 0$  such that for all  $\gamma \in (0, \gamma^*]$ ,  $\tilde{Q}^{(-\mathcal{J})}$  is  $(1_{\{0,1,\dots,p\}} \otimes V_{0,\gamma}^{\otimes(q-\#\mathcal{J})} \otimes V_{1,\gamma'})$ -geometrically ergodic.

(B-2) The stationary distribution  $\tilde{\pi}^{(-\mathcal{J})}$  of  $\tilde{Q}^{(-\mathcal{J})}$  satisfies  $\tilde{\pi}^{(-\mathcal{J})} \left[ \mathbf{w} \otimes 1_{\mathbb{Z}_+^q \times \mathbb{R}_+^p} \right] < 0$ .

Then, for all  $\gamma_1 > 0$  small enough, there exists  $\gamma_0^* > 0$  such that for all  $\gamma_0 \in (0, \gamma_0^*]$ , the transition kernel  $\tilde{Q}$  is  $(1_{\{0,1,\dots,p\}} \otimes V_{0,\gamma_0}^{\otimes q} \otimes V_{1,\gamma_1})$ -geometrically ergodic.

In practice, Theorem 5.8 should be applied by induction, so that only Conditions of the form (B-2) have to be checked out. The case  $q = 0$  is treated in Proposition 3.3. We treat the case  $q = 1$  in the following section. Then we explain in Section 3.5 how to perform this induction for  $q \geq 2$ .

*Proof of Theorem 5.8.* By Conditions (B-1) and (B-2), we may apply Lemma 5.7 to show that Condition (C) of Proposition 5.5 holds. Hence, applying this proposition, we get the drift condition (5.25).

The small sets of Theorem 3.1 are of the form  $\{1, \dots, K\}^q \times (0, M]^p$ . Observe that the sublevel sets of  $V_{0,\gamma_0}^q \otimes V_{1,\gamma_1}$

$$C_{\gamma_0, \gamma_1}(r) = \{ \mathbf{x} \in \mathbb{Z}_+ \times \mathbb{R}_+^p : V_{0,\gamma_0}^q \otimes V_{1,\gamma_1}(\mathbf{x}) \leq r \}, \quad r > 0,$$

are included in such sets. Hence the drift condition (5.25) applies to show that the chain  $Q$  is  $(V_{0,\gamma_0} \otimes V_{1,\gamma_1})$ -geometrically ergodic, see [28, Chapter 15]. The fact that  $\tilde{Q}$  is  $(1_{\{0,\dots,p\}} \otimes V_{0,\gamma_0} \otimes V_{1,\gamma_1})$ -geometrically ergodic follows as in the proof of Proposition 3.3.  $\square$

*Proof of Theorem 3.7.* This result is obtained by induction on  $q$  using Theorem 5.8 and Theorem 3.5 to initiate the induction at  $q = 1$ .  $\square$

## 5.6 Proof of Corollary 3.8

We first show that for all  $D > 0$ ,  $K \geq 1$  and  $M > 0$ , the set  $C = (0, D] \times \{0, \dots, p\} \times \{1, \dots, K\}^q \otimes (0, M]^p$  is a petite set. By Corollary 5.4, there exists  $m \geq 1$ ,  $\epsilon > 0$  and a probability measure  $\tilde{n}\nu$  on  $\mathcal{Y}$  such that  $\{0, \dots, p\} \times \{1, \dots, K\}^q \otimes (0, M]^p$  is an  $(m, \epsilon, \tilde{\nu})$  and  $(m+1, \epsilon, \tilde{\nu})$ -small set for  $\tilde{Q}$ . For all  $\mathbf{z} = (\delta, i, \mathbf{s}, \boldsymbol{\ell}) \in C$ , under the initial condition  $\mathbf{Z}_0 = \mathbf{z}$ , we have  $\boldsymbol{\lambda}_{m-1} \in (0, M + (m-1)\bar{\alpha}\beta]^p$ . It follows that the density of  $\Delta_m$  is bounded from below by  $c_0 e^{-(m-1)c_1 t}$  over  $t \in \mathbb{R}_+$  for some positive constants  $c_0, c_1$  (see the proof of Lemma 5.1). We get that for all Borel sets  $A \subset \mathbb{R}_+$  and  $B \subset \mathcal{Y}$ ,

$$\begin{aligned} \bar{Q}^m(\mathbf{z}, A \times B) &\geq c_0 \int_A e^{-c_1 t} dt \tilde{Q}^m(\mathbf{y}, B) \\ &\geq \epsilon(c_0/c_1)\nu(A \times B), \end{aligned}$$

where we used that  $\{0, \dots, p\} \times \{1, \dots, K\}^q \otimes (0, M]^p$  is an  $(m, \epsilon, \tilde{\nu})$ -small set for  $\tilde{Q}$  and  $\nu$  is the probability measure defined by  $\nu(A \times B) = \int_A c_1 e^{-c_1 t} dt \tilde{\nu}(B)$ . Hence  $C$  is an  $(m, \epsilon, \nu)$ -small set for  $\bar{Q}$ . Similarly, we have that  $C$  is also an  $(m+1, \epsilon, \tilde{\nu})$ -small set. Therefore, we conclude that  $\bar{Q}$  is  $\psi$ -irreducible and aperiodic and that all bounded subsets of  $Z$  are petite sets.

Next we show that the drift condition obtained in Proposition 5.5 for  $\tilde{Q}$  extends to  $\bar{Q}$ . As explained in Section 3.5, the assumptions of Theorem 3.7 imply that of Theorem 5.8 and thus of Proposition 5.5 (see the proof of Theorem 3.7). Hence, for all  $\gamma_1 > 0$  small enough, there exists  $\gamma_0^* > 0$ ,  $\theta \in (0, 1)$ ,  $b > 0$  and  $m \geq 1$  such that for all  $\gamma_0 \in (0, \gamma_0^*]$  such that for any initial condition  $\mathbf{x} \in \mathbf{X}$ , we have

$$\left[ \tilde{Q}^m(V_{0, \gamma_0} \otimes V_{1, \gamma_1}) \right] (\mathbf{x}) \leq \theta (V_{0, \gamma_0} \otimes V_{1, \gamma_1})(\mathbf{x}) + b.$$

By the Cauchy-Schwartz inequality and using that for all  $\mathbf{x} \in \mathbf{X}$ ,  $\Delta_m$  is stochastically smaller than an exponential distribution with mean  $(\mathbf{1}_p^T \boldsymbol{\mu}_0)^{-1}$  (see the proof of Lemma 5.1), we have

$$\begin{aligned} &\bar{Q}^m(V_{2, \gamma_2} \otimes \mathbf{1}_{\{0, \dots, p\}} \otimes V_{0, \gamma_0/2} \otimes V_{1, \gamma_1/2})(\Delta, i, \mathbf{x}) \\ &\leq \left( \mathbb{E}^{\mathbf{x}} [V_{2, 2\gamma_2}(\Delta_m)] \right)^{1/2} \left( \left[ \bar{Q}^m(\mathbf{1}_{\{0, \dots, p\}} \otimes V_{0, \gamma_0} \otimes V_{1, \gamma_1}) \right] (\Delta, i, \mathbf{x}) \right)^{1/2} \\ &\leq \left( \int_0^{+\infty} (\mathbf{1}_p^T \boldsymbol{\mu}_0) e^{2\gamma_2 r} e^{-\mathbf{1}_p^T \boldsymbol{\mu}_0 r} dr \right)^{1/2} (\theta (V_{0, \gamma_0} \otimes V_{1, \gamma_1})(\mathbf{x}) + b)^{1/2}. \end{aligned}$$

Provided that  $\gamma_2 < \mathbf{1}_p^T \boldsymbol{\mu}_0/2$ , we thus get

$$\begin{aligned} &\bar{Q}^m(V_{2, \gamma_2} \otimes \mathbf{1}_{\{0, \dots, p\}} \otimes V_{0, \gamma_0/2} \otimes V_{1, \gamma_1/2})(\Delta, i, \mathbf{x}) \\ &\leq (1 - 2\gamma_2/\mathbf{1}_p^T \boldsymbol{\mu}_0)^{-1/2} \left[ \theta^{1/2} (V_{0, \gamma_0/2} \otimes V_{1, \gamma_1/2})(\mathbf{x}) + b^{1/2} \right]. \end{aligned}$$

Hence, for  $\gamma_2 > 0$  small enough, we have, for some  $\theta' \in (0, 1)$  and  $b' > 0$ ,

$$\bar{Q}^m(V_{2, \gamma_2} \otimes \mathbf{1}_{\{0, \dots, p\}} \otimes V_{0, \gamma_0/2} \otimes V_{1, \gamma_1/2})(\Delta, i, \mathbf{x}) \leq \theta' (V_{0, \gamma_0/2} \otimes V_{1, \gamma_1/2})(\mathbf{x}) + b'.$$

Hence we obtain a drift condition with an unbounded off petite set function  $V_{2, \gamma_2} \otimes \mathbf{1}_{\{0, \dots, p\}} \otimes V_{0, \gamma_0/2} \otimes V_{1, \gamma_1/2}$ . The geometric ergodicity follows (see [28, Section 15]).

## 5.7 Proof of Theorem 3.9

We need the following lemmas to complete the proof for this theorem. Denote the number of arrivals on  $[0, u]$  by  $\mathbf{N}_u = \mathbf{N}(\mathbf{1}_{(0, u]} \otimes \mathbf{1}_{\{0, \dots, p\}})$  and the time interval between the last

arrival before  $u$  and  $u$  by  $\Gamma_u = u - \sum_{k=1}^{\mathbf{N}_u} \Delta_k$ .

**Lemma 5.9.** *We have*

$$\lim_{u \rightarrow \infty} \frac{\mathbf{N}_u}{u} \rightarrow \frac{1}{\overline{\mathbb{E}}(\Delta_1)}, \text{ a.s. } [P_*].$$

*Proof.* By definition of  $\mathbf{N}_u$ , we have

$$\sum_{k=1}^{\mathbf{N}_u} \Delta_k \leq u \leq \sum_{k=1}^{\mathbf{N}_u+1} \Delta_k.$$

Since  $\overline{Q}$  is positive Harris, using [28, Theorem 17.3.2], we get

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \Delta_k = \overline{\mathbb{E}}(\Delta_1), \text{ a.s. } [P_*].$$

We deduce that

$$\lim_{u \rightarrow +\infty} \frac{u}{\mathbf{N}_u} = \lim_{u \rightarrow +\infty} \frac{\sum_{k=1}^{\mathbf{N}_u} \Delta_k}{\mathbf{N}_u} = \lim_{u \rightarrow +\infty} \frac{\sum_{k=1}^{\mathbf{N}_u+1} \Delta_k}{\mathbf{N}_u} = \overline{\mathbb{E}}(\Delta_1), \text{ a.s. } [P_*].$$

Hence the result.  $\square$

**Lemma 5.10.** *For all  $T \in \mathbb{R}_+$ , we have*

$$\sup_{\tau \in [0,1]} \Gamma_{\tau T}^{\mathbf{x}} = o_{P_*}(\sqrt{T}),$$

where as in the notation *a.s.*[ $P_*$ ], the subscript  $P_*$  in  $o_{P_*}$  indicates that the result holds for any initial distribution.

*Proof.* Let us define

$$n_T = \left\lceil \frac{T}{\overline{\mathbb{E}}(\Delta_1)} \right\rceil. \quad (5.43)$$

By definition of  $\Gamma_u$ , for all  $\tau \in [0, 1]$ , we have that  $\Gamma_{\tau T} \leq \max_{k=1, \dots, \mathbf{N}_T+1} \Delta_k$ .

For any  $\tau \in [0, 1]$ ,  $\epsilon > 0$  and  $T \in \mathbb{R}_+$ , we get, for all  $\mathbf{x} \in \mathbf{X}$ ,

$$\mathbb{P}(\Gamma_{\tau T} \geq \epsilon \sqrt{T}) \leq \mathbb{P}^{\mathbf{x}} \left( \max_{k=1, \dots, 2n_T} \Delta_k \geq \epsilon \sqrt{T} \right) + \mathbb{P}^{\mathbf{x}}(\mathbf{N}_T + 1 > 2n_T).$$

Using Lemma 5.9, we have  $\lim_{T \rightarrow \infty} \mathbb{P}^{\mathbf{x}}(\mathbf{N}_T + 1 > 2n_T) = 0$ . Observe that we have, for all

$k \geq 1$ ,  $\Delta_k \stackrel{st}{\leq} E_k$  where  $\{E_i\}_{i \geq 1}$  are the inter-arrivals of a Poisson point process with intensity  $1_p^T \ell$ . We get, for all  $\mathbf{x} \in \mathbf{X}$ ,

$$\lim_{T \rightarrow \infty} \mathbb{P}^{\mathbf{x}} \left( \max_{k=1, \dots, 2n_T} \Delta_k \geq \epsilon \sqrt{T} \right) \leq \lim_{T \rightarrow \infty} \mathbb{P}^{\mathbf{x}} \left( \max_{k=1, \dots, 2n_T} E_k \geq \epsilon \sqrt{T} \right) = 0,$$

which concludes the proof.  $\square$

We now provide a proof of Theorem 3.9 using Lemmas 5.9 and 5.10.

*Proof of Theorem 3.9.* For all  $t \in [0, 1]$ ,  $T \in \mathbb{R}_+$  and  $\mathbf{x} \in \mathbf{X}$ , we have

$$(\mathbf{N}^{\mathbf{x}}(1_{[0,tT]} \otimes \mathbf{w}) - tT E(\mathbf{w})) = \sum_{k=1}^{\mathbf{N}_{tT}^{\mathbf{x}}} (\mathbf{w}(I_k) - E(\mathbf{w})\Delta_k) - E(\mathbf{w})\Gamma_{tT}^{\mathbf{x}}.$$



We also observe that, defining  $n_T$  as in (5.43),

$$\sum_{k=1}^{\mathbf{N}_{tT}^x} (\mathbf{w}(I_k) - E(\mathbf{w})\Delta_k) = \Sigma_{n_T} \left( \frac{\mathbf{N}_{tT}^x}{n_T}, g \right) = s_{n_T} \left( \frac{\mathbf{N}_{tT}^x}{n_T}, g \right),$$

where we used that, by definition of  $E(\mathbf{w})$ ,  $\bar{\pi}(g) = 0$  and that  $\mathbf{N}_{tT}$  is an integer. Hence by Lemma 5.10, we get

$$\sup_{t \in [0,1]} \left| (\mathbf{N}^x(1_{[0,tT]} \otimes \mathbf{w}) - tT E(\mathbf{w})) - s_{n_T} \left( \frac{\mathbf{N}_{tT}^x}{n_T}, g \right) \right| = o_{P^*}(\sqrt{T}), \quad (5.44)$$

Now, by Lemma 5.9, we have, for all  $t \in [0, 1]$ ,  $\mathbf{N}_{tT}^x/n_T \rightarrow t$  as  $T \rightarrow \infty$  a.s. $[P^*]$ . By the Dini theorem, this convergence holds uniformly over  $t \in [0, 1]$ . Applying [34, Theorem 13.2.1], this convergence and the one in (3.15) implies that, for any initial distribution,

$$(n_T \sigma_g^2)^{-1/2} s_{n_T} \left( \frac{\mathbf{N}_{tT}^x}{n_T}, g \right) \xrightarrow{d} B_t,$$

where the weak convergence holds in  $D([0, 1])$ . This, (5.44),  $\sigma_g^2 = v(\mathbf{w})$  and the fact that  $n_T/T \rightarrow 1/\mathbb{E}[\Delta_1]$  yield (3.18), which achieves the proof.  $\square$

## 6 Conclusion

In this paper we have introduced and studied constrained multivariate Hawkes processes. The constraints are expressed using a multidimensional variable, whose evolution is driven by the point process. Under the Markov setting (exponential fertility functions), we have proven that the underlying Markov chain is  $V$ -geometrically ergodic under some conditions on the parameters. A converse result leading to the transience of the chain in the case where an univariate spread variable is used ( $q = 1$ ) illustrates the sharpness of our conditions. Moreover, in the general case, we used a functional central limit theorem applying to the chain to derive the scaling limit of the integrated point process in physical time. Finally we have briefly explained how the constrained multivariate Hawkes process can be applied to model the dynamics of a limit order book. The scaling limit of the mid-price can be deduced from our findings.

We only presented the case of the order book of one asset with only two limits. This is clearly not a restriction. Multi-assets limit order books can be considered, yielding multivariate boundary conditions (hence  $q \geq 2$ ). In a forthcoming paper, we will use this model to provide an empirical study from real data and discuss the potential application of this model to the dynamics of limit order books. We believe that the applicability of the constrained multivariate Hawkes process extends well beyond financial applications, as it could model the motion of an object on a discrete net with some boundary conditions.

## 7 Acknowledgements

This research is supported by NATIXIS quantitative research department.

## A Additional technical lemmas

The following result is used in the proof of Proposition 5.3.

**Lemma A.1.** *Let  $p \geq 1$ , and  $\Gamma$  is a  $p \times p$  invertible matrix. Then there exists a probability measure  $\nu$  such that, for all  $M > 0$ ,  $\gamma > 0$ , there exists  $\epsilon > 0$  such that for all  $g : \mathbb{R}^p \rightarrow \mathbb{R}_+$  and  $\ell \in (0, M]^p$ ,*

$$\int_D g(u\ell + \Gamma\vartheta)u^\gamma \mathrm{d}u \mathrm{d}\vartheta \geq \epsilon \int g \, \mathrm{d}\nu ,$$

where  $D = \{(u, \vartheta) \in (0, 1] \times (0, 1]^p, 0 < u < \vartheta(1) < \dots < \vartheta(p)\}$ .

*Proof.* Setting  $\omega = u\ell + \Gamma\vartheta$ , we have

$$\int_D g(u\ell + \Gamma\vartheta)u^\gamma \mathrm{d}u \mathrm{d}\vartheta = \frac{1}{\det \Gamma} \int g(\omega)u^\gamma 1_D((u, \Gamma^{-1}\omega - u\Gamma^{-1}\ell)) \mathrm{d}u \mathrm{d}\omega .$$

We choose any  $0 < \eta < 1/(p+1)$  so that the open set  $B_\eta = \{\vartheta \in (0, 1]^p : \eta < \vartheta(1), \eta + \vartheta(1) < \vartheta(2), \dots, \eta + \vartheta(p-1) < \vartheta(p), \vartheta(p) < 1 - \eta\}$  is not empty.

Denoting  $\Gamma B_\eta = \{\Gamma\vartheta : \vartheta \in B_\eta\}$  and setting

$$\eta' = \frac{1}{2}\eta \min \left( 1, \left( \max_{i=1, \dots, p} \sup_{\ell \in (0, M]^p} |[\Gamma^{-1}\ell](i)| \right)^{-1} \right)$$

we have that for all  $u \in (0, \eta']$ ,  $\ell \in (0, M]^p$  and  $\omega \in \Gamma B_\eta$ ,  $[u, \Gamma^{-1}\omega - u\Gamma^{-1}\ell] \in D$ . We obtain that

$$\begin{aligned} \int_D g(u\ell + \Gamma\vartheta)u^\gamma \mathrm{d}u \mathrm{d}\vartheta &\geq \frac{1}{\det \Gamma} \int_{\Gamma B_\eta} g(\omega) \left( \int_0^{\eta'} u^\gamma \mathrm{d}u \right) \mathrm{d}\omega \\ &= \frac{\mu^{Leb}(\Gamma B_\eta)(\eta')^{\gamma+1}}{(\gamma+1)\det \Gamma} \int g \mathrm{d}\nu , \end{aligned}$$

where  $\mu^{Leb}$  is Lebesgue measure,  $\nu$  is the uniform probability measure on  $\Gamma B_\eta$ .  $\square$

The following result is used in the proof of Proposition 3.2.

**Lemma A.2.** *Let  $\beta > 0$  and  $\beta' > 0$ . Then, as  $a \rightarrow \infty$ ,*

$$\int_0^\infty e^{a(e^{-\beta t} - 1) - \beta' t} \mathrm{d}t \rightarrow 0 . \tag{A.1}$$

*If moreover  $\beta' > \beta$ , then we have the following asymptotic equivalence as  $a \rightarrow \infty$ ,*

$$\int_0^\infty e^{a(e^{-\beta t} - 1) - \beta' t} \mathrm{d}t \sim \frac{1}{a\beta} . \tag{A.2}$$

*Proof.* Setting  $\vartheta = e^{-\beta t}$ , we get

$$\int_0^\infty e^{a(e^{-\beta t} - 1) - \beta' t} \mathrm{d}t = \frac{1}{\beta} \int_0^1 e^{a(\vartheta - 1)} \vartheta^{\frac{\beta'}{\beta} - 1} \mathrm{d}\vartheta .$$

Letting  $a \rightarrow \infty$ , by dominated convergence, we get (A.1).

Now we set  $\omega = -a(\vartheta - 1)$  and obtain that

$$\frac{1}{\beta} \int_0^1 e^{a(\vartheta - 1)} \vartheta^{\frac{\beta'}{\beta} - 1} \mathrm{d}\vartheta = \frac{1}{a\beta} \int_0^a e^{-\omega} \left(1 - \frac{\omega}{a}\right)^{\frac{\beta'}{\beta} - 1} \mathrm{d}\omega .$$

Letting  $a \rightarrow \infty$ , we obtain (A.2) by dominated convergence.  $\square$

## References

- [1] F. Abergel and A. Jedidi. A mathematical approach to order book modeling. *SSRN eLibrary*, 2011.
- [2] Y. Aït-Sahalia, J. Cacho-Diaz, and R.J.A. Laeven. Modeling financial contagion using mutually exciting jump processes. Working Paper 15850, National Bureau of Economic Research, mar 2010.
- [3] S. Asmussen. *Applied probability and queues*, volume 51 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 2003.
- [4] E. Bacry, K. Dayri, and J.F. Muzy. Non-parametric kernel estimation for symmetric hawkes processes. application to high frequency financial data. Papers, arXiv.org, 2011.
- [5] E. Bacry, S. Delattre, M. Hoffmann, and J. F. Muzy. Modelling microstructure noise with mutually exciting point processes. submitted to quantitative finance. 2010.
- [6] E. Bacry, S. Delattre, M. Hoffmann, and J.F. Muzy. Scaling limits for hawkes processes and application to financial statistics. *Preprint*, 2012.
- [7] L. Bauwens and N. Hautsch. Dynamic latent factor models for intensity processes. CORE Discussion Papers 2003103, Université catholique de Louvain, Center for Operations Research and Econometrics (CORE), 2003.
- [8] L. Bauwens and N. Hautsch. Modelling financial high frequency data using point processes. *Social Science Research Network Working Paper Series*, nov 2006.
- [9] B. Biais, P. Hillion, and C. Spatt. An empirical analysis of the limit order book and the order flow in the paris bourse. *The Journal of Finance*, 50:1655–1689, 1995.
- [10] J.P. Bouchaud, J. Doyne Farmer, and F. Lillo. How markets slowly digest changes in supply and demand. *ArXiv e-prints*, sep 2008.
- [11] C. Bowsher. Modelling security market events in continuous time: Intensity based, multivariate point process models. *Nuffield College Economics Discussion Papers*, nov 2002.
- [12] C. Bowsher. Modelling security market events in continuous time: Intensity based, multivariate point process models. *Journal of Econometrics*, 141:876–912, dec 2007.
- [13] P. Brémaud and L. Massoulié. Stability of nonlinear Hawkes processes. *Ann. Probab.*, 24:1563–1588, 1996.
- [14] A. Chakraborti, I. Muni Toke, M. Patriarca, and F. Abergel. Econophysics review: I. empirical facts. *Quantitative Finance*, 11:991–1012, 2011.
- [15] A. Chakraborti, I. Muni Toke, M. Patriarca, and F. Abergel. Econophysics review: Ii. agent-based models. *Quantitative Finance*, 11:1013–1041, 2011.
- [16] R. Cont and A. De Larrard. Price dynamics in a markovian limit order book market. *Social Science Research Network Working Paper Series*, jan 2011.
- [17] R. Cont, S. Stoikov, and R. Talreja. A stochastic model for order book dynamics. *Oper. Res.*, 58:549–563, 2010.
- [18] D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes. Vol. I. Probability and its Applications (New York)*. Springer-Verlag, New York, second edition, 2003. Elementary theory and methods.
- [19] J. Doyne Farmer, L. Gillemot, F. Lillo, S. Mike, and A. Sen. What really causes large price changes? Technical report, apr 2004.
- [20] P. Embrechts, T. Liniger, and L. Lin. Multivariate hawkes processes : an application to financial data. *Journal of Applied Probability*, 48A:367–378, 2011.

- [21] E. Errais, K. Giesecke, and L. R. Goldberg. Affine point processes and portfolio credit risk. *SIAM J. Financial Math.*, 1:642–665, 2010.
- [22] T. Fletcher, Z. Hussain, and J. Shawe-Taylor. Multiple kernel learning on the limit order book. *Quantitative Finance, preprint*, 2010.
- [23] C. Gouriéroux, J. Jasiak, and G. Le Fol. Intra-day market activity. *Journal of Financial Markets*, 2(3):193–226, aug 1999.
- [24] P. Hewlett. Clustering of order arrivals, price impact and trade path optimisation. Workshop on Financial Modeling, 2006.
- [25] B. Hollifield, R. A. Miller, and P. Sandås. Empirical analysis of limit order markets. *The Review of Economic Studies*, 71(4):1027–1063, oct 2004.
- [26] J.H. Large. Measuring the resiliency of an electronic limit order book. *Journal of Financial Markets*, 10(1):1–25, 2007.
- [27] T.F. Liniger. *Multivariate Hawkes Processes*. PhD thesis, ETH Zürich, 2009.
- [28] S. Meyn and R. L. Tweedie. *Markov chains and stochastic stability*. Cambridge University Press, Cambridge, second edition, 2009. With a prologue by Peter W. Glynn.
- [29] I. Muni Toke and F. Pomponio. Modelling trades-through in a limited order book using hawkes processes. *SSRN eLibrary*, 2011.
- [30] D. Oakes. The Markovian self-exciting process. *J. Appl. Probability*, 12:69–77, 1975.
- [31] C.A. Parlour. Price dynamics in limit order markets. *Review of Financial Studies*, 11(4):789–816, 1998.
- [32] I. Rocu. A dynamic model of the limit order book. *Review of Financial Studies*, 22(11):4601–4641, 2009.
- [33] E. Smith, J. Doyne Farmer, L. Gillemot, and S. Krishnamurthy. Statistical theory of the continuous double auction. *Quantitative Finance*, 3:481–514, 2003.
- [34] W. Whitt. *Stochastic-process limits*. Springer Series in Operations Research. Springer-Verlag, New York, 2002. An introduction to stochastic-process limits and their application to queues.
- [35] B. Zheng. *Détection d'événements rares dans les données hautes fréquences et applications au trading algorithmique*. PhD thesis, Télécom ParisTech, 2013.
- [36] B. Zheng, E. Moulines, and F. Abergel. Price jump prediction in limit order book. *SSRN eLibrary*, 2012.