# Limit theorems for nearly unstable Hawkes processes

Thibault Jaisson
CMAP, École Polytechnique Paris
thibault.jaisson@polytechnique.edu

Mathieu Rosenbaum
LPMA, Université Pierre et Marie Curie (Paris 6)
mathieu.rosenbaum@upmc.fr

October 8, 2013

**Abstract**

Because of their tractability and their natural interpretations in term of market quantities, Hawkes processes are nowadays widely used in high frequency finance. However, in practice, the statistical estimation results seem to show that very often, only *nearly unstable Hawkes processes* are able to fit the data properly. By nearly unstable, we mean that the $L^1$ norm of their kernel is close to unity. We study in this work such processes for which the stability condition is almost violated. Our main result states that after suitable rescaling, they asymptotically behave like integrated Cox Ingersoll Ross models. Thus, modeling financial order flows as nearly unstable Hawkes processes may be a good way to reproduce both their high and low frequency stylized facts. We then extend this result to the Hawkes based price model introduced by Bacry *et al.* in [6]. We show that under a similar criticality condition, this process converges to a Heston model. Again, we recover well known stylized facts of prices, both at the microstructure level and at the macroscopic scale.

**Keywords:** Point processes, Hawkes processes, limit theorems, microstructure modeling, high frequency data, order flows, Cox Ingersoll Ross model, Heston model.

## 1 Introduction

A Hawkes process $(N_t)_{t\geq 0}$ is a self exciting point process, whose intensity at time $t$, denoted by $\lambda_t$, is of the form

$$\lambda_t = \mu + \sum_{0<J_i<t} \phi(t-J_i) = \mu + \int_{(0,t)} \phi(t-s)dN_s,$$

where $\mu$ is a positive real number, $\phi$ a regression kernel and the $J_i$ are the points of the process before time $t$ (see Section 2 for more accurate definitions). These processes have been introduced in 1971 by Hawkes, see [22, 23, 24], in the purpose of modeling earthquakes and their aftershocks, see [1]. However, they are also used in various other disciplines. In particular, in recent years, with the availability of (ultra) high frequency data, finance has

become one of the main domains of application of Hawkes processes.

The introduction of Hawkes processes in finance is probably due to Bowsher, see [12], who jointly studied transaction times and midquote changes, using the Hawkes framework. Then, in [9], Bauwens and Hautsch built so-called latent factor intensity Hawkes models and applied them to transaction data. Another pioneer in this type of approach is Hewlett. He considered in [26] the particular case of the foreign exchange rates market for which he fitted a bivariate Hawkes process on buy and sell transaction data. More recently, Bacry *et al.* have developed a microstructure model for midquote prices based on the difference of two Hawkes processes, see [6]. Moreover, Bacry and Muzy have extended this approach in [7] where they design a framework enabling to study market impact. Beyond midquotes and transaction prices, full limit order book data (not only market orders but also limit orders and cancellations) have also been investigated through the lenses of Hawkes processes. In particular, Large uses in [33] a ten-variate multidimensional Hawkes process to this purpose. Note that besides microstruture problems, Hawkes processes have also been introduced in the study of other financial issues such as daily data analysis, see [17], financial contagion, see [2], or Credit Risk, see [18].

The popularity of Hawkes processes in financial modeling is probably due to two main reasons. First, these processes represent a very natural and tractable extension of Poisson processes. In fact, comparing point processes and conventional time series, Poisson processes are often viewed as the counterpart of iid random variables whereas Hawkes processes play the role of autoregressive processes, see [15] for more details about this analogy. Another explanation for the appeal of Hawkes processes is that it is often easy to give a convincing interpretation to such modeling. To do so, the branching structure of Hawkes processes is quite helpful. Recall that under the assumption $\|\phi\|_1 < 1$, where $\|\phi\|_1$ denotes the $L^1$ norm of $\phi$, Hawkes processes can be represented as a population process where migrants arrive according to a Poisson process with parameter $\mu$. Then each migrant gives birth to children according to a non homogeneous Poisson process with intensity function $\phi$, these children also giving birth to children according to the same non homogeneous Poisson process, see [24]. Now consider for example the classical case of buy (or sell) market orders, as studied in several of the papers mentioned above. Then migrants can be seen as exogenous orders whereas children are viewed as orders triggered by other orders.

Beyond enabling to build this population dynamics interpretation, the assumption $\|\phi\|_1 < 1$ is crucial in the study of Hawkes processes. To fix ideas, let us place ourselves in the classical framework where the Hawkes process $(N_t)$ starts at $-\infty$. In that case, if one wants to get a stationary intensity with finite first moment, then the condition $\|\phi\|_1 < 1$ is required. Furthermore, even in the non stationary setting, this condition seems to be necessary in order to obtain classical ergodic properties for the process, see [5]. For these reasons, this condition is often called stability condition in the Hawkes literature.

From a practical point of view, a lot of interest has been recently devoted to the parameter $\|\phi\|_1$. For example, Hardiman, Bercot and Bouchaud, see [21], and Filimonov and Sornette, see [19, 20], use the branching interpretation of Hawkes processes on midquote data in order to measure the so-called degree of endogeneity of the market. This degree is simply defined by $\|\phi\|_1$, which is also called branching ratio. The intuition behind this interpretation of $\|\phi\|_1$ goes as follows: The parameter $\|\phi\|_1$ corresponds to the average number of children of

an individual, $\|\phi\|_1^2$ to the average number of grandchildren of an individual,... Therefore, if we call cluster the descendants of a migrant, then the average size of a cluster is given by $\sum_{k\geq 1}\|\phi\|_1^k = \|\phi\|_1/(1-\|\phi\|_1)$. Thus, in the financial interpretation, the average proportion of endogenously triggered events is $\|\phi\|_1/(1-\|\phi\|_1)$ divided by $1+\|\phi\|_1/(1-\|\phi\|_1)$, which is equal to $\|\phi\|_1$.

This branching ratio can be measured using parametric and non parametric estimation methods for Hawkes processes, see [35, 36] for likelihood based methods and [4, 38] for functional estimators of the function $\phi$. In [21], very stable estimations of $\|\phi\|_1$ are reported for the E mini S&P futures between 1998 and 2012, the results being systematically close to one. In [19], values of order 0.7-0.8 are obtained on several assets. A debate on the validity of these results is currently ongoing between the two groups. In particular, it is argued in [21] that the choice of exponential kernels in [19] may lead to spurious results, whereas various bias that could affect the study in [21] are underlined in [20]. In any case, we can remark that both groups find values close to one for $\|\phi\|_1$, which is consistent with the results of [4], where estimations are performed on Bund and Dax futures.

This seemingly persistent statistical result should definitely worry users of Hawkes processes. Indeed, it is rarely suitable to apply a statistical model where the parameters are pushed to their limits. In fact, these obtained values for $\|\phi\|_1$ on empirical data are not really surprising. Indeed, one of the most well documented stylized fact in high frequency finance is the persistence (or long memory) in flows and market activity measures, see for example [11, 34]. Usual Hawkes processes, in the same way as autoregressive processes, can only exhibit short range dependence, failing to reproduce this classical empirical feature, see [29] for details.

In spite of their relative inadequacy with market data, Hawkes processes possess so many appealing properties that one could still try to apply them in some specific situations. In [21], it is suggested to use the without ancestors version of Hawkes processes introduced by Brémaud and Massoulié in [13]. For such processes, $\|\phi\|_1 = 1$ but, in order to preserve stationarity and a finite expectation for the intensity, one needs to have $\mu = 0$. This is probably a relevant approach. However setting the parameter $\mu$ to 0 is not completely satisfying since this parameter has a nice interpretation (exogenous orders). Moreover it is not found to be equal to zero in practice, see [21]. Finally, a time-varying $\mu$ is an easy way to reproduce seasonalities observed on the market, see [7] (however, for simplicity, we work in this paper with a constant $\mu > 0$).

These empirical measures of $\|\phi\|_1$, close to one, are the starting point of this work. Indeed, our aim is to study the behavior at large time scales of nearly unstable Hawkes processes, which correspond to these estimations. More precisely, we consider a sequence of Hawkes processes observed on $[0, T]$, where $T$ goes to infinity. In the case of a fixed kernel (not depending on $T$) with norm strictly smaller than one, scaling limits of Hawkes processes have been investigated in [5]. In this framework, Bacry et al. obtain a deterministic limit for the properly normalized sequence of Hawkes processes, as it is the case for suitably rescaled Poisson processes. In their price model consisting in the difference of two Hawkes processes, a Brownian motion (with some volatility) is found at the limit. These two results are in fact quite intuitive. Indeed, in the same way as Poisson processes and autoregressive models, Hawkes processes enjoy short memory properties. In this work, we show that when the

Hawkes processes are nearly unstable, these weakly dependent-like behaviors are no longer observed at intermediate time scales. To do so, we consider that the kernels of the Hawkes processes depend on $T$. More precisely, we translate the near instability condition into the assumption that the norm of the kernels tends to one as the observation scale $T$ goes to infinity.

Our main theorem states that when the norm of the kernel tends to one at the right speed (meaning that the observation scale and kernel's norm balance in a suitable way), the limit of our sequence of Hawkes processes is no longer a deterministic process, but an integrated Cox Ingersoll Ross process (CIR for short), as introduced in [14]. In practice, it means that when observing a Hawkes process with kernel's norm close to one at appropriate time scale, it looks like an integrated CIR. Furthermore, for the price model defined in [6], in the limit, the Brownian motion obtained in [5] is replaced by a Heston model, see [25] for definition. This is probably more in agreement with empirical data.

The paper is organized as follows. The assumptions and main results, notably the convergence towards an integrated CIR are given in Section 2. The case of the difference of two Hawkes processes is studied in Section 3. The proofs are relegated to Section 4 except some auxiliary results which can be found in an appendix.

## 2 Scaling limits of nearly unstable Hawkes processes

We give in this section our main results about the limiting behavior of a sequence of nearly unstable Hawkes processes. We start by presenting our assumptions and defining our asymptotic setting.

### 2.1 Assumptions and asymptotic framework

We consider a sequence of point processes $(N_t^T)_{t\geq 0}$ indexed by $T$[1]. For a given $T$, $(N_t^T)$ satisfies $N_0^T = 0$ and the process is observed on the time interval $[0, T]$. Furthermore, our asymptotic setting is that the observation scale $T$ goes to infinity. The intensity process $(\lambda_t^T)$ is defined for $t \geq 0$ by

$$\lambda_t^T = \mu + \int_0^t \phi^T(t-s)dN_s^T,$$

where $\mu$ is a positive real number and $\phi^T$ a non negative measurable function on $\mathbb{R}^+$ which satisfies $\|\phi\|_1 < +\infty$. For a given $T$, the process $(N_t^T)$ is defined on a probability space $(\Omega^T, \mathcal{F}^T, \mathbb{P}^T)$ equipped with the filtration $(\mathcal{F}_t^T)_{t\in[0,T]}$, where $\mathcal{F}_t^T$ is the $\sigma$-algebra generated by $(N_s^T)_{s\leq t}$. Moreover we assume that for any $0 \leq a < b \leq T$ and $A \in \mathcal{F}_a^T$

$$\mathbb{E}[(N_b^T - N_{a-}^T)1_A] = \mathbb{E}[\int_a^b \lambda_s^T 1_A ds],$$

which sets $\lambda^T$ as the intensity of $N^T$. In particular, if we denote by $(J_n^T)_{n\geq 1}$ the jump times of $(N_t^T)$, the process

$$N_{t\wedge J_n^T}^T - \int_0^{t\wedge J_n^T} \lambda_s^T ds$$

[1]Of course by $T$ we implicitly means $T_n$ with $n \in \mathbb{N}$ tending to infinity.

is a martingale and the law of $N^T$ is characterized by $\lambda^T$. From Jacod [27], such construction can be done. The process $N^T$ is called a Hawkes process.

Let us now give more specific assumptions on the function $\phi^T$. We denote by $\|.\|_\infty$ the $L^\infty$ norm on $\mathbb{R}^+$.

**Assumption 1.** *For $t \in \mathbb{R}^+$,*

$$\phi^T(t) = a_T \phi(t),$$

*where $(a_T)_{T\geq 0}$ is a sequence of positive numbers converging to one such that for all $T$, $a_T < 1$ and $\phi$ is a non negative measurable function such that*

$$\int_0^{+\infty} \phi(s)ds = 1 \text{ and } \int_0^{+\infty} s\phi(s)ds = m < \infty.$$

*Moreover, $\phi$ is differentiable with derivative $\phi'$ such that $\|\phi'\|_\infty < +\infty$ and $\|\phi'\|_1 < +\infty$.*

**Remark 2.1.** *Note that under Assumption 1, $\|\phi\|_\infty$ is finite.*

Thus, the form of the function $\phi^T$ depends on $T$ so that its shape is fixed but its $L^1$ norm varies with $T$. For a given $T$, this $L^1$ norm is equal to $a_T$ and so is smaller than one, implying that the stability condition is in force. Note that in this framework, we have almost surely no explosion[2]:

$$\lim_{n\to+\infty} J_n^T = +\infty.$$

However, remark that we do not work in the stationary setting since our process starts at time $t = 0$ and not at $t = -\infty$.

The case where $\|\phi^T\|_1$ is larger than one corresponds to the situation where the stability condition is violated. Since $a_T = \|\phi^T\|_1 < 1$ tends to one, our framework is a way to get close to instability. Therefore we call our processes nearly unstable Hawkes processes. There are of course many other ways to make the $L^1$ norm of $\phi^T$ converge to one than the multiplicative manner used here. However, this parametrization is sufficient for applications and very convenient to illustrate the different regimes that can be obtained.

## 2.2 Observation scales

In our framework, two parameters degenerate at infinity: $T$ and $(1 - a_T)^{-1}$. The relationship between these two sequences will determine the scaling behavior of the sequence of Hawkes processes. Recall that it is shown in [5] that when $\|\phi\|_1$ is fixed and smaller than one, after appropriate scaling, the limit of the sequence of Hawkes processes is deterministic, as it is for example the case for Poisson processes. In our setting, if $1 - a_T$ tends "slowly" to zero, we can expect the same result. Indeed, we may have $T$ large enough so that we reach the asymptotic regime and for such $T$, $a_T$ is still sufficiently far from unity. This is precisely what happens, as stated in the next theorem.

[2] In fact, for a Hawkes process, the no explosion property can be obtained under weaker conditions, for example $\int_0^t \phi(s)ds < \infty$ for any $t > 0$, see [5].

**Theorem 2.1.** *Assume $T(1-a_T) \to +\infty$. Then, under Assumption 1, the sequence of Hawkes processes is asymptotically deterministic, in the sense that the following convergence in $L^2$ holds:*

$$\sup_{v\in[0,1]} \frac{1-a_T}{T} |N^T_{Tv} - \mathbb{E}[N^T_{Tv}]| \to 0.$$

On the contrary, if $1-a_T$ tends too rapidly to zero, the situation is likely to be quite intricate. Indeed, for given $T$, the Hawkes process may already be very close to instability whereas $T$ is not large enough to reach the asymptotic regime. The last case, which is probably the most interesting one, is the intermediate case, where $1-a_T$ tends to zero in such a manner that a non deterministic scaling limit is obtained, while not being in the preceding degenerate setting. We largely detail this situation in the next subsection.

### 2.3 Non degenerate scaling limit for nearly unstable Hawkes processes

We give in this section our main result, that is a non degenerate scaling limit for a sequence of properly renormalized nearly unstable Hawkes processes. Before giving this theorem, we wish to provide intuitions on how it is derived. Let $M^T$ be the martingale process associated to $N^T$, that is, for $t \geq 0$,

$$M^T_t = N^T_t - \int_0^t \lambda^T_s ds.$$

We also set $\psi^T$ the function defined on $\mathbb{R}^+$ by

$$\psi^T(t) = \sum_{k=1}^{\infty} (\phi^T)^{*k}(t),$$

where $(\phi^T)^{*1} = \phi^T$ and for $k \geq 2$, $(\phi^T)^{*k}$ denotes the convolution product of $(\phi^T)^{*(k-1)}$ with the function $\phi^T$. Note that $\psi^T(t)$ is well defined since $\|\phi^T\|_1 < 1$. In the sequel, it will be convenient to work with another form for the intensity. We have the following result, whose proof is given in Section 4.

**Proposition 2.1.** *For all $t \geq 0$, we have*

$$\lambda^T_t = \mu + \int_0^t \psi^T(t-s)\mu ds + \int_0^t \psi^T(t-s) dM^T_s.$$

Now recall that we observe the process $(N^T_t)$ on $[0,T]$. In order to be able to give a proper limit theorem, where the processes live on the same time interval, we rescale our processes so that they are defined on $[0,1]$. To do that, we consider for $t \in [0,1]$

$$\lambda^T_{tT} = \mu + \int_0^{tT} \psi^T(Tt-s)\mu ds + \int_0^{tT} \psi^T(Tt-s) dM^T_s.$$

For the scaling in space, a natural multiplicative factor is $(1-a_T)$. Indeed, in the stationary case, the expectation of $\lambda^T_t$ is $\mu/(1-\|\phi^T\|_1)$. Thus, the order of magnitude of the intensity is $(1-a_T)^{-1}$. This is why we define

$$C^T_t = \lambda^T_{tT}(1-a_T). \tag{1}$$

Understanding the asymptotic behavior of $C_t^T$ will be the key to the derivation of a suitable scaling limit for our sequence of renormalized processes. We will see that this behavior is closely connected to that of the function $\psi^T$. About $\psi^T$, one can first remark that the function defined for $x \geq 0$ by

$$\rho^T(x) = T\frac{\psi^T}{\|\psi^T\|_1}(Tx) \tag{2}$$

is the density of the random variable

$$X^T = \frac{1}{T}\sum_{i=1}^{I^T} X_i,$$

where the $(X_i)$ are iid random variables with density $\phi$ and $I^T$ is a geometric random variable with parameter $1 - a_T$ ($\forall k > 0, \mathbb{P}[I^T = k] = (1 - a_T)(a_T)^{k-1}$). Now let $z \in \mathbb{R}$. The characteristic function of the random variable $X^T$, denoted by $\widehat{\rho}^T$, satisfies

$$\begin{aligned}\widehat{\rho}^T(z) = \mathbb{E}[e^{izX^T}] &= \sum_{k=1}^{\infty}(1 - a_T)(a_T)^{k-1}\mathbb{E}[e^{i\frac{z}{T}\sum_{i=1}^{k} X_i}] \\ &= \sum_{k=1}^{\infty}(1 - a_T)(a_T)^{k-1}(\hat{\phi}(\frac{z}{T}))^k = \frac{\hat{\phi}(\frac{z}{T})}{1 - \frac{a_T}{1-a_T}(\hat{\phi}(\frac{z}{T}) - 1)},\end{aligned}$$

where $\hat{\phi}$ denotes the characteristic function of $X_1$. Since

$$\int_0^{+\infty} s\phi(s)ds = m < \infty,$$

the function $\hat{\phi}$ is continuously differentiable with first derivative at point zero equal to $im$. Therefore, using that $a_T$ and $\hat{\phi}(\frac{z}{T})$ both tend to one as $T$ goes to infinity, $\widehat{\rho}^T(z)$ is equivalent to

$$\frac{1}{1 - \frac{izm}{T(1-a_T)}}.$$

Thus, we precisely see here that the suitable regime so that we get a non trivial limiting law for $X^T$ is that there exists $\lambda > 0$ such that

$$T(1 - a_T) \underset{T\to+\infty}{\to} \lambda. \tag{3}$$

When (3) holds, we write $d_0 = m/\lambda$. In fact we have just proved the following result.

**Proposition 2.2.** *Assume that (3) holds. Under Assumption 1, the sequence of random variable $X^T$ converges in law towards an exponential random variable with parameter $1/d_0$.*

This simple result is of course not new. For example this type of geometric sums of random variable is studied in detail in [31]. Note also that when $X_1$ is exponentially distributed, $X^T$ is also exponentially distributed, even for a fixed $T$.

Assume from now on that (3) holds and set $u_T = T(1-a_T)/\lambda$ (so that $u_T$ goes to one). Proposition 2.2 is particularly important since it gives us the asymptotic behavior of $\psi^T$ in this setting. Indeed, it tells us that

$$\psi^T(Tx) = \rho^T(x)\frac{a_T}{\lambda u_T} \approx \frac{\lambda}{m}e^{-x\frac{\lambda}{m}}\frac{1}{\lambda} = \frac{1}{m}e^{-x\frac{\lambda}{m}}.$$

Let us now come back to the process $C_t^T$, which can be written

$$C_t^T = (1-a_T)\mu + \mu\int_0^t u_T\lambda\psi^T(Ts)ds + \int_0^t \sqrt{\lambda}\psi^T(T(t-s))\sqrt{C_s^T}dB_s^T, \tag{4}$$

with

$$B_t^T = \frac{1}{\sqrt{T}}\sqrt{u_T}\int_0^{tT}\frac{dM_s^T}{\sqrt{\lambda_s^T}}. \tag{5}$$

By studying its quadratic variation, we will show that $B^T$ represents a sequence of martingales which converges to a Brownian motion. So, heuristically replacing $B^T$ by a Brownian motion $B$ and $\psi^T(Tx)$ by $\frac{1}{m}e^{-x\frac{\lambda}{m}}$ in (4), we get

$$C_t^\infty = \mu(1-e^{-t\frac{\lambda}{m}}) + \frac{\sqrt{\lambda}}{m}\int_0^t e^{-(t-s)\frac{\lambda}{m}}\sqrt{C_s^\infty}dB_s.$$

Applying Itô's formula, this gives

$$C_t^\infty = \int_0^t(\mu - C_s^\infty)\frac{\lambda}{m}ds + \frac{\sqrt{\lambda}}{m}\int_0^t\sqrt{C_s^\infty}dB_s,$$

which precisely corresponds to the stochastic differential equation (SDE) satisfied by a CIR process.

Before stating the theorem which makes the preceding heuristic derivation rigorous, we consider an additional assumption.

**Assumption 2.** *There exists $K_\rho > 0$ such that for all $x \geq 0$ and $T > 0$,*

$$|\rho^T(x)| \leq K_\rho.$$

Note that Assumption 2 is in fact not really restrictive. Indeed, from [37] (Page 214, point 5), we get that if $\|\phi\|_\infty < \infty$ and $\int_0^{+\infty}|s|^3\phi(s)ds < +\infty$, then Assumption 2 holds. From [31] (Chapter 5, Lemma 4.1), it also holds if the random variable $X_1$ with density $\phi$ can be written (in law) under the form $X_1 = E + Y$, where $E$ follows an exponential law with parameter $\gamma > 0$ and $Y$ is independent of $E$. We now give our main theorem.

**Theorem 2.2.** *Assume that (3) holds. Under Assumptions 1 and 2, the sequence of renormalized Hawkes intensities $(C_t^T)$ defined in (1) converges in law, for the Skorohod topology, towards the law of the unique strong solution of the following Cox Ingersoll Ross stochastic differential equation on $[0,1]$:*

$$X_t = \int_0^t(\mu - X_s)\frac{\lambda}{m}ds + \frac{\sqrt{\lambda}}{m}\int_0^t\sqrt{X_s}dB_s.$$

*Furthermore, the sequence of renormalized Hawkes process*

$$V_t^T = \frac{1-a_T}{T} N_{tT}^T$$

*converges in law, for the Skorohod topology, towards the process*

$$\int_0^t X_s ds, \ t \in [0,1].$$

## 2.4 Discussion

- Theorem 2.2 implies that when $\|\phi\|_1$ is close to 1, if the observation time $T$ is suitably chosen (that is of order $1/(1-\|\phi\|_1)$), a non degenerate behavior (neither explosive, nor deterministic) can be obtained for a rescaled Hawkes process.

- This can for example be useful for the statistical estimation of the parameters of a Hawkes process. Indeed, designing an estimating procedure based on the fine scale properties of a Hawkes process is a very hard task: Non parametric methods are difficult to use and present various instabilities, see [4, 20], whereas parametric approaches are of course very sensitive to model specifications, see [20, 21]. Considering an intermediate scale, where the process behaves like a CIR model, one can use statistical methods specifically developed in order to estimate CIR parameters, see [3] for a survey. Of course, only the parameters $\lambda$, $m$ and $\mu$ can be recovered this way. Therefore, there is clearly an information loss in this approach. However, it still enables to get access to quantities which are important in practice, see Section 1. In some sense, it can be compared to the extreme value theory based method for extreme quantile estimation, where one assumes that the random variables of an iid sample belong to some max stable attraction domain. Indeed, these two methods lie between a fully parametric one, where a parametric form is assumed (for the law of the random variables or the function $\phi$), and a fully non parametric one, where a functional estimator (of the repartition function or of $\phi$) is used in order to reach the quantity of interest (the quantile or the $L^1$ norm of $\phi$).

- CIR processes are a very classical way to model stochastic (squared) volatilities in finance, see the celebrated Heston model [25]. Also, it is widely acknowledged that there exists a linear relationship between the cumulated order flow and the integrated squared volatility, see for example [40]. Therefore, our setting where $\|\phi\|_1$ is close to one and the limiting behavior obtained in Theorem 2.2 seem in good agreement with market data.

- For the stationary version of a Hawkes process, one can show that the variance of $N_T^T$ is of order $T(1-\|\phi^T\|_1)^{-3}$, see for example [13]. Therefore, if $T(1-a_T)$ tends to zero, that is $\|\phi^T\|_1$ goes rapidly to one, then the variance of $\frac{(1-a_T)}{T} N_T^T$ blows up as $T$ goes to infinity. This situation is therefore very different from the one studied here and out of the scope of this paper.

- The assumption $\int_0^{+\infty} s\phi(s)ds < +\infty$ is crucial in order to approximate $\psi^T$ by an exponential function using Proposition 2.2. Let us now consider the fat tail case where the preceding integral is infinite. More precisely, let us take a function $\phi$ which is of order

$\frac{1}{x^{1+\alpha}}$, $0 < \alpha < 1$, as $x$ goes to infinity. In that case, following the proof of Proposition 2.2, we can show the following result, where we borrow the notations of Proposition 2.2.

**Proposition 2.3.** *Let $E_C^\alpha$ be a random variable whose characteristic function satisfies*

$$\mathbb{E}[e^{izE_C^\alpha}] = \frac{1}{1 - C(iz)^\alpha}.$$

*Assume $\hat{\phi}(z) - 1 \sim_0 \sigma(iz)^\alpha$ for some $\sigma > 0$, $0 < \alpha < 1$, and $(1 - a_T)T^\alpha \to \lambda > 0$. Then $X^T$ converges in law towards the random variable $E_{\frac{\sigma}{\lambda}}^\alpha$.*

Thus, when the shape of the kernel is of order $x^{-(1+\alpha)}$, the "right" observation scale is no longer $T \sim 1/(1 - \|\phi\|_1)$ but $T \sim 1/(1 - \|\phi\|_1)^{\frac{1}{\alpha}}$. Remark also that if we denote by $E_{\alpha,\beta}$ the $(\alpha, \beta)$ Mittag-Leffler function, that is

$$E_{\alpha,\beta}(z) = \sum_{n=1}^{\infty} \frac{z^n}{\Gamma(\alpha n + \beta)},$$

see for example [39], then the density $\phi_C^\alpha$ of $E_C^\alpha$ is linked to this function since

$$\phi_1^\alpha(x) = x^{\alpha-1} E_{\alpha,\alpha}(-x^\alpha).$$

Now let us consider the asymptotic setting where $\mu^T = \mu T^{\alpha-1}$,$\phi^T = a_T \phi$ with $a_T = 1 - \frac{\lambda}{T^\alpha}$ and $\phi$ as in Proposition 2.3. If we apply the same heuristic arguments as those used in Section 2 to the renormalized intensity

$$C_t^T = \frac{\lambda_{tT}^T (1 - a_T)}{T^{\alpha-1}},$$

we get the following type of limiting law for our sequence of Hawkes intensities:

$$X_t = \mu \int_0^t \phi_{\frac{\sigma}{\lambda}}^\alpha (t - s) ds + \int_0^t \phi_{\frac{\sigma}{\lambda}}^\alpha (t - s) \frac{1}{\sqrt{\lambda}} \sqrt{X_s} dB_s.$$

These heuristic arguments are however far from a proof. Indeed, in this case, we probably have to deal with a non semi-martingale limit. Furthermore, tightness properties which are important in the proofs of this paper are much harder to show (in particular the function $\phi_C^\alpha$ is not bounded). We leave this case for further research.

- In the classical time series setting let us mention the paper [8] where the authors study the asymptotic behavior of unstable integer-valued autoregressive model (INAR processes). In this case, CIR processes also appear in the limit. This is in fact not so surprising since INAR processes share some similarities with Hawkes processes. In particular, they can somehow be viewed as Hawkes processes for which the kernel would be a sum of Dirac functions.

## 3 Extension of Theorem 2.2 to a price model

In the previous section, we have studied one-dimensional nearly unstable Hawkes processes. For financial applications, they can for example be used to model the arrival of orders when the number of endogenous orders is much larger than the number of exogenous orders, which seems to be the case in practice, see [19, 21]. In this section, we consider the high frequency price model introduced in [6], which is essentially defined as a difference of two Hawkes processes. Using the same approach as for Theorem 2.2, we investigate the limiting behavior of this model when the stability condition is close to saturation.

### 3.1 A Hawkes based price model

In [6], tick by tick moves of the midprice $(P_t)_{t\geq 0}$ are modeled thanks to a two dimensional Hawkes process in the following way: For $t \geq 0$,

$$P_t = N_t^+ - N_t^-,$$

where $(N^+, N^-)$ is a two dimensional Hawkes process with intensity

$$\begin{pmatrix} \lambda_t^+ \\ \lambda_t^- \end{pmatrix} = \begin{pmatrix} \mu \\ \mu \end{pmatrix} + \int_0^t \begin{pmatrix} \phi_1(t-s) & \phi_2(t-s) \\ \phi_2(t-s) & \phi_1(t-s) \end{pmatrix} \begin{pmatrix} dN_s^+ \\ dN_s^- \end{pmatrix},$$

with $\phi_1$ and $\phi_2$ two non negative measurable functions such that the stability condition

$$\int_0^{+\infty} \phi_1(s)ds + \int_0^{+\infty} \phi_2(s)ds < 1$$

is satisfied.

This model takes into account the discreteness and the negative autocorrelation of prices at the microstructure level. Moreover, it is shown in [5] that when one considers this price at large time scales, the stability condition implies that after suitable renormalization, it converges towards a Brownian motion (with a given volatility).

### 3.2 Scaling limit

In the same spirit as in Section 2, we consider the scaling limit of the Hawkes based price process when the stability condition becomes almost violated. More precisely, following the construction of multivariate Hawkes processes of [5], for every observation interval $[0, T]$, we define the Hawkes process $(N^{T+}, N^{T-})$ with intensity

$$\begin{pmatrix} \lambda_t^{T+} \\ \lambda_t^{T-} \end{pmatrix} = \begin{pmatrix} \mu \\ \mu \end{pmatrix} + \int_0^t \begin{pmatrix} \phi_1^T(t-s) & \phi_2^T(t-s) \\ \phi_2^T(t-s) & \phi_1^T(t-s) \end{pmatrix} \begin{pmatrix} dN_s^{T+} \\ dN_s^{T-} \end{pmatrix},$$

with $\phi_1^T$ and $\phi_2^T$ two non negative measurable functions. Note that in this construction, $N^{T+}$ and $N^{T-}$ do not have common jumps, see [5] for details. We consider the following assumption.

**Assumption 3.** *For $i = 1, 2$ and $t \in \mathbb{R}^+$,*

$$\phi_i^T(t) = a_T \phi_i(t),$$

*where $(a_T)_{T\geq 0}$ is a sequence of positive numbers converging to one such that for all $T$, $a_T < 1$ and $\phi_1$ and $\phi_2$ are two non negative measurable functions such that*

$$\int_0^{+\infty} \phi_1(s) + \phi_2(s) ds = 1 \text{ and } \int_0^{+\infty} s\big(\phi_1(s) + \phi_2(s)\big) ds = m < \infty.$$

*Moreover, for $i = 1, 2$, $\phi_i$ is differentiable with derivative $\phi_i'$ such that $\|\phi_i'\|_\infty < +\infty$ and $\|\phi_i'\|_1 < +\infty$.*

We will also make the following technical assumption.

**Assumption 4.** *Let*

$$\psi_+^T = \sum_{k\geq 1} (a_T(\phi_1 + \phi_2))^{*k} \text{ and } \rho^T(x) = T \frac{\psi_+^T}{\|\psi_+^T\|_1}(Tx).$$

*There exists $K_\rho > 0$ such that for all $x \geq 0$ and $T > 0$,*

$$|\rho^T(x)| \leq K_\rho.$$

We work with the renormalized price process

$$P_t^T = \frac{1}{T}(N_{Tt}^{T+} - N_{Tt}^{T-}). \tag{6}$$

The following theorem states that if we consider the rescaled price process over the right time interval, that is if we take $T$ of order $1/(1 - \|\phi_1\|_1 - \|\phi_1\|_2)$, it asymptotically behaves like a Heston model, see [25].

**Theorem 3.1.** *Let $\phi = \phi_1 - \phi_2$. Assume that (3) holds. Under Assumptions 3 and 4, the sequence of Hawkes based price models $(P_t^T)$ converges in law, for the Skorohod topology, towards a Heston type process $P$ on $[0, 1]$ defined by:*

$$\begin{cases} dC_t = (\frac{2\mu}{\lambda_1} - C_t)\frac{\lambda}{m} dt + \frac{1}{m}\sqrt{C_t} dB_t^1 & C_0 = 0 \\ dP_t = \frac{1}{1-\|\phi\|_1}\sqrt{C_t} dB_t^2 & P_0 = 0, \end{cases}$$

*with $(B^1, B^2)$ a bidimensional Brownian motion.*

# 4 Proofs

We gather in this section the proofs of Theorem 2.1, Proposition 2.1, Theorem 2.2 and Theorem 3.1. In the following, $c$ denotes a constant that may vary from line to line.

## 4.1 Proof of Theorem 2.1

Let $v \in [0, 1]$. From Lemma 4 in [5], we get

$$\mathbb{E}[N_{Tv}^T] = \mu T v + \mu \int_0^{Tv} \psi^T(Tv - s) s ds$$

and

$$N^T_{Tv} - \mathbb{E}[N^T_{Tv}] = M^T_{Tv} + \int_0^{Tv} \psi^T(Tv - s) M^T_s ds.$$

Thus, using that

$$\|\psi^T\|_1 = \frac{\|\phi^T\|_1}{1 - \|\phi^T\|_1},$$

we deduce

$$\frac{1 - \|\phi^T\|_1}{T}(N^T_{Tv} - \mathbb{E}[N^T_{Tv}]) \leq \frac{1 - \|\phi^T\|_1}{T}(1 + \|\psi^T\|_1) \sup_{t\in[0,T]} |M^T_t| \leq \frac{1}{T} \sup_{t\in[0,T]} |M^T_t|.$$

Now recall that $M^T$ is a square integrable martingale with quadratic variation process $N^T$. Thus we can apply Doob's inequality which gives

$$\mathbb{E}[(\sup_{t\in[0,T]} M^T_t)^2] \leq 4 \sup_{t\in[0,T]} \mathbb{E}[(M^T_t)^2] \leq 4\mathbb{E}[N^T_T] \leq 4\mu \frac{T}{1 - \|\phi^T\|_1}.$$

Therefore, we finally obtain

$$\mathbb{E}\Big[\sup_{v\in[0,1]} \Big(\frac{1 - \|\phi^T\|_1}{T}(N^T_{Tv} - \mathbb{E}[N^T_{Tv}])\Big)^2\Big] \leq \frac{4\mu}{T(1 - \|\phi^T\|_1)},$$

which gives the result since $T(1 - \|\phi^T\|_1)$ tends to infinity.

### 4.2 Proof of Proposition 2.1

From the definition of $\lambda^T$, using the fact that $\phi$ is bounded on $[0, t]$, we can write

$$\lambda^T_t = \mu + \int_0^t \phi^T(t - s) dM^T_s + \int_0^t \phi^T(t - s)\lambda^T_s ds.$$

We now recall the following classical lemma, see for example [5] for a proof.

**Lemma 4.1.** *If $f(t) = h(t) + \int_0^t \phi^T(t-s) f(s) ds$ with $h$ a measurable locally bounded function, then*

$$f(t) = h(t) + \int_0^t \psi^T(t - s) h(s) ds.$$

We apply this lemma to the function $h$ defined by

$$h(t) = \mu + \int_0^t \phi^T(t - s) dM^T_s.$$

Thus, we obtain

$$\lambda^T_t = \mu + \int_0^t \phi^T(t - s) dM^T_s + \int_0^t \psi^T(t - s)\big(\mu + \int_0^s \phi^T(s - r) dM^T_r\big) ds. \tag{7}$$

Now remark that using Fubini theorem and the fact that

$$\psi^T * \phi^T = \psi^T - \phi^T,$$

we get

$$\begin{aligned}\int_0^t \psi^T(t-s)\int_0^s \phi^T(s-r)dM_r^T ds &= \int_0^t\int_0^t 1_{r\le s}\psi^T(t-s)\phi^T(s-r)dsdM_r^T\\ &= \int_0^t\int_0^{t-r} \psi^T(t-r-s)\phi^T(s)dsdM_r^T\\ &= \int_0^t \psi^T * \phi^T(t-r)dM_r^T\\ &= \int_0^t \psi^T(t-r)dM_r^T - \int_0^t \phi^T(t-r)dM_r^T.\end{aligned}$$

We conclude the proof rewriting (7) using this last equality.

## 4.3 Proof of Theorem 2.2

Before starting the proof of Theorem 2.2, we give some preliminary lemmas.

### 4.3.1 Preliminary lemmas

We start with some lemmas on $\phi$ and its Fourier transform $\hat{\phi}$ (the associated characteristic function).

**Lemma 4.2.** *Let $\delta > 0$. There exists $\varepsilon > 0$ such that for any real number $z$ with $|z| \ge \delta$,*

$$|1-\hat{\phi}(z)| \ge \varepsilon.$$

*Proof.* Since $\phi$ is bounded, $\hat{\phi}(z)$ tends to zero as $z$ tends to infinity. Consequently, there exists $b > \delta$ such that for all $z$ such that $|z| \ge b$,

$$|\hat{\phi}(z)| \le \frac{1}{2}.$$

Now, let $M$ denote the supremum of the real part of $\hat{\phi}$ on $[-b,-\delta]\cup[\delta,b]$, since $\hat{\phi}$ is continuous this supremum is attained at some point $z_0$. We have $M = \mathrm{Re}(\hat{\phi}(z_0)) = \mathbb{E}[\cos(z_0X)]$, with $X$ a random variable with density $\phi$. Since $\phi$ is continuous, almost surely, $X$ does not belong to $2\pi/z_0\mathbb{Z}$. Thus $M = \mathbb{E}[\cos(z_0X)] < 1$. Therefore, taking $\varepsilon = \min(\frac{1}{2}, 1-M)$ we have the lemma. □

Using that $\|\phi'\|_1 < +\infty$, integrating by parts, we immediately get the following lemma.

**Lemma 4.3.** *Let $z \in \mathbb{R}$. We have $|\hat{\phi}(z)| \le c/|z|$.*

We now turn to the function $\rho^T$ defined in (2). We have the following result.

**Lemma 4.4.** *There exist $c > 0$ such that for all real $z$ and $T \ge 1$,*

$$|\widehat{\rho^T}(z)| \le c(1 \wedge |\frac{1}{z}|).$$

*Proof.* First note that as the Fourier transform of a random variable, $|\widehat{\rho^T}| \leq 1$. Furthermore, using Lemma 4.2 together with the fact that

$$\int_0^{+\infty} x\phi(x)dx = m < +\infty,$$

we get that there exist $\delta > 0$ and $\varepsilon > 0$ such that if $|x| \leq \delta$,

$$|\mathrm{Im}(\hat{\phi})(x)| \geq \frac{m}{2}|x|$$

and if $|x| \geq \delta$,

$$|1 - \hat{\phi}(x)| \geq \varepsilon.$$

Therefore, we deduce that if $|z/T| \leq \delta$,

$$|\widehat{\rho^T}(z)| = |\frac{(1-a_T)\hat{\phi}(\frac{z}{T})}{1 - a_T\hat{\phi}(\frac{z}{T})}| \leq \frac{(1-a_T)}{a_T|\mathrm{Im}(\hat{\phi})(\frac{z}{T})|} \leq \frac{2(1-a_T)T}{a_T m|z|} \leq c/|z|$$

and, thanks to Lemma 4.3, if $|z/T| \geq \delta$

$$|\widehat{\rho^T}(z)| \leq \frac{(1-a_T)|\hat{\phi}(\frac{z}{T})|}{|1 - \hat{\phi}(\frac{z}{T})|} \leq \frac{c(1-a_T)T}{|z|\varepsilon} \leq c/|z|.$$

□

The next lemma gives us the $L^2$ convergence of $\rho^T$.

**Lemma 4.5.** *Let $\rho(x) = \frac{\lambda}{m}e^{-\frac{x\lambda}{m}}$ be the density of the exponential random variable with parameter $\lambda/m$. We have the following convergence, where $|.|_2$ denotes the $L^2$ norm on $\mathbb{R}^+$:*

$$|\rho^T - \rho|_2 \to 0.$$

*Proof.* Using the Fourier isometry, we get

$$|\rho^T - \rho|_2 = \frac{1}{2\pi}|\widehat{\rho^T} - \widehat{\rho}|_2.$$

From Proposition 2.2, for given $z$, we have $(\widehat{\rho^T}(z) - \widehat{\rho}(z)) \to 0$. Thanks to Lemma 4.4, we can apply the dominated convergence theorem which gives that this convergence also takes place in $L^2$. □

We now give a Lipschitz type property for $\rho^T$.

**Lemma 4.6.** *There exists $c > 0$ such that for all $x \geq 0$, $y \geq 0$ and $T \geq 1$,*

$$|\rho^T(x) - \rho^T(y)| \leq cT|x - y|.$$

*Proof.* We simply compute the derivative of $\rho^T$ on $\mathbb{R}_+$, which is given by

$$(\rho^T)'(x) = T\big(\phi'(Tx)\frac{T}{\|\psi^T\|_1} + \phi' * \rho^T(Tx)\big).$$

Using that $\|\psi^T\|_1 = a_T/(1-a_T)$ together with the fact that $T(1-a_T) \to \lambda$, we get

$$|(\rho^T)'(x)| \leq T(c\|\phi'\|_\infty + \|\phi'\|_1\|\rho^T\|_\infty).$$

□

We now consider the function $f^T$ defined for $x \geq 0$ by

$$f^T(x) = \frac{m}{\lambda}\frac{a_T}{u_T}\rho^T(x) - e^{-\frac{x}{d_0}}.$$

We have the following obvious corollaries.

**Corollary 4.1.** *We have*

$$\int |f^T(x)|^2 dx \to 0.$$

**Corollary 4.2.** *There exists $c > 0$ such that for any $z \geq 0$,*

$$|f^T(z)| \leq c.$$

**Corollary 4.3.** *There exists $c > 0$ such that for any $z \geq 0$,*

$$|\widehat{f^T}(z)| \leq c(|\frac{1}{z}| \wedge 1).$$

**Corollary 4.4.** *There exists $c > 0$ such that for all $x \geq 0$, $y \geq 0$ and $T \geq 1$,*

$$|f^T(x) - f^T(y)| \leq cT|x - y|.$$

We finally give a lemma on the integrated difference associated to the function $f^T$.

**Lemma 4.7.** *For any $0 < \varepsilon < 1$, there exists $c_\varepsilon$ so that for all $t, s \geq 0$,*

$$\int_{\mathbb{R}} (f^T(t-u) - f^T(s-u))^2 du \leq c_\varepsilon |t-s|^{1-\varepsilon}.$$

*Proof.* Defining $g^T_{t,s}(u) = f^T(t-u) - f^T(s-u)$, we easily get

$$|\widehat{g^T_{t,s}}(w)| = |e^{-iwt} - e^{-iws}||\widehat{f^T}(w)|.$$

Thus, from Corollary 4.3 together with the fact that

$$|\frac{e^{-iwt} - e^{-iws}}{w(t-s)}| \leq 1,$$

we get

$$\begin{aligned}
\int_{\mathbb{R}} (f^T(t-u) - f^T(s-u))^2 du &\leq c \int_{\mathbb{R}} |\widehat{g^T_{t,s}}(w)|^2 dw \\
&\leq c \int_{\mathbb{R}} |e^{-iwt} - e^{-iws}|^2 (|\frac{1}{w^2}| \wedge 1) dw \\
&\leq c \int_{\mathbb{R}} 2^{1+\varepsilon} |\frac{e^{-iwt} - e^{-iws}}{w(t-s)}|^{1-\varepsilon} (|\frac{1}{w^2}| \wedge 1) w^{1-\varepsilon} dw |t-s|^{1-\varepsilon} \\
&\leq c_\varepsilon |t-s|^{1-\varepsilon}.
\end{aligned}$$

□

#### 4.3.2 Proof of the first part of Theorem 2.2

We now begin with the proof of the first assertion in Theorem 2.2. We split this proof into several steps.

**Step 1: Convenient rewriting of $C^T$**

In this step, our goal is to obtain a suitable expression for $C_t^T$. Let $d_0 = m/\lambda$. Inspired by the limiting behavior of $\psi^T$ given in Proposition 2.2, we write Equation (4) under the form

$$C_t^T = R_t^T + \mu(1 - e^{-\frac{t}{d_0}}) + \frac{\sqrt{\lambda}}{m} \int_0^t e^{-\frac{t-s}{d_0}} \sqrt{C_s^T} dB_s^T,$$

where $R_t^T$ is obviously defined. Using integration by parts (for finite variation processes), we get

$$C_t^T = R_t^T + \frac{\mu}{d_0} \int_0^t e^{-\frac{v}{d_0}} dv + \frac{\sqrt{\lambda}}{m} \int_0^t \sqrt{C_v^T} dB_v^T - \frac{\sqrt{\lambda}}{md_0} \int_0^t \Big( \int_0^v e^{-\frac{v-s}{d_0}} \sqrt{C_s^T} dB_s^T \Big) dv.$$

Then remarking that

$$\frac{\sqrt{\lambda}}{md_0} \int_0^v e^{-\frac{v-s}{d_0}} \sqrt{C_s^T} dB_s^T = \frac{1}{d_0} \big( C_v^T - R_v^T - \mu(1 - e^{-\frac{v}{d_0}}) \big),$$

we finally derive

$$C_t^T = U_t^T + \frac{1}{d_0} \int_0^t (\mu - C_s^T) ds + \frac{\sqrt{\lambda}}{m} \int_0^t \sqrt{C_s^T} dB_s^T, \tag{8}$$

with

$$U_t^T = R_t^T + \frac{1}{d_0} \int_0^t R_s^T ds.$$

The form (8) will be quite convenient in order to study the asymptotic behavior of $C_t^T$. Indeed, we will show that $U_t^T$ vanishes so that (8) almost represents a stochastic differential equation.

**Step 2: Preliminaries for the convergence of $U^T$**

We now want to prove that the sequence of processes $(U_t^T)_{t\in[0,1]}$ converges to zero in law, for the Skorohod topology, and therefore uniformly on compact sets on $[0,1]$ (ucp). We show here that to do so, it is enough to study a (slightly) simpler process than $U^T$. First, it is clear that showing the convergence of $(R_t^T)_{t\in[0,1]}$ to zero gives also the convergence of $U^T$. Now recall that

$$R_t^T = \mu(1-a_T) - \mu((1-e^{-\frac{t}{d_0}}) - \int_0^t a_T T \frac{\psi^T}{\|\psi^T\|_1}(Ts) ds) + \sqrt{\lambda} \int_0^t \big( \psi^T(T(t-s)) - \frac{1}{m} e^{-\frac{t-s}{d_0}} \big) \sqrt{C_t^T} dB_s^T.$$

Since $a_T$ tends to one, the first term tends to zero. For $t \in [0,1]$, Proposition 2.2 gives us the convergence of

$$\int_0^t a_T T \frac{\psi^T}{\|\psi^T\|_1}(Ts) ds$$

towards $1 - e^{-\frac{t}{d_0}}$. Using Dini's theorem, we get that this convergence is in fact uniform over $[0,1]$. Thus, using Equation (5), we see that it remains to show that $(Y_t^T)_{t\in[0,1]}$ goes to zero, with

$$Y_t^T = \int_0^t (m\psi^T(T(t-u)) - e^{-\frac{t-u}{d_0}})d\overline{M}_t^T,$$

where $\overline{M}_t^T = M_{tT}^T/T$.

**Step 3: Finite dimensional convergence of $Y^T$**

We now show the finite dimensional convergence of $(Y_t^T)_{t\in[0,1]}$.

**Lemma 4.8.** *For any $(t_1, ..., t_n) \in [0,1]^n$, we have the following convergence in law:*

$$(Y_{t_1}^T, ..., Y_{t_n}^T) \to 0.$$

*Proof.* First note that the quadratic variation of $\overline{M}^T$ at time $t$ is given by $N_{tT}^T/T^2$, whose predicable compensator process at time $t$ is simply equal to

$$\frac{1}{T^2}\int_0^{tT} \lambda_s^T ds.$$

Using this together with the fact that

$$\mathbb{E}[\lambda_t^T] = \mu + \mu\int_0^t \psi^T(t-s)ds \leq \mu + \mu\frac{a_T}{1-a_T} \leq cT,$$

we get

$$\mathbb{E}[(Y_t^T)^2] \leq c\int_0^t (m\psi^T(T(t-s)) - e^{-\frac{t-s}{d_0}})^2 ds.$$

Now remark that

$$m\psi^T(T(t-s)) - e^{-\frac{t-s}{d_0}} = f^T(t-s),$$

where $f^T$ is defined by $f^T(x) = 0$ for $x < 0$ and

$$f^T(x) = \frac{m}{\lambda}\frac{a_T}{u_T}\rho^T(x) - e^{-\frac{x}{d_0}}$$

for $x \geq 0$, with $\rho^T$ the function introduced in Equation (2). From Corollary 4.1,

$$\mathbb{E}[(Y_t^T)^2] \to 0,$$

which gives the result. □

**Step 4: A Kolmogorov type inequality for $Y^T$**

To prove the convergence of $Y^T$ towards 0, it remains to show its tightness. We have the following Kolmogorov type inequality on the moments of the increments of $Y^T$, which is a first step in order to get the tightness.

**Lemma 4.9.** *For any $\varepsilon > 0$, there exists $c_\varepsilon > 0$ such that for all $T \geq 1$, $0 \leq t, s \leq 1$,*

$$\mathbb{E}[(Y_t^T - Y_s^T)^4] \leq c_\varepsilon(|t-s|^{3/2-\varepsilon} + \frac{1}{T^2}|t-s|^{1-\varepsilon}). \tag{9}$$

*Proof.* Let $\mu^{\mathbb{E}[M_4^T]}$ denote the fourth moment measure of $M^T$, see Appendix for definition and properties. We have

$$\mathbb{E}[(Y_t^T - Y_s^T)^4] = \frac{1}{T^4}\int_{[0,T]^4}\Big(\prod_{i=1}^{4}\big[f^T(t-\frac{t_i}{T}) - f^T(s-\frac{t_i}{T})\big]\Big)\mu^{\mathbb{E}[M_4^T]}(dt_1,dt_2,dt_3,dt_4).$$

Therefore, using Lemma A.17, we obtain

$$\begin{aligned}\mathbb{E}[(Y_t^T - Y_s^T)^4] \leq{}& \frac{c}{T^3}\int_0^T |f^T(t-\frac{u}{T}) - f^T(s-\frac{u}{T})|^4 du\\ &+ \frac{c}{T^3}\int_0^T |f^T(t-\frac{u}{T}) - f^T(s-\frac{u}{T})|^3 du \int_0^T |f^T(t-\frac{u}{T}) - f^T(s-\frac{u}{T})| du\\ &+ \frac{c}{T^2}\int_0^T |f^T(t-\frac{u}{T}) - f^T(s-\frac{u}{T})|^2 du \int_0^T |f^T(t-\frac{u}{T}) - f^T(s-\frac{u}{T})|^2 du\\ &+ \frac{c}{T^3}\Big(\int_0^T |f^T(t-\frac{u}{T}) - f^T(s-\frac{u}{T})| du\Big)^2 \int_0^T |f^T(t-\frac{u}{T}) - f^T(s-\frac{u}{T})|^2 du.\end{aligned}$$

Then, using Cauchy Schwarz inequality together with Corollary 4.2 and Lemma 4.7, we get

$$\int_0^T |f^T(t-\frac{u}{T}) - f^T(s-\frac{u}{T})| du \leq c_\varepsilon T\sqrt{|t-s|^{1-\varepsilon}}$$

and for $p = 2, 3, 4$,

$$\int_0^T |f^T(t-\frac{u}{T}) - f^T(s-\frac{u}{T})|^p du \leq c_\varepsilon T|t-s|^{1-\varepsilon},$$

which enables to conclude the proof. □

**Step 5: Tightness**

Let us define $\tilde{Y}^T$ the linear interpolation of $Y^T$ with mesh $1/T^4$:

$$\tilde{Y}_t^T = Y^T_{\frac{\lfloor tT^4\rfloor}{T^4}} + (tT^4 - \lfloor tT^4\rfloor)(Y^T_{\frac{\lfloor tT^4\rfloor+1}{T^4}} - Y^T_{\frac{\lfloor tT^4\rfloor}{T^4}}).$$

We use this interpolation since for $t - s = 1/T^4$, both terms on the right hand side of (9) have the same order of magnitude and for $t - s > 1/T^4$ the second term becomes negligible. We have the following lemma.

**Lemma 4.10.** *The sequence* $(\tilde{Y}^T)$ *is tight.*

*Proof.* We want to apply the classical Kolmogorov tightness criterion, see [10], that states that if there exist $\gamma > 1$ and $c > 0$ such that for any $0 \leq s \leq t \leq 1$,

$$\mathbb{E}|\tilde{Y}_t^T - \tilde{Y}_s^T|^4 \leq c|t-s|^\gamma,$$

then $\tilde{Y}^T$ is tight. Remark that such inequality can of course not hold for $Y^T$ since it is not continuous. Let $n_t^T = \lfloor tT^4\rfloor$ and $n_s^T = \lfloor sT^4\rfloor$. Let $0 < \varepsilon, \varepsilon' \leq 1/4$ and $T \geq 1$. There are three cases:

• If $n_t^T = n_s^T$, using Lemma 4.9, we obtain that

$$\mathbb{E}[(\tilde{Y}_t^T - \tilde{Y}_s^T)^4]$$

is smaller than

$$|t-s|^4 T^{16} \mathbb{E}[(Y_{\frac{n_t^T+1}{T^4}} - Y_{\frac{n_t^T}{T^4}})^4] \leq c_\varepsilon \frac{1}{T^{4(3/2-\varepsilon)}} T^{16} |t-s|^4 \leq c_\varepsilon \frac{1}{T^{4(3/2-\varepsilon)}} T^{16} |t-s|^{1+\varepsilon'} \frac{1}{T^{4(3-\varepsilon')}}.$$

Since $0 < \varepsilon, \varepsilon' \leq 1/4$, this leads to

$$\mathbb{E}[(\tilde{Y}_t^T - \tilde{Y}_s^T)^4] \leq c_\varepsilon |t-s|^{1+\varepsilon'}.$$

• If $n_t^T = n_s^T + 1$,

$$\mathbb{E}[(\tilde{Y}_t^T - \tilde{Y}_s^T)^4] \leq c\mathbb{E}[(\tilde{Y}_t^T - \tilde{Y}^T_{\frac{n_t^T}{T^4}})^4] + c\mathbb{E}[(\tilde{Y}^T_{\frac{n_t^T}{T^4}} - \tilde{Y}_s^T)^4] \leq c_\varepsilon |t-s|^{1+\varepsilon'}.$$

• If $n_t^T \geq n_s^T + 2$, using again Lemma 4.9, we get

$$\begin{aligned}\mathbb{E}[(\tilde{Y}_t^T - \tilde{Y}_s^T)^4] &\leq c\mathbb{E}[(\tilde{Y}_t^T - \tilde{Y}^T_{\frac{n_t^T}{T^4}})^4] + c\mathbb{E}[(\tilde{Y}^T_{\frac{n_s^T+1}{T^4}} - \tilde{Y}_s^T)^4] + c\mathbb{E}[(\tilde{Y}^T_{\frac{n_t^T}{T^4}} - \tilde{Y}^T_{\frac{n_s^T+1}{T^4}})^4] \\ &\leq c_\varepsilon (\frac{1}{T^4})^{1+\varepsilon'} + c_\varepsilon |\frac{n_t^T}{T^4} - \frac{n_s^T+1}{T^4}|^{\frac{3}{2}-\varepsilon} \leq c_\varepsilon |t-s|^{\min(\frac{3}{2}-\varepsilon, 1+\varepsilon')}.\end{aligned}$$

Hence the Kolmogorov criterion holds, which implies the tightness of $\tilde{Y}^T$. □

We now show that the difference between $Y^T$ and $\tilde{Y}^T$ tends uniformly to zero.

**Lemma 4.11.** *We have the following convergence in probability:*

$$\sup_{|t-s| \leq \frac{1}{T^4}} |Y_t^T - Y_s^T| \to 0.$$

*Proof.* Recall that for $0 \leq s \leq t \leq 1$,

$$|Y_t^T - Y_s^T| = |\int_0^s f^T(t-u) - f^T(s-u) d\overline{M}_u^T + \int_s^t f^T(t-u) d\overline{M}_u^T|.$$

Thus, we have that $|Y_t^T - Y_s^T|$ is smaller than

$$\int_0^{sT} |f^T(t-u/T) - f^T(s-u/T)| (dN_u^T + \lambda_u du) \frac{1}{T} + \int_{sT}^{tT} |f^T(t-u/T)| (dN_u^T + \lambda_u du) \frac{1}{T}.$$

Using Corollaries 4.2 and 4.4, we obtain

$$|Y_t^T - Y_s^T| \leq c|t-s| (N_T^T + \int_0^T \lambda_u^T du) + c(N_{tT}^T - N_{sT}^T + \int_{sT}^{tT} \lambda_u^T du) \frac{1}{T}.$$

Consider now

$$\sup_{|t-s| \leq 1/T^4} |Y_t^T - Y_s^T|.$$

This is smaller than

$$c\frac{1}{T^4}(N_T^T + \int_0^T \lambda_u^T du) + 2c \max_{i=0,\dots,\lfloor T^4 \rfloor} \frac{1}{T}(N^T_{\frac{i+1}{T^4}T} - N^T_{\frac{i}{T^4}T} + \int_{\frac{i}{T^4}T}^{\frac{i+1}{T^4}T} \lambda_u^T du). \tag{10}$$

From Lemma A.5, we have

$$\mathbb{E}\big[N_T^T + \int_0^T \lambda_u^T du\big] \leq cT^2.$$

Thus, the first term on the right hand side of (10) tends to zero. For the second term, we use Lemma A.15 (with $t = \frac{i+1}{T^4}T$ and $s = \frac{i}{T^4}T$) which gives that

$$\mathbb{E}\big[\big(\frac{1}{T}(N^T_{\frac{i+1}{T^4}T} - N^T_{\frac{i}{T^4}T} + \int_{\frac{i}{T^4}T}^{\frac{i+1}{T^4}T} \lambda_u^T du)\big)^3\big] \leq \frac{c}{T^5}.$$

So, for any $\varepsilon > 0$, using Markov inequality, we get

$$\mathbb{P}\big[\frac{1}{T}(N^T_{\frac{i+1}{T^4}T} - N^T_{\frac{i}{T^4}T} + \int_{\frac{i}{T^4}T}^{\frac{i+1}{T^4}T} \lambda_u^T du) \geq \varepsilon\big] \leq \frac{c}{T^5\varepsilon^3}.$$

From this inequality, since the maximum is taken over a number of terms of order $T^4$, we easily deduce that the second term on the right hand side of (10) tends to zero in probability. □

We end this step by the proposition stating the convergence of $Y^T$.

**Proposition 4.1.** *The process $Y^T$ converges ucp to 0 on $[0,1]$.*

*Proof.* We have

$$\sup_{t\in[0,1]} |Y_t^T| \leq \sup_{t\in[0,1]} |\tilde{Y}_t^T| + \sup_{t\in[0,1]} |\tilde{Y}_t^T - Y_t^T|.$$

From Lemma 4.8 and Lemma 4.10 we get that $\tilde{Y}^T$ tends to zero, in law for the Skorohod topology. This implies the ucp convergence. Applying Lemma 4.11 we get the result.

□

### Step 6: Limit of a sequence of SDEs

In this last step, we show the convergence of the process $(C_t^T)_{t\in[0,1]}$ towards a CIR process. To do so, we use the fact that $C^T$ can almost be written under the form of a stochastic differential equation. Indeed, recall that

$$C_t^T = U_t^T + \frac{1}{d_0}\int_0^t (\mu - C_s^T)ds + \frac{\sqrt{\lambda}}{m}\int_0^t \sqrt{C_s^T} dB_s^T,$$

with

$$B_t^T = \frac{1}{\sqrt{T}}\sqrt{u_T}\int_0^{tT} \frac{dM_s^T}{\sqrt{\lambda_s^T}}.$$

Then we aim at applying Theorem 5.4 in [32] to $C^T$. This result essentially says that for a sequence of SDEs where the functions and processes defining the equations satisfy some convergence properties, the laws of the solutions of the SDEs converge to the law of the solution

of the limiting SDE. We now check these convergence properties.

The sequence of processes $(B^T)$ is a sequence of martingales with jumps uniformly bounded by $c/\sqrt{\mu}$. Furthermore, for $t \in [0,1]$, the quadratic variation of $(B^T)$ at point $t$ is equal to

$$\frac{u_T}{T}\int_0^{tT}\frac{dN_s^T}{\lambda_s^T} = u_T\big(t + \int_0^{tT}\frac{dM_s^T}{T\lambda_s^T}\big).$$

Now, remark that

$$\mathbb{E}\big[\big(\int_0^{tT}\frac{dM_s^T}{T\lambda_s^T}\big)^2\big] \leq \mathbb{E}\big[\int_0^T \frac{1}{T^2\lambda_s^T}ds\big] \leq c/(T\mu).$$

Therefore, we get that for any $t \in [0,1]$, the quadratic variation of $(B^T)$ at point $t$ converges in probability to $t$. Thus, we can apply Theorem VIII.3.11 in [28] to deduce that $(B_t^T)_{t\in[0,1]}$ converges in law for the Skorohod topology towards a Brownian motion.

Since $U^T$ converges to a deterministic limit, we get the convergence in law, for the product topology, of the couple $(U_t^T, B_t^T)_{\in[0,1]}$ to $(0, B_t)_{\in[0,1]}$, with $B$ a Brownian motion. The components of $(0, B_t)$ being continuous, the last convergence also takes place for the Skorohod topology on the product space.

Finally, recall that the (CIR) stochastic differential equation

$$X_t = \int_0^t (\mu - X_s)\frac{1}{d_0}ds + \frac{\sqrt{\lambda}}{m}\int_0^t \sqrt{X_s}dB_s$$

admits a unique strong solution on $[0,1]$. This, together with the preceding elements enables us to readily apply Theorem 5.4 in [32] to the sequence $C^T$, which gives the result.

#### 4.3.3 Proof of the second part of Theorem 2.2

We now give the proof of the second part of Theorem 2.2 which deals with the sequence of Hawkes processes $N^T$. Let

$$V_t^T = \frac{(1-a_T)}{T}N_{tT}^T.$$

We write

$$V_t^T = \int_0^t C_s^T ds + \hat{M}_t^T,$$

where

$$\hat{M}_t^T = \frac{(1-a_T)}{T}(N_{tT}^T - \int_0^{tT}\lambda_s^T ds)$$

is a martingale. Using Doob's inequality, we obtain

$$\mathbb{E}[(\sup_{t\in[0,1]}\hat{M}_t^T)^2] \leq 4\mathbb{E}[(\hat{M}_1^T)^2] \leq 4(\frac{(1-a_T)}{T})^2\mathbb{E}[N_T^T] \leq \frac{4\mu(1-a_T)}{T} \to 0.$$

Moreover, $(C^T, t)$ converges in law over $[0,1]$ to $(C, t)$ for the Skorokod topology. This last remark and Theorem 2.6 in [30] on the limit of sequences of stochastic integrals give the result.

### 4.4 Proof of Theorem 3.1

We first introduce some notations. In this proof, we write

$$\phi^T = \phi_1^T - \phi_2^T \text{ and } \psi^T = \sum_{k=1}^{+\infty} (\phi^T)^{*k}.$$

Moreover, we set

$$C_t^T = \frac{\lambda_{tT}^{T+} + \lambda_{tT}^{T-}}{T}$$

and define

$$(B^1)_t^T = \int_0^{tT} \frac{dM_s^{T+} + dM_s^{T-}}{\sqrt{T(\lambda_s^{T+} + \lambda_s^{T-})}}, \quad (B^2)_t^T = \int_0^{tT} \frac{dM_s^{T+} - dM_s^{T-}}{\sqrt{T(\lambda_s^{T+} + \lambda_s^{T-})}},$$

with

$$M_s^{T+} = N_s^{T+} - \int_0^s \lambda_s^{T+} ds, \quad M_s^{T-} = N_s^{T-} - \int_0^s \lambda_s^{T-} ds.$$

Finally, we set

$$\overline{M}_t^{T+} = \frac{M_{Tt}^{T+}}{T}, \quad \overline{M}_t^{T-} = \frac{M_{Tt}^{T-}}{T}.$$

We split the proof of Theorem 3.1 into several steps.

#### Step 1: Convenient rewriting

In this first step, we rewrite the price, intensity and martingale processes under more convenient forms. We have

$$\lambda_t^{T+} - \lambda_t^{T-} = \int_0^t \phi^T(t-s)(\lambda_s^{T+} - \lambda_s^{T-}) ds + \int_0^t \phi^T(t-s)(dM_s^{T+} - dM_s^{T-}).$$

Therefore, in the same way as for the proof of Proposition 2.1, we get

$$\lambda_t^{T+} - \lambda_t^{T-} = \int_0^t \psi^T(t-s)(dM_s^{T+} - dM_s^{T-}).$$

From this last expression, we easily obtain

$$N_t^{T+} - N_t^{T-} = \int_0^t \big(1 + \Psi^T(t-u)\big)(dM_u^{T+} - dM_u^{T-}), \tag{11}$$

with

$$\Psi^T(x) = \int_0^x \psi^T(s) ds.$$

Finally, note that

$$\overline{M}_t^{T+} - \overline{M}_t^{T-} = \frac{1}{T}(M_{Tt}^{T+} - M_{Tt}^{T-}) = \int_0^t \sqrt{C_s^T} d(B^2)_s^T. \tag{12}$$

**Step 2: Preliminary result**

For $s \in [0,1]$, we define

$$X_s^T = \frac{\lambda_{sT}^{T+} - \lambda_{sT}^{T-}}{T}.$$

We have the following important result.

**Lemma 4.12.** *The process $X^T$ converges ucp to 0 on $[0,1]$.*

*Proof.* We write

$$X_t^T = \int_0^t f_1^T(t-s) d(\overline{M}_s^{T+} - \overline{M}_s^{T-}),$$

with $f_1^T(x) = \psi^T(Tx)$. Remark that Corollaries 4.1, 4.2, 4.3 and 4.4 are valid if in their statement, $f^T$ is replaced by $f_1^T$. In the proof of Theorem 2.2, we have shown the convergence to zero of the process

$$Y_t^T = \int_0^T f^T(t-s) d\overline{M}_s^T.$$

Therefore, applying the same strategy but replacing $f^T$ by $f_1^T$ and $\overline{M}^T$ by $\overline{M}^{T+} - \overline{M}^{T-}$, it is clear that we get the result. □

**Step 3: Convergence of $(B^1, B^2)$**

In this step, we prove the convergence of $(B^1, B^2)$ towards a two-dimensional Brownian motion. To do so, we study the quadratic (co-)variations of the processes. Let $i \in \{1,2\}$, $j \in \{1,2\}$. We denote by $[(B^i)^T, (B^j)^T]_t$ the quadratic co-variation of $B^i$ and $B^j$ at time $t$.

**Lemma 4.13.** *We have the following convergence in probability:*

$$[(B^i)^T, (B^j)^T]_t \to t 1_{i=j}.$$

*Proof.* There are three cases:

- If $i = j = 1$, using that $N^{T+}$ and $N^{T-}$ have no common jumps, we get,

$$[(B^1)^T, (B^1)^T]_t = \int_0^{tT} \frac{dN_s^{T+} + dN_s^{T-}}{T(\lambda_s^{T+} + \lambda_s^{T-})} = t + \int_0^{tT} \frac{dM_s^{T+} + dM_s^{T-}}{T(\lambda_s^{T+} + \lambda_s^{T-})}.$$

Furthermore,

$$\mathbb{E}\Big[\Big(\int_0^{tT} \frac{dM_s^{T+} + dM_s^{T-}}{T(\lambda_s^{T+} + \lambda_s^{T-})}\Big)^2\Big] \leq \frac{ct}{T\mu} \to 0.$$

Therefore we have the result for $i = j = 1$.

- If $i = j = 2$, the proof goes similarly.

- If $i = 1$ and $j = 2$,

$$[(B^1)^T, (B^2)^T]_t = \int_0^{tT} \frac{dN_s^{T+} - dN_s^{T-}}{T(\lambda_s^{T+} + \lambda_s^{T-})} = \int_0^{tT} \frac{dM_s^{T+} - dM_s^{T-} + \lambda_s^{T+} ds - \lambda_s^{T-} ds}{T(\lambda_s^{T+} + \lambda_s^{T-})}.$$

As for the case $i = j = 1$, we easily get

$$\int_0^{tT} \frac{dM_s^{T+} - dM_s^{T-}}{T(\lambda_s^{T+} + \lambda_s^{T-})} \to 0.$$

It remains to show the convergence to zero of $Z_t^T$ defined by

$$Z_t^T = \int_0^t \frac{X_s^T}{C_s^T} ds.$$

For any $\varepsilon > 0$, we have

$$|Z_t^T| \leq \int_0^t (1 \wedge |\frac{X_s^T}{\varepsilon}|) ds + \int_0^t 1_{C_s^T < \varepsilon} ds.$$

From Lemma 4.12, we have the convergence of the process $X^T$ to zero. Furthermore, in Lemma 4.15 we will show that $C^T$ converge in law over $[0, 1]$ towards a CIR process denoted by $C$. Therefore, since the limiting processes are continuous, we have the joint convergence of $(X_T, C^T)$ to $(0, C)$. We now use Skorohod representation theorem (without changing notations). Almost surely, for $T$ large enough, we have

$$\sup_{s \in [0,1]} |X_s^T| \leq \varepsilon^2, \quad \sup_{s \in [0,1]} |C_s^T - C_s| \leq \varepsilon.$$

This implies

$$\int_0^t (1 \wedge |\frac{X_s^T}{\varepsilon}|) ds + \int_0^t 1_{C_s^T < \varepsilon} ds \leq \varepsilon + \int_0^1 1_{C_s < 2\varepsilon} ds.$$

Recall that the set of zeros of a CIR process on a finite time interval has zero Lebesgue measure. Thus, using the dominated convergence theorem, we easily see that choosing $\varepsilon$ conveniently, the second term in the preceding inequality can be made arbitrarily small, which ends the proof. □

Thus for any $T$, $(B^1)^T$ and $(B^2)^T$ are two martingales with uniformly bounded jumps and their quadratic (co-)variations satisfy Lemma 4.13. Consequently, Theorem VIII.3.11 of [28] gives us the following lemma.

**Lemma 4.14.** *We have*

$$((B^1)^T, (B^2)^T) \to (B^1, B^2),$$

*in law, for the Skorohod topology, where $(B^1, B^2)$ is a two-dimensional Brownian motion.*

**Step 4: Convergence of $(C^T, (B^2)^T)$**

The aim of this step is to prove that the couple $(C^T, (B^2)^T)$ converges in law towards $(C, (B^2))$, with $C$ a CIR process and $B^2$ a Brownian motion, independent of $C$. More precisely, we have the following lemma.

**Lemma 4.15.** *The couple of process $(C^T, (B^2)^T)$ converges in law, for the Skorohod topology, over $[0, 1]$, towards $(C, B^2)$, where $B^2$ is a Brownian motion independent of $C$ and $C$ is a CIR process satisfying*

$$C_t = \int_0^t (\frac{2\mu}{\lambda} - C_s) \frac{\lambda}{m} ds + \frac{1}{m} \int_0^t \sqrt{C_s} dW_s,$$

*with $W$ another Brownian motion, independent of $B^2$.*

*Proof.* Let us consider the process $N^T = N^{T+} + N^{T-}$. It is a point process with intensity

$$\lambda_t^T = \lambda_t^{T+} + \lambda_t^{T-} = 2\mu + a_T \int_0^t (\phi_1 + \phi_2)(t-s) dN_s^T.$$

Therefore, we are in the framework of Theorem 2.2: $N^T$ is a Hawkes process whose kernel has a norm that tends to 1 at the right speed and its renormalized intensity $C^T$ converges towards a CIR. Remark that the renormalizing factor here is $1/T$ and not $(1 - a_T)$, which is not an issue since (3) holds. Thus we get the convergence of $C^T$ towards a CIR. To obtain the joint convergence, we just need to write the same proof as for Theorem 2.2 (up to obvious changes), but using this time Theorem 5.4 in [32] together with Lemma 4.14. □

**Step 5: Technical results**

This fifth step consists in proving two technical results. The first one is the following.

**Lemma 4.16.** *The process*

$$R_t^T = \int_0^t \int_{T(t-u)}^{+\infty} \psi^T(s) ds d(\overline{M}_u^{T+} - \overline{M}_u^{T-})$$

*converges ucp to* 0 *on* $[0, 1]$.

*Proof.* We write

$$R_t^T = \int_0^t f_2^T(t-u) d(\overline{M}_u^{T+} - \overline{M}_u^{T-}),$$

with

$$f_2^T(x) = \int_{Tx}^{+\infty} \psi^T(s) ds.$$

The result follows in the same way as in the proof of Lemma 4.12. □

We now give the last lemma of this step.

**Lemma 4.17.** *We have*

$$\int_0^\infty \int_x^\infty \phi_i(s) ds dx < \infty.$$

*Proof.* Using integration by parts together with Assumption 4, we get

$$\int_0^\infty \int_x^\infty \phi_i(s) ds dx = \int_0^\infty x\phi_i(x) dx + \lim_{x\to\infty} x \int_x^\infty \phi_i(s) ds \leq 2m.$$

□

**Step 6: End of the proof**

We finally show Theorem 3.1 in this step. Using (11) we write

$$\begin{aligned} P_t^T &= (1 + \frac{\|\phi\|_1}{1 - \|\phi\|_1})(\overline{M}_t^{T+} - \overline{M}_t^{T-}) \\ &\quad - \int_0^t \int_{T(t-u)}^{+\infty} \psi^T(s) ds d(\overline{M}_u^{T+} - \overline{M}_u^{T-}) - (\frac{\|\phi\|_1}{1 - \|\phi\|_1} - \frac{a_T \|\phi\|_1}{1 - a^T \|\phi\|_1})(\overline{M}_t^{T+} - \overline{M}_t^{T-}). \end{aligned}$$

Using Theorem 2.6 in [30] together with Lemma 4.15 and Equation (12), we get the convergence of the process $\overline{M}^{T+} - \overline{M}^{T-}$, over $[0,1]$, for the Skorohod topology, towards

$$\int_0^t \sqrt{C_s} dB_s^2.$$

Moreover, in Lemma 4.16, we have shown that the second term in the decomposition of $P_t^T$ tends to zero. Finally, the third term also vanishes since $\|\phi\|_1 < 1$. This concludes the proof.

# A Appendix: Lemmas on the moment measures of Hawkes processes

We give in this appendix some useful formulas related to the moment measures of Hawkes processes. As previously, Hawkes processes are denoted by $N$ (or $N^T$) and we keep the same notations for the associated quantities $M$ and $\lambda$ (or $M^T$ and $\lambda^T$). We also introduce the cumulated intensities:

$$\Lambda_t = \int_0^t \lambda_t(s)ds, \quad \Lambda_t^T = \int_0^t \lambda^T(s)ds.$$

## A.1 Expectation of product measures

Let us first recall the standard definitions of product measures and their expectations. For a random, real valued, increasing process $X$ on $[0,T]$, we denote by $\mu^{X_1}$ its associated random Stieltjes measure on $[0,T]$ and for $k \in \mathbb{N}^*$, we write $\mu^{X_k}$ the product measure on $[0,T]^k$ built from $\mu^{X_1}$. More generaly, when $X^1, ..., X^n$ are $n$ increasing processes and $k^1, ..., k^n$ are $n$ integers such that $k^1 + ... + k^n = k$, we denote by $\mu^{X^1_{k^1} \otimes ... \otimes X^n_{k^n}}$ the product measure on $[0,T]^k$ built from $\mu^{X^1_{k^1}}$,..., $\mu^{X^n_{k^n}}$.

If $X^1, ..., X^n$ are now random, real valued, finite variation processes, we can write them as the difference of two random, real valued, increasing processes: $X^i = (X^i)^+ - (X^i)^-$, $(X^i)^+$ being the total variation of the process and $(X^i)^- = (X^i)^+ - X^i$. We have the following definition.

**Definition A.1.** *The product measures $\mu^{X_k}$ and $\mu^{X^1_{k^1} \otimes ... \otimes X^n_{k^n}}$ are defined as the signed sums*

$$\mu^{X_k} = \sum_{I \in \{-,+\}^k} (-1)^{\#\{i; I_i = -\}} \mu^{X_1^{I_1} \otimes ... \otimes X_1^{I_k}},$$

*and*

$$\mu^{X^1_{k^1} \otimes ... \otimes X^n_{k^n}} = \sum_{I \in \{-,+\}^k} (-1)^{\#\{i; I_i = -\}} \mu^{(X^1)_1^{I_1} \otimes ... \otimes (X^1)_1^{I_{k_1}} \otimes ... \otimes (X^n)_1^{I_k}}.$$

We now define the expectation measure.

**Definition A.2.** *Let $\mu$ be a signed random measure on $(\Omega, \Sigma)$ which writes $\mu = \mu^+ - \mu^-$ with $\mu^+$ and $\mu^-$ two finite positive random measures such that their expectation measures $\mathbb{E}[\mu^+]$ and $\mathbb{E}[\mu^-]$ (defined as $A \mapsto \mathbb{E}[\mu^+(A)]$ and $A \mapsto \mathbb{E}[\mu^-(A)]$, see [16]) are both finite. We define its expectation measure $\mathbb{E}[\mu]$ as*

$$\forall A \in \Sigma, \quad \mathbb{E}[\mu](A) = \mathbb{E}[\mu^+(A)] - \mathbb{E}[\mu^-(A)],$$

**Remark A.1.** *The latter definition does not depend on the pair $(\mu^+, \mu^-)$.*

**Remark A.2.** *From Definition A.1, we have the decomposition*

$$\mu^{X_k} = \mu^{X_k^+} - \mu^{X_k^-}$$

*with $\mu^{X_k^+}$ and $\mu^{X_k^-}$ two finite positive random measures. In the same way,*

$$\mu^{X^1_{k^1} \otimes ... \otimes X^n_{k^n}} = \mu^{(X^1_{k^1} \otimes ... \otimes X^n_{k^n})^+} - \mu^{(X^1_{k^1} \otimes ... \otimes X^n_{k^n})^-}$$

*with* $\mu^{(X^1_{k^1}\otimes\ldots\otimes X^n_{k^n})^+}$ *and* $\mu^{(X^1_{k^1}\otimes\ldots\otimes X^n_{k^n})^-}$ *two finite positive random measures. We will later show that the expectations of* $\mu^{X^+_k}$, $\mu^{X^-_k}$, $\mu^{(X^1_{k^1}\otimes\ldots\otimes X^n_{k^n})^+}$ *and* $\mu^{(X^1_{k^1}\otimes\ldots\otimes X^n_{k^n})^-}$ *are finite for the processes that we will consider (*$\Lambda$*,* $N$ *or* $M$*). Therefore, we can define*

$$\mu^{\mathbb{E}[X_k]} = \mathbb{E}[\mu^{X_k}] \text{ and } \mu^{\mathbb{E}[X^1_{k^1}\otimes\ldots\otimes X^n_{k^n}]} = \mathbb{E}[\mu^{X^1_{k^1}\otimes\ldots\otimes X^n_{k^n}}].$$

**Lemma A.1.** *Let* $\mu$ *be a signed random measure on* $(\Omega,\Sigma)$*, which writes* $\mu = \mu^+ - \mu^-$ *with* $\mu^+$ *and* $\mu^-$ *two finite positive random measures whose expectations are finite on* $\Omega$*. For any (deterministic)* $\Sigma$*-measurable bounded function* $g : \Omega \to \mathbb{R}$*, we have*

$$\mathbb{E}[\int g d\mu] = \int g d\mathbb{E}[\mu].$$

*Proof.* If $\mu$ is a positive random measure, $g \mapsto \mathbb{E}[\int g d\mu]$ coincides with $g \mapsto \int g d\mathbb{E}[\mu]$ for $g = 1_S$ with $S \in \Sigma$. Therefore, using Beppo Levy Theorem, these two linear forms coincide for any positive measurable function $g$ and so for any bounded measurable function $g$. We then only need to decompose $\mu = \mu^+ - \mu^-$. □

**Lemma A.2.** *In the above framework, the expectations of the product measures can be characterized on the products of closed intervals*[3] *as:*

$$\mu^{\mathbb{E}[X_k]}([r_1,s_1]\times\ldots\times[r_k,s_k]) = \mathbb{E}[(X_{s_1} - X_{r_1^-})\ldots(X_{s_k} - X_{r_k^-})], \tag{13}$$

*and*

$$\mu^{\mathbb{E}[X^1_{k^1}\otimes\ldots\otimes X^n_{k^n}]}([r_1,s_1]\times\ldots\times[r_k,s_k]) = \mathbb{E}[(X^1_{s_1} - X^1_{r_1^-})\ldots(X^1_{s_{k_1}} - X^1_{r^-_{k_1}})\ldots(X^n_{s_k} - X^n_{r_k^-})].$$

*Proof.* Let us prove (13). The measure $\mu^{X_k}$ is a sum of $2^k$ terms:

$$\mu^{X_k} = \mu^{X^+_k} - \mu^{X^-_1\otimes X^+_{k-1}} - \mu^{X^+_1\otimes X^-_1\otimes X^+_{k-2}} + \ldots + (-1)^k\mu^{X^-_k}.$$

Applying it to a product of closed intervals and taking expectation, we get

$$\mu^{\mathbb{E}[X_k]}([r_1,s_1]\times\ldots\times[r_k,s_k]) = \mu^{E[X^+_k]}([r_1,s_1]\times\ldots\times[r_k,s_k])+\ldots+(-1)^k\mu^{E[X^-_k]}([r_1,s_1]\times\ldots\times[r_k,s_k]),$$

which is precisely equal to

$$\mathbb{E}[(X_{s_1} - X_{r_1^-})\times\ldots\times(X_{s_k} - X_{r_k^-})].$$

□

**Remark A.3.** *For the processes* $X$ *that we will consider here, for any* $t \geq 0$*, almost surely,* $X_t = X_{t^-}$*, therefore*

$$\mu^{\mathbb{E}[X_k]}([r_1,s_1]\times\ldots\times[r_k,s_k]) = \mathbb{E}[(X_{s_1} - X_{r_1})\ldots(X_{s_k} - X_{r_k})],$$

*and*

$$\mu^{\mathbb{E}[X^1_{k^1}\otimes\ldots\otimes X^n_{k^n}]}([r_1,s_1]\times\ldots\times[r_k,s_k]) = \mathbb{E}[(X^1_{s_1} - X^1_{r_1})\ldots(X^1_{s_{k_1}} - X^1_{r_{k_1}})\ldots(X^n_{s_k} - X^n_{r_k})].$$

[3]Note that when we write an interval $[r,s]$, we always assume that $r \leq s$.

These concepts will be useful since they will allow us to compute from the expectations of product measures important expressions. Indeed, we have the following lemma.

**Lemma A.3.** *We have*

$$\mathbb{E}[(X_s - X_r)^k] = \int_{I^k} \mu^{\mathbb{E}[X_k]}(dt_1, ..., dt_k),$$

*with* $I = [r, s]$*, and for* $f$ *bounded measurable*

$$\mathbb{E}[(\int_I f(t)dX_t)^k] = \int_{I^k} f(t_1)...f(t_k)\mu^{\mathbb{E}[X_k]}(dt_1, ..., dt_k).$$

Furthermore, we have the following characterization result.

**Lemma A.4.** *Let* $\mu_1$ *and* $\mu_2$ *be two signed (random) measures on* $([0,T]^k, B([0,T]^k))$*, which write* $\mu_i = \mu_i^+ - \mu_i^-$ *with* $\mu_i^+$ *and* $\mu_i^-$ *two finite positive random measures. If* $\mu_1$ *and* $\mu_2$ *coincide on the products of closed intervals, then* $\mu_1$ *and* $\mu_2$ *are equal.*

*Proof.* For any product of closed intervals $R$,

$$\mu_1^+(R) - \mu_1^-(R) = \mu_2^+(R) - \mu_2^-(R)$$

Therefore,

$$\mu_1^+(R) + \mu_2^-(R) = \mu_2^+(R) + \mu_1^-(R).$$

$\mu_1^+ + \mu_2^-$ and $\mu_2^+ + \mu_1^-$ are two finite positive measures which coincide on the products of closed intervals therefore they are equal and for every Borel set $A$:

$$\mu_1^+(A) + \mu_2^-(A) = \mu_2^+(A) + \mu_1^-(A),$$

Therefore,

$$\mu_1^+(A) - \mu_1^-(A) = \mu_2^+(A) - \mu_2^-(A),$$

which ends the proof. □

In the next paragraphs, we will compute some of these expectations of product measures for $X = N$ or $M$ and $k = 1, 2, 3$ or 4.

## A.2 First order properties

We start from a result which we borrow from [5].

**Lemma A.5.** *There exists* $c$ *such that for all* $t, s \geq 0$ *and* $T \geq 1$*,*

$$\mathbb{E}[\lambda_t^T] = \mu(1 + \int_0^t \psi^T(t-s)ds) \leq cT.$$

$$\mathbb{E}[N_t^T - N_s^T] = \int_s^t \mathbb{E}[\lambda_u^T]du \leq cT(t-s).$$

## A.3 Second order properties

**Lemma A.6.** *The second moment measure of the martingale $M$ satisfies*

$$\mu^{\mathbb{E}[M_2]}(dt_1, dt_2) = \delta_{t_1}(dt_2)\mathbb{E}[\lambda(t_1)]dt_1.$$

**Remark A.4.** *By Lemma A.6, we mean that for every measurable function $h : [0,T]^2 \to \mathbb{R}$:*

$$\int_{[0,T]^2} h(t_1, t_2)\mu^{\mathbb{E}[M_2]}(dt_1, dt_2) = \int_{[0,T]^2} h(t_1, t_2)\delta_{t_1}(dt_2)\mathbb{E}[\lambda(t_1)]dt_1.$$

**Remark A.5.** *By symmetry between $t_1$ and $t_2$ this lemma also writes*

$$\mu^{\mathbb{E}[M_2]}(dt_1, dt_2) = \delta_{t_2}(dt_1)\mathbb{E}[\lambda(t_2)]dt_2.$$

*Proof.* From Remark A.2, $\mu^{M_2}$ is equal to the difference of two positive finite random measures and thus, by definition, so is $\mu^{\mathbb{E}[M_2]}$. We prove that for a partition $(E_-, E_+, E_0)$ of $[0,T]^2$ that we define below, these two measures restricted to $E_-$, $E_+$ and $E_0$ coincide. To do so, we show that the value of these two measures applied to any product of closed intervals included into $E_-$, $E_+$ or $E_0$ are equal and use Lemma A.4.

- Let $R_- = [r_1, s_1] \times [r_2, s_2]$ be a product of closed intervals included in $E_- = \{(x, y) \in [0,T]^2; y > x \geq 0\}$:

  $$\mu^{\mathbb{E}[M_2]}(R_-) = \mathbb{E}[(M_{s_1} - M_{r_1})(M_{s_2} - M_{r_2})] = 0.$$

  because $M$ is a martingale and (since $R_- \subset E_-$) $s_1 < r_2$.

  Therefore $\mu^{\mathbb{E}[M_2]}$ restricted to $E_-$ is null.

- In the same way, $\mu^{\mathbb{E}[M_2]}$ restricted to $E_+ = \{(x, y); x > y \geq 0\}$ is null.

- Let $R_0 = \{(x, y) \in [0,T]^2; t_1 \leq y = x \leq t_2\}$ be a product of closed intervals of $E_0 = \{(x, y) \in [0,T]^2; y = x\}$, then almost surely:

  $$\begin{aligned} \mu^{M_2}(R_0) &= (\mu^{N_2} + \mu^{\Lambda_2} - \mu^{N_1 \otimes \Lambda_1} - \mu^{\Lambda_1 \otimes N_1})(R_0) \\ &= \int_{t_1}^{t_2} (N_{s^+} - N_{s^-})dN_s \\ &= N_{t_2} - N_{t_1}. \end{aligned}$$

  Therefore:

  $$\mu^{\mathbb{E}[M_2]}(R_0) = \mathbb{E}[N_{t_1} - N_{t_2}] = \int_{R_0} \delta_{t_1}(dt_2)\mathbb{E}[\lambda(t_1)]dt_1.$$

This implies the lemma.

□

**Lemma A.7.** *Let us now consider $M_1 \otimes \Lambda_1$. We have*

$$\mu^{\mathbb{E}[M_1 \otimes \Lambda_1]}(dt_1, dt_2) = \mathbb{E}[\lambda(t_1)]\psi(t_2 - t_1)dt_1 dt_2.$$

*Proof.* From Remark A.2, $\mu^{M_1\otimes\Lambda_1}$ is equal to the difference of two positive finite random measures and thus, by definition, so does $\mu^{\mathbb{E}[M_1\otimes\Lambda_1]}$. Let us prove this equality on any product of closed intervals of $[0,T]^2$: $R=[r_1,s_1]\times[r_2,s_2]$ and use Lemma A.4. Using the decomposition

$$\lambda_u=\mathbb{E}[\lambda_u]+\int_0^u\psi(u-s)dM_s,$$

we get

$$\begin{aligned}\mu^{\mathbb{E}[M_1\otimes\Lambda_1]}(R) &= \mathbb{E}[(M(s_1)-M(r_1))\int_{r_2}^{s_2}\lambda_u du]\\ &= \mathbb{E}[\int_{r_1}^{s_1}dM_v(\int_{r_2}^{s_2}(\mathbb{E}[\lambda_u]+\int_0^u\psi(u-s)dM_s)du)].\end{aligned}$$

Now since if $x<0$ then $\psi(x)=0$, we have

$$\begin{aligned}&= \mathbb{E}[\int_{r_1}^{s_1}\int_0^{s_2}(\int_{r_2}^{s_2}\psi(u-s)du)dM_s dM_v]\\ &= \int_{r_1}^{s_1}\int_0^{s_2}(\int_{r_2}^{s_2}\psi(u-s)du)\mu^{\mathbb{E}[M_2]}(ds,dv)\\ &= \int_{r_2}^{s_2}\int_{r_1}^{s_1}\psi(u-v)\mathbb{E}[\lambda_v]dvdu.\end{aligned}$$

This ends the proof. □

**Lemma A.8.** *We have*

$$\mu^{\mathbb{E}[\Lambda_2]}(dt_1,dt_2)=\mathbb{E}[\lambda(t_1)]\mathbb{E}[\lambda(t_2)]dt_1dt_2+\big(\int_0^{t_1\wedge t_2}\mathbb{E}[\lambda(u)]\psi(t_1-u)\psi(t_2-u)du\big)dt_1dt_2. \quad (14)$$

*Proof.* As in the previous proof, we will prove the equality of the two measures on any product of closed intervals of $[0,T]^2$: $R=[r_1,s_1]\times[r_2,s_2]$ and use the decomposition

$$\lambda_{t_1}=\mathbb{E}[\lambda_{t_1}]+\int_0^{t_1}\psi(t_1-u)dM_u.$$

We have

$$\begin{aligned}\mu^{\mathbb{E}[\Lambda_2]}(R) &= \mathbb{E}[\int_{r_1}^{s_1}\lambda_{t_1}dt_1\int_{r_2}^{s_2}\lambda_{t_2}dt_2]\\ &= \int_{r_1}^{s_1}\int_{r_2}^{s_2}\mathbb{E}[\lambda_{t_1}]\mathbb{E}[\lambda_{t_2}]dt_1dt_2+\mathbb{E}[\int_{r_2}^{s_2}\int_0^{s_1}\int_{r_1}^{s_1}\psi(t_1-u)dt_1dM_u d\Lambda_{t_2}]\\ &= \int_{r_1}^{s_1}\int_{r_2}^{s_2}\mathbb{E}[\lambda_{t_1}]\mathbb{E}[\lambda_{t_2}]dt_1dt_2+\int_{r_2}^{s_2}\int_0^{s_1}\int_{r_1}^{s_1}\psi(t_1-u)dt_1\mu^{\mathbb{E}[M_1\otimes\Lambda_1]}(du,dt_2)\\ &= \int_{r_1}^{s_1}\int_{r_2}^{s_2}\mathbb{E}[\lambda_{t_1}]\mathbb{E}[\lambda_{t_2}]dt_1dt_2+\int_{r_1}^{s_1}\int_{r_2}^{s_2}(\int_0^{t_1\wedge t_2}\mathbb{E}[\lambda(u)]\psi(t_1-u)\psi(t_2-u)du)dt_1dt_2\end{aligned}$$

□

Using the previous lemmas, we obtain the following result.

**Lemma A.9.** *We have*

$$\begin{aligned}\mu^{\mathbb{E}[N_2]}(dt_1,dt_2) &= \delta_{t_1}(dt_2)\mathbb{E}[\lambda(t_1)]dt_1 + \mathbb{E}[\lambda(t_1)]\psi(t_2-t_1)dt_1dt_2 \\ &+ \mathbb{E}[\lambda(t_2)]\psi(t_1-t_2)dt_1dt_2 + \Big(\int_0^{t_1\wedge t_2}\mathbb{E}[\lambda(u)]\psi(t_1-u)\psi(t_2-u)du\Big)dt_1dt_2\end{aligned}$$

*Proof.* Indeed, we only need to notice that

$$\mu^{\mathbb{E}[N_2]} = \mu^{\mathbb{E}[M_2]} + \mu^{\mathbb{E}[M_1\otimes\Lambda_1]} + \mu^{\mathbb{E}[\Lambda_1\otimes M_1]} + \mu^{\mathbb{E}[\Lambda_2]}$$

and use the symmetry property of product measures:

$$\mu^{\mathbb{E}[X_1\otimes X_2]}(dt_1,dt_2) = \mu^{\mathbb{E}[X_2\otimes X_1]}(dt_2,dt_1)$$

to prove the lemma. □

Applied to our sequence of Hawkes processes, this gives us the following second order bound.

**Lemma A.10.** *There exists $c$ such that for all $t > s \geq 0$, $T \geq 1$:*

$$\mathbb{E}[(N_t^T - N_s^T)^2 + (\Lambda_t^T - \Lambda_s^T)^2] \leq c(T(t-s) + T^2(t-s)^2).$$

*Proof.* We use Lemma A.3 and the bounds $\|\psi^T\|_1 \leq cT$, $\|\psi^T\|_\infty \leq c$ and $\mathbb{E}[\lambda_t^T] \leq cT$. □

## A.4 Third order properties

**Lemma A.11.** *The expectation of the third order product measure of $M$ satisfies*

$$\begin{aligned}\mu^{\mathbb{E}[M_3]}(dt_1,dt_2,dt_3) &= \delta_{t_3}(dt_2)\delta_{t_2}(dt_1)\mathbb{E}[\lambda(t_1)]dt_3 + \mathbb{E}[\lambda(t_2)]\delta_{t_3}(dt_2)\psi(t_2-t_1)dt_1dt_3 \\ &+ \mathbb{E}[\lambda(t_3)]\delta_{t_1}(dt_3)\psi(t_3-t_2)dt_2dt_1 + \mathbb{E}[\lambda(t_1)]\delta_{t_2}(dt_1)\psi(t_1-t_3)dt_3dt_2.\end{aligned}$$

*Proof.* We proceed as in the proof of Lemma A.6 considering $[0,T]^3$ as the disjoint union of

- $\{(x,y,z)\in[0,T]^3; x,y<z\}\cup\{(x,y,z)\in[0,T]^3; x,z<y\}\cup\{(x,y,z)\in[0,T]^3; z,y<x\}$ on which $\mu^{\mathbb{E}[M_3]}$ is null. Indeed, if $[s_1,t_1]\times[s_2,t_2]\times[s_3,t_3]\subset\{(x,y,z)\in[0,T]^3; x,y<z\}$ (then $s_3>t_2,t_1$), $\mu^{\mathbb{E}[M_3]}([s_1,t_1]\times[s_2,t_2]\times[s_3,t_3]) = E[(M_{t_1}-M_{s_1})(M_{t_2}-M_{s_2})(M_{t_3}-M_{s_3})] = 0$ since $M$ is a martingale.

- $\{(x,y,z)\in[0,T]^3; x<y=z\}$ on which $\mu^{\mathbb{E}[M_3]}(dt_1,dt_2,dt_3) = \mathbb{E}[\lambda(t_2)]\delta_{t_3}(dt_2)\psi(t_2-t_1)dt_1dt_3$

  Indeed, let us consider a product of closed intervals $R=\{(x,y,z)\in[0,T]^3; r_1\leq x\leq s_1<r_2\leq y=z\leq s_2\}$ of this set. Almost surely,

$$\begin{aligned}\mu^{M_3}(R) &= (M_{s_1}-M_{r_1})\times\int_{r_2}^{s_2}(M_{x^+}-M_{x^-})dM_x \\ &= (M_{s_1}-M_{r_1})\times(N_{s_2}-N_{r_2}) \\ &= (M_{s_1}-M_{r_1})\times((M_{s_2}-M_{r_2})+(\Lambda_{s_2}-\Lambda_{r_2})).\end{aligned}$$

  Therefore, taking expectations and using Lemmas A.6 and A.7, we get our expression.

- $\{(x,y,z) \in [0,T]^3; y < x = z\}$ on which $\mu^{\mathbb{E}[M_3]}(dt_1, dt_2, dt_3) = \mathbb{E}[\lambda(t_3)]\delta_{t_1}(dt_3)\psi(t_3 - t_2)dt_2dt_1$ in the same way.
- $\{(x,y,z) \in [0,T]^3; z < y = x\}$ on which $\mu^{\mathbb{E}[M_3]}(dt_1, dt_2, dt_3) = \mathbb{E}[\lambda(t_1)]\delta_{t_2}(dt_1)\psi(t_1 - t_3)dt_3dt_2$ in the same way.
- $\{(x,y,z) \in [0,T]^3; x = y = z\}$ on which $\mu^{\mathbb{E}[M_3]}(dt_1, dt_2, dt_3) = \delta_{t_3}(dt_2)\delta_{t_2}(dt_1)\mathbb{E}[\lambda(t_1)]dt_3$.

□

**Lemma A.12.** *Let $|\mu|$ be the absolute variation of the measure $\mu$. We have*

$$\begin{aligned} |\mu^{\mathbb{E}[M_2\otimes\Lambda_1]}|(dt_1, dt_2, dt_2) &\leq \delta_{t_2}(dt_1)dt_2dt_3(\mu^2(1+\|\psi\|_1)^2 + \mu\|\psi\|_\infty(1+\|\psi\|_1) \\ &+ \mu(1+\|\psi\|_1)^2\|\psi\|_\infty) + dt_1dt_2dt_3(2(1+\|\psi\|_1)\|\psi\|_\infty). \end{aligned}$$

*Proof.* We write

$$\lambda_u = \mathbb{E}[\lambda_u] + \int_0^u \psi(u-v)dM_v.$$

Then, we proceed as in the proof of Lemma A.7 to get

$$\begin{aligned} \mu^{\mathbb{E}[M_2\otimes\Lambda_1]}(dt_1, dt_2, dt_3) &= \mathbb{E}[\lambda(t_1)]\delta_{t_2}(dt_1)\mathbb{E}[\lambda(t_3)]dt_2dt_3 + \psi(t_3-t_2)\delta_{t_2}(dt_1)\mathbb{E}[\lambda(t_1)]dt_2dt_3 \\ &+ \mathbb{E}[\lambda(t_1)]\psi(t_3-t_2)\psi(t_2-t_1)dt_1dt_2dt_3 \\ &+ \mathbb{E}[\lambda(t_2)]\psi(t_3-t_1)\psi(t_1-t_2)dt_1dt_2dt_3 \\ &+ \delta_{t_2}(dt_1)[\int_0^{t_3} \mathbb{E}[\lambda(s)]\psi(t_3-s)\psi(t_1-s)ds]dt_2dt_3. \end{aligned}$$

Finally, we use that for all $t$, $\mathbb{E}[\lambda(t)] \leq \mu(1+\|\psi\|_1)$.

□

In the same way we get the following lemmas.

**Lemma A.13.** *We have*

$$\begin{aligned} |\mu^{\mathbb{E}[M_1\otimes\Lambda_2]}|(dt_1, dt_2, dt_3) &\leq dt_1dt_2dt_3(\|\psi\|_\infty\mu(1+\|\psi\|_1) + \|\psi\|_\infty(\mu^2(1+\|\psi\|_1)^2 \\ &+ \mu\|\psi\|_\infty(1+\|\psi\|_1) + \mu(1+\|\psi\|_1)^2\|\psi\|_\infty) + 2(1+\|\psi\|_1)^2\|\psi\|_\infty). \end{aligned}$$

**Lemma A.14.** *The quantity $|\mu^{\mathbb{E}[\Lambda_3]}|(dt_1, dt_2, dt_3)$ is smaller than*

$$\begin{aligned} &dt_1dt_2dt_3[\mu(1+\|\psi\|_1)^2\|\psi\|_\infty + (1+\|\psi\|_1)(\|\psi\|_\infty\mu(1+\|\psi\|_1) + \|\psi\|_\infty(\mu^2(1+\|\psi\|_1)^2 \\ &+ \mu\|\psi\|_\infty(1+\|\psi\|_1) + \mu(1+\|\psi\|_1)^2\|\psi\|_\infty) + 2(1+\|\psi\|_1)^2\|\psi\|_\infty)]. \end{aligned}$$

This gives us the following third order bound.

**Lemma A.15.** *There exists $c$ such that for all, $t, s \geq 0$ and $T \geq 1$,*

$$\mathbb{E}[(N_t^T - N_s^T)^3 + (\Lambda_t^T - \Lambda_s^T)^3] \leq c(T(t-s) + T^3(t-s)^3).$$

*Proof.* Let us write $I = [s,t]$. Using Lemma A.3, we have

$$\mathbb{E}[|(N_t^T - N_s^T)^3|] \leq \int_{I^3} |\mu^{\mathbb{E}[N_3^T]}|(dt_1, dt_2, dt_3).$$

Moreover,

$$\begin{aligned} |\mu^{\mathbb{E}[N_3^T]}| &\leq |\mu^{\mathbb{E}[M_3^T]}| + |\mu^{\mathbb{E}[M_2^T \otimes \Lambda_1^T]}| + |\mu^{\mathbb{E}[\Lambda_1^T \otimes M_2^T]}| + |\mu^{\mathbb{E}[M_1^T \otimes \Lambda_1^T \otimes M_1^T]}| \\ &+ |\mu^{\mathbb{E}[M_1^T \otimes \Lambda_2^T]}| + |\mu^{\mathbb{E}[\Lambda_2^T \otimes M_1^T]}| + |\mu^{\mathbb{E}[\Lambda_1^T \otimes M_1^T \otimes \Lambda_1^T]}| + |\mu^{\mathbb{E}[\Lambda_3^T]}|. \end{aligned}$$

Therefore, from the previous Lemmas and simple symmetry properties of product measures, we have (using that $(1 + \|\psi^T\|_1) = cT$ and $\|\psi^T\|_\infty \leq c$)

$$\begin{aligned} |\mu^{\mathbb{E}[N_3^T]}|(dt_1, dt_2, dt_3) &\leq c(T\delta_{t_2}(dt_1)\delta_{t_3}(dt_2)dt_3 + T^2\delta_{t_2}(dt_1)dt_2dt_3 \\ &+ T^2\delta_{t_3}(dt_2)dt_3dt_1 + T^2\delta_{t_1}(dt_3)dt_1dt_2 + T^3dt_1dt_2dt_3). \end{aligned}$$

The integration of this bound on $[s,t]^3$ gives the result. □

## A.5 Fourth order properties

Proceeding as in the proof of Lemmas A.6 and A.11, we get the following result.

**Lemma A.16.** *For $0 \leq t_1 \leq t_2 \leq t_3 \leq t_4$, $\mu^{\mathbb{E}[M_4]}(dt_1, dt_2, dt_3, dt_4)$ is equal to*

$$\begin{aligned} & \delta_{t_3}(dt_4)[\delta_{t_3}(dt_2)\delta_{t_2}(dt_1)\mathbb{E}[\lambda(t_1)]dt_3 + \mathbb{E}[\lambda(t_1)]\delta_{t_3}(dt_2)\psi(t_2 - t_1)dt_1dt_3 \\ + & \mathbb{E}[\lambda(t_2)]\delta_{t_1}(dt_3)\psi(t_3 - t_2)dt_2dt_1 + \mathbb{E}[\lambda(t_3)]\delta_{t_2}(dt_1)\psi(t_1 - t_3)dt_3dt_2 \\ + & \mathbb{E}[\lambda(t_1)]\delta_{t_2}(dt_1)\mathbb{E}[\lambda(t_3)]dt_2dt_3 + \psi(t_3 - t_2)\delta_{t_2}(dt_1)\mathbb{E}[\lambda(t_1)]dt_2dt_3 \\ + & \mathbb{E}[\lambda(t_1)]\psi(t_3 - t_2)\psi(t_2 - t_1)dt_1dt_2dt_3 + \mathbb{E}[\lambda(t_2)]\psi(t_3 - t_1)\psi(t_1 - t_2)dt_1dt_2dt_3 \\ + & \delta_{t_2}(dt_1)[\int_0^{t_3} \mathbb{E}[\lambda(s)]\psi(t_3 - s)\psi(t_1 - s)ds]dt_2dt_3]. \end{aligned}$$

Denoting by $S_n$ the $n$-permutation group, this gives us the following bound.

**Lemma A.17.** *There exists $c$ such that, for any positive $t_1, t_2, t_3, t_4 \geq 0$ and $T \geq 1$,*

$$\begin{aligned} |\mu^{\mathbb{E}[M_4^T]}|(dt_1, dt_2, dt_3, dt_4) &\leq c \sum_{\sigma \in S_4} [T\delta_{t_{\sigma(1)}}(dt_{\sigma(2)})dt_{\sigma(1)}\delta_{t_{\sigma(2)}}(dt_{\sigma(3)})\delta_{t_{\sigma(3)}}(dt_{\sigma(4)}) \\ &+ T\delta_{t_{\sigma(1)}}(dt_{\sigma(2)})dt_{\sigma(1)}dt_{\sigma(3)}dt_{\sigma(4)} \\ &+ T^2\delta_{t_{\sigma(1)}}(dt_{\sigma(2)})dt_{\sigma(1)}\delta_{t_{\sigma(3)}}(dt_{\sigma(4)})dt_{\sigma(3)} \\ &+ T\delta_{t_{\sigma(1)}}(dt_{\sigma(2)})dt_{\sigma(1)}\delta_{t_{\sigma(2)}}(dt_{\sigma(3)})dt_{\sigma(4)}] \end{aligned}$$

*Proof.* From Lemma A.16 and using the bounds $\|\psi^T\|_1 \leq cT$, $\|\psi^T\|_\infty \leq c$ and $\mathbb{E}[\lambda_t^T] \leq cT$, we have that for $t_1 \leq t_2 \leq t_3 \leq t_4$,

$$\begin{aligned} |\mu^{\mathbb{E}[M_4^T]}|(dt_1, dt_2, dt_3, dt_4) &\leq \delta_{t_3}(dt_4)c[T\delta_{t_2}(dt_3)\delta_{t_1}(dt_2)dt_1 \\ &+ T\delta_{t_2}(dt_3)dt_1dt_2 + T\delta_{t_3}(dt_1)dt_2dt_3 \\ &+ T\delta_{t_1}(dt_2)dt_3dt_1 + Tdt_1dt_2dt_3 + T^2\delta_{t_1}(dt_2)dt_1dt_3] \end{aligned}$$

Therefore, by considering $[0,T]^4$ as the union of the 4! sets of the form $\{(t_1, t_2, t_3, t_4); t_{\sigma(1)} \leq t_{\sigma(2)} \leq t_{\sigma(3)} \leq t_{\sigma(4)}\}$ where $\sigma$ is a permutation, we get the result.

□

## Acknowledgments

We thank Emmanuel Bacry for several interesting discussions.

## References

[1] L. Adamopoulos. Cluster models for earthquakes: Regional comparisons. *Journal of the International Association for Mathematical Geology*, 8(4):463–475, 1976.

[2] Y. Aït-Sahalia, J. Cacho-Diaz, and R. J. Laeven. Modeling financial contagion using mutually exciting jump processes. Technical report, National Bureau of Economic Research, 2010.

[3] M. B. Alaya and A. Kebaier. Parameter estimation for the square-root diffusions: Ergodic and nonergodic cases. *Stochastic Models*, 28(4):609–634, 2012.

[4] E. Bacry, K. Dayri, and J.-F. Muzy. Non-parametric kernel estimation for symmetric Hawkes processes. application to high frequency financial data. *The European Physical Journal B*, 85(5):1–12, 2012.

[5] E. Bacry, S. Delattre, M. Hoffmann, and J. F. Muzy. Scaling limits for Hawkes processes and application to financial statistics. *arXiv preprint arXiv:1202.0842*, 2012.

[6] E. Bacry, S. Delattre, M. Hoffmann, and J.-F. Muzy. Modelling microstructure noise with mutually exciting point processes. *Quantitative Finance*, 13(1):65–77, 2013.

[7] E. Bacry and J.-F. Muzy. Hawkes model for price and trades high-frequency dynamics. *arXiv preprint arXiv:1301.1135*, 2013.

[8] M. Barczy, M. Ispany, and G. Pap. Asymptotic behavior of unstable INAR (p) processes. *Stochastic Processes and their Applications*, 121(3):583–608, 2011.

[9] L. Bauwens and N. Hautsch. Dynamic latent factor models for intensity processes. 2004.

[10] P. Billingsley. *Convergence of probability measures*, volume 493. Wiley. com, 2009.

[11] J.-P. Bouchaud, Y. Gefen, M. Potters, and M. Wyart. Fluctuations and response in financial markets: the subtle nature of "random" price changes. *Quantitative Finance*, 4(2):176–190, 2004.

[12] C. G. Bowsher. Modelling security market events in continuous time: Intensity based, multivariate point process models. *Journal of Econometrics*, 141(2):876–912, 2007.

[13] P. Brémaud and L. Massoulié. Hawkes branching point processes without ancestors. *Journal of applied probability*, 38(1):122–135, 2001.

[14] J. C. Cox, J. E. Ingersoll Jr, and S. A. Ross. A theory of the term structure of interest rates. *Econometrica: Journal of the Econometric Society*, pages 385–407, 1985.

[15] D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes: volume I*, volume 1. Springer, 2002.

[16] D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes: volume II: general theory and structure*, volume 2. Springer, 2007.

[17] P. Embrechts, T. Liniger, and L. Lin. Multivariate Hawkes processes: an application to financial data. *Journal of Applied Probability*, 48:367–378, 2011.

[18] E. Errais, K. Giesecke, and L. R. Goldberg. Affine point processes and portfolio credit risk. *SIAM Journal on Financial Mathematics*, 1(1):642–665, 2010.

[19] V. Filimonov and D. Sornette. Quantifying reflexivity in financial markets: Toward a prediction of flash crashes. *Physical Review E*, 85(5):056108, 2012.

[20] V. Filimonov and D. Sornette. Apparent criticality and calibration issues in the Hawkes self-excited point process model: application to high-frequency financial data. *arXiv preprint arXiv:1308.6756*, 2013.

[21] S. J. Hardiman, N. Bercot, and J.-P. Bouchaud. Critical reflexivity in financial markets: a Hawkes process analysis. *arXiv preprint arXiv:1302.1405*, 2013.

[22] A. G. Hawkes. Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 438–443, 1971.

[23] A. G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971.

[24] A. G. Hawkes and D. Oakes. A cluster process representation of a self-exciting process. *Journal of Applied Probability*, pages 493–503, 1974.

[25] S. L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of financial studies*, 6(2):327–343, 1993.

[26] P. Hewlett. Clustering of order arrivals, price impact and trade path optimisation. In *Workshop on Financial Modeling with Jump processes, Ecole Polytechnique*, pages 6–8, 2006.

[27] J. Jacod. Multivariate point processes: predictable projection, Radon-Nikodym derivatives, representation of martingales. *Probability Theory and Related Fields*, 31(3):235–253, 1975.

[28] J. Jacod and A. N. Shiryaev. *Limit theorems for stochastic processes*, volume 288. Springer-Verlag Berlin, 1987.

[29] T. Jaisson. Market impact as anticipation of the order flow imbalance. *Working paper*, 2013.

[30] A. Jakubowski, J. Mémin, and G. Pagès. Convergence en loi des suites d'intégrales stochastiques sur l'espace 1 de Skorokhod. *Probability Theory and Related Fields*, 81(1):111–137, 1989.

[31] V. V. Kalashnikov. *Geometric sums: bounds for rare events with applications: risk analysis, reliability, queueing*, volume 413. Springer, 1997.

[32] T. G. Kurtz and P. Protter. Weak limit theorems for stochastic integrals and stochastic differential equations. *The Annals of Probability*, 19(3):1035–1070, 1991.

[33] J. Large. Measuring the resiliency of an electronic limit order book. *Journal of Financial Markets*, 10(1):1–25, 2007.

[34] F. Lillo and J. D. Farmer. The long memory of the efficient market. *Studies in Nonlinear Dynamics & Econometrics*, 8(3), 2004.

[35] Y. Ogata. The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Annals of the Institute of Statistical Mathematics*, 30(1):243–261, 1978.

[36] Y. Ogata. Likelihood analysis of point processes and its application to seismological data. *Bull. Int. Statist. Inst*, 50:943–961, 1983.

[37] V. V. Petrov. *Sums of independent random variables.* Berlin, 1975.

[38] P. Reynaud-Bouret and S. Schbath. Adaptive estimation for Hawkes processes; application to genome analysis. *The Annals of Statistics*, 38(5):2781–2822, 2010.

[39] P. Shah and R. Jana. Results on generalized Mittag-Leffler function via Laplace transform. *Applied Mathematical Sciences*, 7(12):567–570, 2013.

[40] M. Wyart, J.-P. Bouchaud, J. Kockelkoren, M. Potters, and M. Vettorazzo. Relation between bid–ask spread, impact and volatility in order-driven markets. *Quantitative Finance*, 8(1):41–57, 2008.