Time-frequency analysis of locally stationary Hawkes processes

FRANÇOIS ROUEFF[1] and RAINER VON SACHS[2]

[1]*LTCI, Télécom ParisTech, Université Paris-Saclay, 46, rue Barrault, 75013, Paris, France.*
E-mail: roueff@telecom-paristech.fr
[2]*Institut de statistique, biostatistique et sciences actuarielles (ISBA) IMMAQ, Université catholique de Louvain, Voie du Roman Pays 20/L1.04.01, 1348 Louvain-la-Neuve. E-mail: rvs@uclouvain.be*

Locally stationary Hawkes processes have been introduced in order to generalise classical Hawkes processes away from stationarity by allowing for a time-varying second-order structure. This class of self-exciting point processes has recently attracted a lot of interest in applications in the life sciences (seismology, genomics, neuro-science, ...), but also in the modeling of high-frequency financial data. In this contribution, we provide a fully developed nonparametric estimation theory of both local mean density and local Bartlett spectra of a locally stationary Hawkes process. In particular, we apply our kernel estimation of the spectrum localised both in time and frequency to two data sets of transaction times revealing pertinent features in the data that had not been made visible by classical non-localised approaches based on models with constant fertility functions over time.

Keywords: high frequency financial data; locally stationary time series; non-parametric kernel estimation; self-exciting point processes; time frequency analysis

1. Introduction

Many recent time series data modelling and analysis problems increasingly face the challenge of occurring time-variations of the underlying probabilistic structure (mean, variance-covariance, spectral structure, ...). This is due to the availability of larger and larger data stretches which can hardly any longer be described by stationary models. Mathematical statisticians (Dahlhaus [7], Zhou and Wu [24], Birr et al. [4], among others) have contributed with time-localised estimation approaches for many of these time series data based on rigorous models of *locally stationary approximations* to the non-stationary data. Often it has been only via these theoretical studies that well-motivated and fully understood *time-dependent estimation* methods could be developed which correctly adapt to the degree of deviation of the underlying data from a stationary situation (e.g., via the modelling of either a slow – or in contrast a rather abrupt – change of the probabilistic structure over time). For many of these situations, the development of an asymptotic theory of doubly-indexed stochastic processes has proven to be useful: the underlying data stretch is considered to be part of a family of processes which asymptotically approaches a limiting process which locally shows all the characteristics of a stationary process (hence "the locally stationary approximation"). Accompanying estimators ought to adapt to this behavior, for example, by introduction of local bandwidths in time. Many of these aforementioned approaches have been achieved in the context of classical real-valued series in discrete time, for example, for (linear) time series via time-varying MA(∞) representations which apply for a large class of models

(see, e.g., Dahlhaus [9]). Once those are available, the development of a rigorous asymptotic estimation theory is achievable.

This approach, however, is not directly applicable to the model of *locally stationary Hawkes processes*, a class of self-exciting point processes introduced in Roueff, von Sachs and Sansonnet [21] with not only time-varying baseline intensity (such as in Chen and Hall [5]) but also time-varying fertility function. The fertility function describes the iterative probabilistic mechanism for Hawkes processes to generate offsprings from each occurrence governed by a conditional Poisson point process, given all previous generations. Often, a fertility function $p(t)$ with exponential decay over time t is assumed. The condition $\int p < 1$ ensures a non-explosive accumulation of consecutive populations of the underlying process. When this fertility function is made varying over time through a second argument, one can still rely on the point process mechanism to derive locally stationary approximations.

In this paper, we treat estimation of the first and second order structure of locally stationary Hawkes processes on the real line, with a (time-dependent) fertility function $p(\cdot; t)$ assumed to be causal, that is, supported on $\mathbb{R}_+$. Whereas for estimating the local mean density of the process it is sufficient to introduce a localisation via a short window (kernel) in time, for estimating the second-order structure, that is, the local Bartlett spectrum of the process, one needs to localise both in time and frequency: this time-frequency analysis will be provided via a pair of kernels which concentrate around a given point (t, ω) in the time-frequency plane. Using an appropriate choice of bandwidths in time and in frequency which tend to zero with rates calibrated to minimize the asymptotic mean-square error between the time-frequency estimator and the true underlying local Bartlett spectrum, one can show consistency of our estimator of the latter one. Note that the use of kernels for non-parametric estimation with counting process is not new. To the best of our knowledge this was introduced first by Ramlau-Hansen [17] for estimating regression point process models such as those introduced in Aalen [1], see also Andersen et al. [2] for a general account on the estimation of such processes.

Self exciting point processes have been recently used for modelling point processes resulting from high frequency financial data such as price jump instants (see, e.g., Bacry et al. [3]) or limit order book events (see, e.g., Zheng, Roueff and Abergel [23]). In this paper, we will illustrate our time frequency analysis approach for point processes on transaction times of two assets (which are Essilor International SA and Total SA).

The rest of the paper is organised as follows. In Section 2, we introduce all necessary preparatory material to develop our estimation theory, including the definition of the population quantities, that is, local mean density and local Bartlett spectra that we wish to estimate. Section 3 defines our kernel estimators and treats asymptotic bias and variance developments of those under regularity assumptions to be given beforehand. Although these developments are far from being direct and straightforward, the resulting rates of convergence are completely intuitive from a usual nonparametric time-frequency estimation point of view. In Section 4, we present, as a result of independent interest in its own, the necessary new techniques of directly controlling moments of non-stationary Hawkes processes. All proofs are deferred to the Appendix. A Supplementary Material is available in Roueff and von Sachs [20]. In particular, Roueff and von Sachs [20], Section 6, presents the analysis of our transaction data sets leading to interesting observations that have not previously been revealed by a classical analysis with time-constant fertility functions.

2. Preparatory material

In this preparatory section, we prepare the ground for developing the presentation of our asymptotic estimation theory of local Bartlett spectra. We do this by introducing a list of useful conventions and definitions, as well as recalling the main concepts and results of Roueff, von Sachs and Sansonnet [21] in as much as they are necessary.

2.1. Conventions and notation

We here set some general conventions and notation adopted all along the paper. Additional ones are introduced in relation with the main assumptions in Section 3.1.

A point process is identified with a random measure with discrete support, $N = \sum_k \delta_{t_k}$ typically, where δ_t is the Dirac measure at point t and $\{t_k\}$ the corresponding (locally finite) random set of points. We use the notation $\mu(g)$ for a measure μ and a function g to express $\int g \, d\mu$ when convenient. In particular, for a measurable set A, $\mu(A) = \mu(\mathbb{1}_A)$ and for a point process N, $N(g) = \sum_k g(t_k)$. The shift operator of lag t is denoted by S^t. For a set A, $S^t(A) = \{x - t, x \in A\}$ and for a function g, $S^t(g) = g(\cdot + t)$, so that $S^t(\mathbb{1}_A) = \mathbb{1}_{S^t(A)}$. One can then compose a measure μ with S^t, yielding for a function g, $\mu \circ S^t(g) = \mu(g(\cdot + t))$.

We also need some notation for the functional norms which we deal with in this work. Usual L^p-norms are denoted by $|h|_p$,

$$|h|_p = \left(\int |h|^p \right)^{1/p}$$

for $p \in [1, \infty)$ and $|h|_\infty$ is the essential supremum on $\mathbb{R}$, $|h|_\infty = \operatorname{ess\,sup}_{t \in \mathbb{R}} |h(t)|$. We also use the following weighted L^p norms which we define to be for any $p \geq 1$, $\beta > 0$, $a \geq 0$ and $h : \mathbb{R} \to \mathbb{C}$,

$$|h|_{(\beta), p} := \left| h \times |\cdot|^\beta \right|_p = \left(\int \left| h(t) t^\beta \right|^p dt \right)^{1/p}, \tag{1}$$

$$|h|_{a, p} := \left| h \times e^{a|\cdot|} \right|_p = \left(\int \left| h(t) \right|^p e^{ap|t|} dt \right)^{1/p} \tag{2}$$

with the above usual essential sup extensions to the case $p = \infty$.

We denote the convolution product by $*$, that is, for any two functions h_1 and h_2,

$$h_1 * h_2(s) = \int h_1(s - t) h_2(t) \, dt.$$

Finally we use for a random variable X the notation

$$\|X\|_p := \left(\mathbb{E} |X|^p \right)^{1/p} \qquad \text{for } p \geq 1. \tag{3}$$

2.2. From stationary to non-stationary and locally stationary Hawkes processes

To start with, we first recall the definition of a stationary (linear) Hawkes process N with immigrant intensity λ_c and fertility function p defined on the positive half-line. The conditional intensity function $\lambda(t)$ of such a process is driven by the fertility function taken at the time distances to previous points of the process, that is, $\lambda(t)$ is given by

$$\lambda(t) = \lambda_c + \int_{-\infty}^{t^-} p(t-s)N(\mathrm{d}s) = \lambda_c + \sum_{t_i < t} p(t-t_i). \tag{4}$$

Here the first integral is to be read as the integral of the "fertility" function p with respect to the counting process N, which is a sum of Dirac masses at (random) points $(t_i)_{i \in \mathbb{Z}}$. The existence of a stationary point processes with conditional intensity (4) holds under the condition $\int p < 1$ and can be constructed as a cluster point process (see Daley and Vere-Jones [11], Example 6.3(c)).

We extend the stationary Hawkes model defined by the conditional intensity (4) to the non-stationary case by authorizing the immigrant intensity λ_c to be a function $\lambda_c(t)$ of time t and also the fertility function p to be time varying, replacing $p(t-s)$ by the more general $p(t-s;t)$. To ensure a locally finite point process in this definition, we impose the two conditions

$$\zeta_1 := \sup_{t \in \mathbb{R}} \int p(s;t)\,\mathrm{d}s < 1 \quad \text{and} \quad |\lambda_c|_\infty < \infty. \tag{5}$$

They yield the existence of a non-stationary point process N with a mean density function which is uniformly bounded by $|\lambda_c|_\infty / (1 - \zeta_1)$ (see Roueff, von Sachs and Sansonnet [21], Definition 1).

As non-stationary Hawkes processes can evolve quite arbitrarily over time, the statistical analysis of them requires to introduce local stationary approximations in the same fashion as time varying autoregressive processes in time series, for which *locally stationary* models have been successfully introduced (see Dahlhaus [7]). Thus, a locally stationary Hawkes process with *local immigrant intensity* $\lambda_c^{<\mathrm{LS}>}(\cdot)$ and *local fertility function* $p^{<\mathrm{LS}>}(\cdot;\cdot)$ is a collection $(N_T)_{T \geq 1}$ of non-stationary Hawkes processes with respective immigrant intensity and fertility function given by $\lambda_{cT}(t) = \lambda_c^{<\mathrm{LS}>}(t/T)$ and varying fertility function given by $p_T(\cdot;t) = p^{<\mathrm{LS}>}(\cdot;t/T)$, see Roueff, von Sachs and Sansonnet [21], Definition 2, where this model is called a *locally stationary* Hawkes process. For a given *real location* t, the scaled location t/T is typically called an *absolute location* and denoted by u or v.

Note that the collection $(N_T)_{T \geq 1}$ of non-stationary Hawkes processes are defined using the same time varying parameters $\lambda_c^{<\mathrm{LS}>}$ and $p^{<\mathrm{LS}>}$ but with the time varying arguments scaled by T. As a result, the larger T is, the slower the parameters evolve along the time.

An assumption corresponding to (5) to guarantee that, for all $T \geq 1$, the non-stationary Hawkes process N_T admits a uniformly bounded mean density function is the following:

$$\zeta_1^{<\mathrm{LS}>} := \sup_{u \in \mathbb{R}} \int p^{<\mathrm{LS}>}(r;u)\,\mathrm{d}r < 1 \quad \text{and} \quad \left|\lambda_c^{<\mathrm{LS}>}\right|_\infty < \infty. \tag{6}$$

Under this assumption, for each absolute location $u \in \mathbb{R}$, the function $r \mapsto p^{<LS>}(r; u)$ satisfies the required condition for the fertility function of a stationary Hawkes process. Hence, assuming (6), for any absolute location u, we denote by $N(\cdot; u)$ a stationary Hawkes process with constant immigrant intensity $\lambda_c^{<LS>}(u)$ and fertility function $r \mapsto p^{<LS>}(r; u)$. In the following subsection, we will include this assumption (6) into a stronger set of assumptions that we use for derivation of the results on asymptotic estimation theory.

We also remark that, for any $T \geq 1$, the conditional intensity function λ_T of the non-stationary Hawkes process N_T takes the form

$$\lambda_T(t) = \lambda_c^{<LS>}(t/T) + \int_{-\infty}^{t^-} p^{<LS>}(t - s; t/T) N_T(ds)$$

$$= \lambda_c^{<LS>}(t/T) + \sum_{t_{i,T} < t} p^{<LS>}(t - t_{i,T}; t/T),$$

where $(t_{i,T})_{i \in \mathbb{Z}}$ denote the points of N_T. This latter formula can also be used to simulate locally stationary Hawkes processes on the real line. The examples for locally stationary Hawkes processes (with time varying Gamma shaped fertility functions) used in Roueff, von Sachs and Sansonnet [21], Section 2.6, were simulated in this way.

First and second order statistics for point processes are of primary importance for statistical inference. As for time series they are conveniently described in the stationary case by a mean parameter for the first order statistics and a spectral representation, the so called Bartlett spectrum (see Daley and Vere-Jones [11], Proposition 8.2.I), for the covariance structure. The locally stationary approach allows us to define such quantities as depending on the absolute time u as introduced in the following section.

2.3. Local mean density and Bartlett spectrum

Consider a locally stationary Hawkes process $(N_T)_{T \geq 1}$ with local immigrant intensity $\lambda_c^{<LS>}$ and local fertility function $p^{<LS>}(\cdot; \cdot)$ satisfying condition (6). Although for a given T, the first and second order statistics of N_T can be quite involved, some intuitive asymptotic approximations are available as T grows to infinity. Namely, for any absolute time u, the *local* statistical behavior of N_T around real time Tu has to be well approximated by that of a stationary Hawkes process with (constant) immigrant intensity $\lambda_c^{<LS>}(u)$ and fertility function $p^{<LS>}(\cdot; u)$. This stationary Hawkes process at absolute location u is denoted in the following by $N(\cdot; u)$. Precise approximation results are provided in Roueff, von Sachs and Sansonnet [21] and recalled in Section 3.4.1 below. Presently, we only need to introduce how to define this local first and second order statistical structure.

We first introduce the *local mean density* function $m_1^{<LS>}(u)$ defined at each absolute location u, as the mean parameter of the stationary Hawkes process $N(\cdot; u)$. By Daley and Vere-Jones [11], Eq. (6.3.26) in Example 6.3(c), it is given by

$$m_1^{<LS>}(u) = \frac{\lambda_c^{<LS>}(u)}{1 - \int p^{<LS>}(\cdot; u)}. \tag{7}$$

A convenient way to describe the covariance structure of a second order stationary point process N on $\mathbb{R}$ is to rely on a spectral representation, the Bartlett spectrum, which is defined as the (unique) non-negative measure Γ on the Borel sets such that, for any bounded and compactly supported function f on $\mathbb{R}$ (see Daley and Vere-Jones [11], Proposition 8.2.I)

$$\text{Var}\big(N(f)\big) = \Gamma\big(|F|^2\big) = \int \big|F(\omega)\big|^2 \Gamma(\mathrm{d}\omega),$$

where F denotes the Fourier transform of f,

$$F(\omega) = \int f(t)\mathrm{e}^{-\mathrm{i}t\omega}\,\mathrm{d}t.$$

For the stationary Hawkes processes $N(\cdot; u)$, the Bartlett spectrum admits a density given by (see Daley and Vere-Jones [11], Example 8.2(e))

$$\gamma^{<\text{LS}>}(u; \omega) = \frac{m_1^{<\text{LS}>}(u)}{2\pi}\big|1 - P^{<\text{LS}>}(\omega; u)\big|^{-2}, \tag{8}$$

where

$$P^{<\text{LS}>}(\omega; u) = \int p^{<\text{LS}>}(t; u)\mathrm{e}^{-\mathrm{i}t\omega}\,\mathrm{d}t.$$

Analogous to the first order structure, we call $\gamma^{<\text{LS}>}(u; \omega)$ the *local Bartlett spectrum density* at frequency ω and absolute location u. This local Bartlett spectrum density plays a role similar to that of the local spectral density $f(u, \lambda)$ introduced in Dahlhaus [7], page 142, for locally stationary time series.

2.4. Estimators

As our approach is local in time and frequency, we rely on two kernels w and q which are required to be compactly supported (see Remark 1 below). More precisely, we have the following assumptions.

   (K-1) Let w be a $\mathbb{R} \to \mathbb{R}_+$ bounded function with compact support such that $\int w = |w|_1 = 1$.
   (K-2) Let q be a $\mathbb{R} \to \mathbb{C}$ bounded function with compact support such that $|q|_2 = \sqrt{2\pi}$.

To localize in time let $b_1 > 0$ be a given time bandwidth and define w_{b_1} and w_{Tb_1} to be the corresponding kernels in absolute time u and real time t, namely,

$$w_{b_1}(u) := b_1^{-1}w(u/b_1) \quad\text{and}\quad w_{Tb_1}(t) := T^{-1}w_{b_1}(t/T) = (Tb_1)^{-1}w\big(t/(Tb_1)\big). \tag{9}$$

Let now u_0 be a fixed absolute time. For estimating the local mean density $m_1^{<\text{LS}>}(u_0)$ given by equation (7), approximating the mean density function $t \mapsto m_{1T}(t)$ of N_T locally in the neighborhood of Tu_0 by $m_1^{<\text{LS}>}(u_0)$ we have, for b_1 small,

$$m_1^{<\text{LS}>}(u_0) \approx \int w_{Tb_1}(t - Tu_0)m_{1T}(t)\,\mathrm{d}t,$$

where we used $\int w_{Tb_1} = 1$ and that the support of $t \mapsto w_{Tb_1}(t - Tu_0)$ essentially lives in the neighborhood of Tu_0 for Tb_1 small. Since the right-hand side of this approximation is $\mathbb{E}[N_T(S^{-Tu_0} w_{Tb_1})]$, this suggests the following estimator of $m_1^{<\text{LS}>}(u_0)$,

$$\widehat{m}_{b_1}(u_0) := N_T\left(S^{-Tu_0} w_{Tb_1}\right) = \int w_{Tb_1}(t - Tu_0) N_T(\mathrm{d}t). \tag{10}$$

For estimation of the second order structure, that is, the local Bartlett spectral density $\gamma^{<\text{LS}>}(u_0; \omega_0)$ for some given point (u_0, ω_0) of the time-frequency plane, we need also to lo-calise in frequency by a kernel which will be given by the (squared) Fourier transform $|Q|^2$ of the kernel q. Then for a given frequency bandwidth $b_2 > 0$, we are looking for an estimator of the auxiliary quantity

$$\gamma_{b_2}^{<\text{LS}>}(u_0; \omega_0) := \int \frac{1}{b_2} \left| Q\left(\frac{\omega - \omega_0}{b_2}\right) \right|^2 \gamma^{<\text{LS}>}(\omega; u_0) \, \mathrm{d}\omega, \tag{11}$$

which in turn, as $b_2 \to 0$, is an approximation of the density $\gamma^{<\text{LS}>}(u_0; \omega_0)$, since (K-2) im-plies $|Q|_2 = 1$. The rate of approximation (i.e., the "bias in frequency direction" of the follow-ing estimator) is established in Theorem 2, equation (27), below. Let us now set $q_{\omega_0, b_2}(t) = b_2^{1/2} e^{i\omega_0 t} q(b_2 t)$ such that the squared modulus of its Fourier transform writes as

$$\left| Q_{\omega_0, b_2}(\omega) \right|^2 = \frac{1}{b_2} \left| Q\left(\frac{\omega - \omega_0}{b_2}\right) \right|^2.$$

Using that $\gamma^{<\text{LS}>}(\omega; u_0) \, \mathrm{d}\omega$ is the Bartlett spectrum of $N(\cdot; u_0)$ as recalled in Section 2.3, we can thus rewrite (11) as

$$\gamma_{b_2}^{<\text{LS}>}(u_0; \omega_0) = \text{Var}\left(N(q_{\omega_0, b_2}; u_0)\right). \tag{12}$$

Since this variance is an approximation of $\text{Var}(N_T(S^{-Tu_0} q_{\omega_0, b_2}))$, where q_{ω_0, b_2} is shifted to be localized around Tu_0, we finally estimate $\gamma_{b_2}^{<\text{LS}>}(u_0; \omega_0)$ by the following moment estimator:

$$\widehat{\gamma}_{b_2, b_1}(u_0; \omega_0) := \widehat{E}\left[\left| N_T\left(S^{-Tu_0} q_{\omega_0, b_2}\right) \right|^2; w_{b_1}\right] - \left| \widehat{E}\left[N_T\left(S^{-Tu_0} q_{\omega_0, b_2}\right); w_{b_1}\right] \right|^2, \tag{13}$$

where for the test function $f = S^{-Tu_0} q_{\omega_0, b_2}$ and taking $\rho(x) = x$ and $\rho(x) = |x|^2$ successively, we have built estimators of $\mathbb{E}[\rho(N_T(f))]$ based on the empirical observations of N_T and defined by

$$\widehat{E}\left[\rho\left(N_T(f)\right); w_{b_1}\right] := \int \rho\left(N_T\left(f(\cdot - t)\right)\right) w_{Tb_1}(t) \, \mathrm{d}t$$

$$= \frac{1}{T} \int \rho\left(\sum_k f(t_{k,T} - t)\right) w_{b_1}(t/T) \, \mathrm{d}t. \tag{14}$$

Note that in (13) the dependence of the estimator on u_0, ω_0 appears in the choice of $f = S^{-Tu_0} q_{\omega_0, b_2}$. By an obvious change of variable, this would be equivalent to let the kernel q_{ω_0, b_2}

unshifted in time, hence take $f = q_{\omega_0, b_2}$, and instead shift $w_{T b_1}(t)$ into $w_{T b_1}(t - T u_0)$, or, in absolute time, shift $w_{b_1}(u)$ into $w_{b_1}(u - u_0)$.

Remark 1. In practice N_T is observed over a finite interval. In order to have estimators $\widehat{m}_{b_1}(u_0)$ and $\widehat{\gamma}_{b_2, b_1}(u_0; \omega_0)$ in (10) and (13) that only use observations within this interval, the supports of w and q must be bounded and some restriction imposed on b_1, b_2 and T. Suppose for instance that N_T is observed on $[0, T]$ (mimicking the usual convention for locally stationary time series of Dahlhaus [8]). The local mean density and Bartlett spectrum can then be estimated at corresponding absolute times $u_0 \in (0, 1)$ and the restrictions on b_1, b_2 and T read as follows. In (10), we must have $u_0 + b_1 \operatorname{Supp}(w) \subseteq [0, 1]$, and in (13), we must have $u_0 + b_1 \operatorname{Supp}(w) + (T b_2)^{-1} \operatorname{Supp}(q) \subseteq [0, 1]$. These two support conditions are always satisfied, eventually as $T \to \infty$, provided that the kernels w and q are compactly supported and that $b_1 \to 0$ and $T b_2 \to \infty$.

In the sequel, we will show that this is a sensible estimator of $\gamma_{b_2}^{<\mathrm{LS}>}(u_0; \omega_0)$ sharing the usual properties of a nonparametric estimator constructed via kernel-smoothing over time and frequency: for sufficiently small bandwidths b_1 in time and b_2 in frequency this estimator becomes well localised around $(u_0; \omega_0)$.

The main results stated hereafter provide asymptotic expansions of its bias and variance behaviour, leading to consistency of this estimator under some asymptotic condition for b_1 and b_2 as $T \to \infty$.

3. Bias and variance bounds

3.1. Main assumptions

The first assumption is akin but stronger than condition (6) above, being in fact equal to assumption (LS-1) of Roueff, von Sachs and Sansonnet [21]. It guarantees that, for all $T \geq 1$, the locally stationary Hawkes process N_T admits a (causal) local fertility function $s \mapsto p^{<\mathrm{LS}>}(s; u)$ which is not only uniformly bounded, but has an exponentially decaying memory (as a function in the first argument, uniformly with respect to its second argument).

(LS-1) Assume that

$$\left| \lambda_c^{<\mathrm{LS}>} \right|_\infty < \infty. \tag{15}$$

Assume moreover that for all $u \in \mathbb{R}$, $p^{<\mathrm{LS}>}(\cdot; u)$ is supported on $\mathbb{R}_+$ and that there exists a $d > 0$ such that $\zeta_1^{<\mathrm{LS}>}(d) < 1$ and $\zeta_\infty^{<\mathrm{LS}>}(d) < \infty$ where

$$\zeta_1^{<\mathrm{LS}>}(d) := \sup_{u \in \mathbb{R}} \left| p^{<\mathrm{LS}>}(\cdot; u) \right|_{d, 1} = \sup_{u \in \mathbb{R}} \int p^{<\mathrm{LS}>}(s; u) \mathrm{e}^{d|s|} \, \mathrm{d}s \tag{16}$$

and

$$\zeta_\infty^{<\mathrm{LS}>}(d) := \sup_{u \in \mathbb{R}} \left| p^{<\mathrm{LS}>}(\cdot; u) \right|_{d, \infty} = \sup_{u \in \mathbb{R}} \operatorname{ess\,sup}_{s \in \mathbb{R}} \left\{ \left| p^{<\mathrm{LS}>}(s; u) \right| \mathrm{e}^{d|s|} \right\}. \tag{17}$$

All the examples considered in Roueff, von Sachs and Sansonnet [21] satisfy this condition. It is also satisfied if the local fertility functions have a (uniformly) bounded compact support (cf. Hansen, Reynaud-Bouret and Rivoirard [14] in the stationary case).

(LS-1) Assume that, for some $\beta \in (0, 1]$,

$$\xi_c^{(\beta)} := \sup_{u \neq v} \frac{|\lambda_c^{<\mathrm{LS}>}(v) - \lambda_c^{<\mathrm{LS}>}(u)|}{|v - u|^\beta} < \infty.$$

(LS-2) Assume that, for some $\beta \in (0, 1]$, $|\xi^{(\beta)}|_1 < \infty$, where

$$\xi^{(\beta)}(r) := \sup_{u \neq v} \frac{|p^{<\mathrm{LS}>}(r; v) - p^{<\mathrm{LS}>}(r; u)|}{|v - u|^\beta}.$$

Assumptions (LS-1) and (LS-2) can be interpreted as smoothness conditions respectively on $\lambda_c^{<\mathrm{LS}>}$ and on $p^{<\mathrm{LS}>}(\cdot; \cdot)$ with respect to its second argument. Note also that Assumptions (16) and (17) imply in particular Assumption (LS-4) of Roueff, von Sachs and Sansonnet [21] which we recall here to be

$$\zeta_\infty^{<\mathrm{LS}>} := \zeta_\infty^{<\mathrm{LS}>}(0) = \sup_{u \in \mathbb{R}} |p^{<\mathrm{LS}>}(\cdot; u)|_\infty < \infty, \tag{18}$$

$$\zeta_{(\beta)}^{<\mathrm{LS}>} := \sup_{u \in \mathbb{R}} |p^{<\mathrm{LS}>}(\cdot; u)|_{(\beta), 1} < \infty. \tag{19}$$

This can be seen simply by noting that $\zeta_1^{<\mathrm{LS}>} := \zeta_1^{<\mathrm{LS}>}(0) \leq \zeta_1^{<\mathrm{LS}>}(d)$ and $\zeta_\infty^{<\mathrm{LS}>} \leq \zeta_\infty^{<\mathrm{LS}>}(d)$ for all $d \geq 0$, with equality for $d = 0$. Similarly, $\zeta_1^{<\mathrm{LS}>}(d) < \infty$ for some $d > 0$ implies $\zeta_{(\beta)}^{<\mathrm{LS}>} < \infty$ for all $\beta > 0$.

Hereafter all the given bounds are uniform upper bounds in the sense that they hold uniformly over parameters $\lambda_c^{<\mathrm{LS}>}$ and $p^{<\mathrm{LS}>}$ satisfying the set of conditions (6), (LS-1), (LS-2), and (18) and (19), as in Theorems 1, or the more restrictive set of conditions (LS-1), (LS-1) and (LS-2), as in Theorem 2 and 3. More specifically, we use the following conventions all along the paper.

Convention 1 (Symbol $\lesssim$). For two nonnegative sequences a_T and b_T indexed by $T \geq 1$, possibly depending on parameters $\lambda_c^{<\mathrm{LS}>}$, $p^{<\mathrm{LS}>}$, b_1 and b_2, we use the notation $a_T \lesssim b_T$ to denote that there exists a constant C such that, for all b_1, b_2 and T satisfying certain conditions $\mathcal{C}(b_1, b_2, T)$, we have $a_T \leq Cb_T$ with C only depending on non-asymptotic quantities and constants such as d, β, $\zeta_1^{<\mathrm{LS}>}(d)$, $\zeta_{(\beta)}^{<\mathrm{LS}>}$, $\zeta_\infty^{<\mathrm{LS}>}(d)$, $\xi_c^{(\beta)}$, $|\xi^{(\beta)}|_1$, $|\lambda_c^{<\mathrm{LS}>}|_\infty$ and the two kernel functions w and q.

The conditions $\mathcal{C}(b_1, b_2, T)$ will be intersections of the following ones:

$$T \geq 1 \quad \text{and} \quad b_1 \in (0, 1], \tag{20}$$

$$b_1, b_2 \in (0, 1] \quad \text{and} \quad Tb_1b_2 \geq 1, \tag{21}$$

$$b_1 \ln(T) \leq 1. \tag{22}$$

Convention 2 (Constants A_1, A_2). We use A_1, A_2 to denote positive constants that can change from one expression to another but always satisfy $A_1^{-1} \lesssim 1$ and $A_2 \lesssim 1$, using Convention 1. In other words A_1 and A_2 are positive constants which can be bounded from below and from above, respectively, using the constants appearing in the assumptions and the chosen kernels w and q.

Convention 2 will be useful to treat exponential terms in a simplified way, that is, without considering unnecessary constants; for instance, we can write $(e^{-A_1 T})^2 \le e^{-A_1 T}$ replacing $2A_1$ by A_1 in the second expression without affecting the property $A_1^{-1} \lesssim 1$ required on A_1.

3.2. Main results

We can now state the main results of this contribution, whose proofs can be found in Appendix B. For the bias and variance of the local mean density estimator $\widehat{m}_{b_1}(u_0)$, we establish the following result.

Theorem 1. *Let the kernel w satisfy* (K-1). *Assume conditions* (6), (LS-1), (LS-2), *and* (18) *and* (19) *to hold. Then, for b_1 and T satisfying* (20), *the bias of the local density estimator satisfies, for all $u_0 \in \mathbb{R}$,*

$$\left| \mathbb{E}\big[\widehat{m}_{b_1}(u_0)\big] - m_1^{<\mathrm{LS}>}(u_0) \right| \lesssim b_1^{\beta} + T^{-\beta}. \tag{23}$$

If moreover (LS-1) *holds, its variance satisfies*

$$\mathrm{Var}\big(\widehat{m}_{b_1}(u_0)\big) \lesssim (Tb_1)^{-1}. \tag{24}$$

Hence, $\widehat{m}_{b_1}(u_0)$ is shown to be a (mean-square) consistent estimator of $m_1^{<\mathrm{LS}>}(u_0)$, and, optimizing the bias and variance bounds, we get the "usual" mean-square error rate $T^{-\frac{2\beta}{2\beta+1}}$ for nonparametric curve estimation with an additive noise structure, achieved for a bandwidth $b_1 \sim T^{-\frac{1}{2\beta+1}}$.

We now treat the bias of the estimator $\widehat{\gamma}_{b_2, b_1}(u_0; \omega_0)$ which can be decomposed as the sum of 1) a bias in the time direction, namely, $\mathbb{E}\widehat{\gamma}_{b_2, b_1}(u_0; \omega_0) - \gamma_{b_2}^{<\mathrm{LS}>}(u_0; \omega_0)$ and 2) a bias in the frequency direction, namely, $\gamma_{b_2}^{<\mathrm{LS}>}(u_0; \omega_0) - \gamma^{<\mathrm{LS}>}(u_0; \omega_0)$.

Theorem 2. *Let the kernels w and q satisfy* (K-1) *and* (K-2). *Assume conditions* (LS-1), (LS-1), *and* (LS-2) *to hold. Then, for all b_1, b_2 and T satisfying* (21) *and* (22) *and for all $u_0, \omega_0 \in \mathbb{R}$, we have*

$$\left| \mathbb{E}\big[\widehat{\gamma}_{b_2, b_1}(u_0; \omega_0)\big] - \gamma_{b_2}^{<\mathrm{LS}>}(u_0; \omega_0) \right| \lesssim b_1^{\beta} + b_1^{2\beta} b_2^{-1} + (Tb_1 b_2)^{-1}. \tag{25}$$

If moreover the squared modulus $|Q(\omega)|^2$ of the Fourier transform of the kernel q satisfies

$$\int \omega^2 |Q(\omega)|^2 \, \mathrm{d}\omega < \infty \quad and \quad \int \omega |Q(\omega)|^2 \, \mathrm{d}\omega = 0, \tag{26}$$

the "bias in frequency direction" fulfills for $b_2 \in (0, 1]$,

$$\left|\gamma_{b_2}^{<\text{LS}>}(u_0; \omega_0) - \gamma^{<\text{LS}>}(u_0; \omega_0)\right| \lesssim b_2^2. \tag{27}$$

Remark 2. Condition (26) is automatically satisfied if q is compactly supported, real valued and even, and admits an L^2 derivative, such as the triangle shape kernel.

We already observe here that for the estimator $\widehat{\gamma}_{b_2, b_1}(u_0; \omega_0)$ to be asymptotically unbiased, equations (25) and (27) require the following conditions on the choice of the two bandwidths b_1 and b_2 to be fulfilled:

$$Tb_1 b_2 \to \infty, \qquad b_2 \to 0 \quad \text{and} \quad b_2^{-1} b_1^{2\beta} \to 0.$$

Note in particular that these conditions for an asymptotically unbiased estimator imply those required for the feasibility of the estimator in Remark 1 ($b_1 \to 0$ and $Tb_2 \to \infty$).

We shall discuss possible compatible bandwidth choices below, following the treatment of the variance of this estimator.

Theorem 3. *Let the kernels w and q satisfy* (K-1) *and* (K-2). *Assume conditions* (LS-1), (LS-1) *and* (LS-2) *to hold. Then, for all b_1, b_2, T satisfying* (21), *and for all $u_0, \omega_0 \in \mathbb{R}$, we have*

$$\text{Var}\big(\widehat{\gamma}_{b_2, b_1}(u_0; \omega_0)\big) \lesssim (Tb_1 b_2)^{-1} + b_1^{2\beta}\big(b_1^{\beta} b_2^{-1}\big)^2. \tag{28}$$

3.3. Immediate consequences and related works

In order to optimize bandwidth choices in time and in frequency to derive an optimal rate of MSE-consistency of the estimator $\widehat{\gamma}_{b_2, b_1}(u_0; \omega_0)$ for a given time-frequency point $(u_0; \omega_0)$ we observe, by equations (25), (27) and (28), that the MSE satisfies

$$\text{MSE}\big(\widehat{\gamma}_{b_2, b_1}(u_0; \omega_0)\big) \lesssim b_2^4 + b_1^{2\beta} + (Tb_1 b_2)^{-1} + \big(b_1^{2\beta} b_2^{-1}\big)^2.$$

The MSE-rate $T^{-\frac{4\beta}{5\beta+2}}$ is achieved by optimizing this upper bound, that is, by imposing $b_2^4 \sim b_1^{2\beta} \sim (Tb_1 b_2)^{-1}$ (leading to $(b_1^{2\beta} b_2^{-1})^2 \sim b_2^6$), and hence by setting $b_1 \sim T^{-\frac{2}{2+5\beta}}$ and $b_2 \sim T^{-\frac{\beta}{2+5\beta}}$. The fact that this rate bound is obtained by balancing the two squared bias terms $b_1^{2\beta}$ and b_2^4 with the variance term $(Tb_1 b_2)^{-1}$ indicates that all the other terms appearing in the upper bounds (25) and (28) are negligible. Thus, since the bias terms b_1^{β} and b_2^2 and the variance term $(Tb_1 b_2)^{-1}$ correspond to the usual bias and variance rates of a kernel estimator of a local spectral density estimator (see Dahlhaus [10], Example 4.2, for locally stationary linear time series), it is clear that the obtained rate is sharp for this moment estimator under our assumptions. Note also that the MSE-rate $T^{-\frac{4\beta}{5\beta+2}}$ corresponds to the minimax lower bound for evolutionary spectrum estimation established in Neumann and von Sachs [16], Theorem 2.1, in the Sobolev space $W_{\infty,\infty}^{\beta,2}$. Although insightful, the comparison is not completely rigorous as their model for

establishing this lower bound is a benchmark for a class of non-stationary time series and the MSE they consider is integrated. Therefore, a particularly interesting problem for future work would be to derive (hopefully large) classes of locally stationary point processes on which our estimator achieves the minimax rate.

In the same line of thoughts about the performance of our kernel estimator, the question naturally arises about the data-driven choice of the bandwidths b_1 and b_2. This question of bandwidth selection in the context of locally stationary time series has been addressed only recently, see Giraud, Roueff and Sanchez-Perez [13], Richter and Dahlhaus [19] for adaptive prediction and parameter curve estimation, respectively. The problem of adaptive kernel estimation of the local spectral density for locally stationary time series has been more specifically addressed in van Delft and Eichler [22], where a practical approach is derived and studied. It is mainly based on the central limit theorem established in Dahlhaus [10], Example 4.2, for this estimator. This methodology could be adapted to the case of locally stationary point processes. A first step in this direction would be to establish a central limit theorem for our estimator $\widehat{\gamma}_{b_2,b_1}(u_0, \omega_0)$ at a given time-frequency point, which is also left for future work. Note also that, in practical time frequency analysis, the bandwidths b_1 and b_2 are often chosen having in mind a physical interpretation. For instance, in our real data example of Roueff and von Sachs [20], Section 6, on each day, transaction data is collected between 9:00 AM and 5:30 PM, hence over 8.5 hours. Our choice $b_1 = 0.15$ and $b_2 = 0.005$ corresponds to saying that we consider finance transactions data as roughly stationary over $8.5 \times 0.15 \approx 1$ hour and 16 minutes of time, and that the spectrum obtained from such data has maximal frequency resolution 0.005 Hz (we may distinguish between two periodic behaviors present in the data only if their frequencies differ by at least this value).

To conclude this section, let us discuss whether our approach could be used for the parameter estimation of locally stationary Hawkes processes. In a fully non-parametric approach, one would be interested in estimating the two (possibly smooth or sparse) unknown functions that are the baseline intensity function $u \mapsto \lambda_c^{<\mathrm{LS}>}(u)$ and the local fertility function $(s, u) \mapsto p^{<\mathrm{LS}>}(s; u)$. In a parametric approach, one would assume these functions to depend on an unknown finite dimensional parameter θ in a (known) form $u \mapsto \lambda_c^{<\mathrm{LS}>}(u|\theta)$ and $(s, u) \mapsto p^{<\mathrm{LS}>}(s; u|\theta)$, and try to estimate θ. An intermediate approach, proposed in Roueff, von Sachs and Sansonnet [21], Section 5.1, to derive simple examples of locally stationary Hawkes processes, is to consider a parametric stationary model for the fertility function, say $s \mapsto p^{<\mathrm{S}>}(s|\theta)$ for $\theta \in \Theta \subset \mathbb{R}^d$, and to deduce a local one of the form $(s, u) \mapsto p^{<\mathrm{LS}>}(s; u) = p^{<\mathrm{S}>}(s|\boldsymbol{\theta}(u))$, where now $\boldsymbol{\theta}$ is a Θ-valued function of the absolute time. In all these cases, the estimators $\widehat{m}_{b_1}$ and $\widehat{\gamma}_{b_2,b_1}$ could be used as empirical moments to estimate $\lambda_c^{<\mathrm{LS}>}$, $p^{<\mathrm{LS}>}$, θ, or the curve $\boldsymbol{\theta}$. Since our estimators are consistent, this method would in principle work whenever the unknown quantities to estimate can be deduced from the local mean density $m_1^{<\mathrm{LS}>}$ and local Bartlett spectrum $\gamma^{<\mathrm{LS}>}$ through Relations (7) and (8), respectively. More direct methods to estimate the parameters of interest for non-stationary Hawkes processes have been proposed in Chen and Hall [5], Chen and Hall [6], Mammen [15]. In the first two references, only the baseline intensity is time varying, and a different asymptotic setting is considered, where this baseline intensity tends to infinity through a multiplicative constant. In the fully non-parametric case, an identifiability problem is pointed out in Chen and Hall [6], Section 2.2. We do not have this problem in our asymptotic scheme, since (7) and (8) show that the base line intensity $\lambda_c^{<\mathrm{LS}>}(u)$ can be completely identified from

the local mean density $m_1^{<\text{LS}>}(u)$ and the local Bartlett spectrum $\gamma^{<\text{LS}>}(u; \cdot)$ alone using the formula

$$\lambda_c^{<\text{LS}>}(u) = m_1^{<\text{LS}>}(u)\left(\frac{m_1^{<\text{LS}>}(u)}{2\pi \gamma^{<\text{LS}>}(u; 0)}\right)^{1/2}.$$

The model considered in Mammen [15] is a multivariate version of the locally stationary Hawkes process with the same asymptotic setting as ours, and additional assumptions on the (multivariate) fertility function. Both the baseline intensity and the (time varying) fertility function are estimated in a non-parametric fashion using a direct method based on localized mean square regression and a decomposition of the local fertility function on a B-spline base. Such methods should be more efficient than using the local mean density and Bartlett spectrum to build moment estimators of these parameters, since they rely on the intrinsic auto-regression structure of the underlying process. In contrast, as far as the time-frequency analysis is concerned, which is the main focus of our contribution, the estimator that we propose should be relevant to estimate the local Bartlett spectrum beyond the case of locally stationary Hawkes processes, namely, for any locally stationary point process for which the general formula (11) and (12) make sense.

3.4. Main ideas of the proofs

3.4.1. *Local approximations of moments*

An essential step for treating the bias terms is to be able to approximate, as $T \to \infty$, in the neighborhood of uT, the first and second moments of N_T by that of the local stationary approximation $N(\cdot; u)$ defined as in Section 2.3. We first state the two approximations that directly follow Roueff, von Sachs and Sansonnet [21], Theorem 4, with $m = 1, 2$.

Theorem 4. *Let $\beta \in (0, 1]$ and let $(N_T)_{T \geq 1}$ be a locally stationary Hawkes process satisfying conditions (6), (LS-1), (LS-2), (18) and (19). Let $(N(\cdot; u))_{u \in \mathbb{R}}$ be the collection of stationary Hawkes process defined as in Section 2.3. Then, for all $T \geq 1$, $u \in \mathbb{R}$ and all bounded integrable functions g, we have*

$$\left|\mathbb{E}[N_T(S^{-Tu}g)] - \mathbb{E}[N(g; u)]\right| \lesssim (|g|_1 + |g|_{(\beta),1})T^{-\beta}, \tag{29}$$

$$\left|\text{Var}(N_T(S^{-Tu}g)) - \text{Var}(N(g; u))\right| \lesssim (|g|_1 + |g|_\infty)(|g|_1 + |g|_{(\beta),1})T^{-\beta}. \tag{30}$$

The control of the bias in Theorem 1 directly follows from (29). However it turns out that (30) is not sharp enough to control the bias of the local Bartlett spectrum and thus to obtain the expected convergence rate. The basic reason is that it involves L^1 (weighted) norms of g in the upper bound instead of L^2 norms. In order to recover the correct rates of convergence, we rely on the following new result where the L^1 (weighted) norms are indeed replaced by L^2 (weighted) norms, or, to be more precise, where the remaining weighted L^1-norms are compensated by an exponentially decreasing term in T, and will thus turn out to be negligible.

Theorem 5. *Let $\beta \in (0, 1]$ and let $(N_T)_{T \geq 1}$ be a locally stationary Hawkes process satisfying conditions (LS-1), (LS-1) and (LS-2). Let $(N(\cdot; u))_{u \in \mathbb{R}}$ be the collection of stationary Hawkes*

process defined as in Section 2.3. *Then, for all bounded and compactly supported functions* g, *for all* $T \geq 1$ *and* $u \in \mathbb{R}$,

$$\left| \mathrm{Var}\big(N_T(g)\big) \right| \lesssim |g|_2^2 + \mathrm{e}^{-A_1 T} |g|_{d,1}^2, \tag{31}$$

$$\left| \mathrm{Var}\big(N_T\big(S^{-Tu}g\big)\big) - \mathrm{Var}\big(N(g;u)\big) \right|$$
$$\lesssim \left\{ |g|_2^2 + \mathrm{e}^{-A_1 T} |g|_{d,1}^2 + |g|_{(\beta),2} \big(|g|_2 + \mathrm{e}^{-A_1 T} |g|_{d,1} \big) \right\} T^{-\beta}. \tag{32}$$

The proof of this theorem can be found in Roueff and von Sachs [20], Section 4. To obtain this new result, we crucially rely on the assumption (LS-1) where controls in exponentially weighted norms (based on $\sup_u \left| p^{<\mathrm{LS}>}(\cdot;u)) \right|_{d,q}$ for $q = 1, \infty$ and some $d > 0$) are assumed to strengthen the assumptions (6), (18) and (19).

3.4.2. *Bias and approximate centering*

To control all the error terms, we found useful to introduce a centered version of N_T with a centering term corresponding to its asymptotic deterministic version. Recalling that N_T behaves in a neighborhood of Tu_0 as $N(\cdot;u_0)$ and that this process admits the mean intensity denoted by $m_1^{<\mathrm{LS}>}(u_0)$, we define, for any test function f,

$$\overline{N}_T(f;u_0) := N_T(f) - \mathbb{E}\big[N(f;u_0)\big] = \int f(s)\big[N_T(\mathrm{d}s) - m_1^{<\mathrm{LS}>}(u_0)\,\mathrm{d}s\big]. \tag{33}$$

It is important to note that this "approximate" centering depends on an absolute location u_0 as it is a good approximation of $\mathbb{E}[N_T(f)]$ only for f localized in a neighborhood of Tu_0. Let us apply this definition. By (10), since w_{Tb_1} integrates to 1, the error of the local mean density estimator can directly be expressed as

$$\widehat{m}_{b_1}(u_0) - m_1^{<\mathrm{LS}>}(u_0) = \overline{N}_T\big(S^{-Tu_0} w_{Tb_1}; u_0\big). \tag{34}$$

Hence controlling the bias of this estimator directly amounts to evaluating the quality of the above centering.

The treatment of the bias of the local Bartlett spectrum is a bit more involved since, as often for spectral estimators, the empirical centering term requires a specific attention. This term appears inside the negated square modulus of the right-hand side of (13). To see why it is indeed a centering term, observe that, using $\int w_{Tb_1} = 1$, we can write

$$\widehat{\gamma}_{b_2,b_1}(u_0;\omega_0) = \widehat{E}\big[\rho\big(N_T\big(S^{-Tu_0} q_{\omega_0,b_2}\big)\big); w_{b_1}\big]$$

with now ρ defined, for any test function f, as the "centered" squared modulus

$$\rho\big(N_T(f)\big) = \big|N_T(f) - \widehat{E}\big[N_T(f); w_{b_1}\big]\big|^2.$$

Using that $\int w_{b_1} = 1$, the centering in (33) can be introduced within this definition of ρ, leading to

$$\rho\big(N_T(f)\big) = \big|\overline{N}_T(f;u_0) - \widehat{E}\big[\overline{N}_T(f;u_0); w_{b_1}\big]\big|^2.$$

By comparison with the previous expression of $\widehat{\gamma}_{b_2,b_1}(u_0;\omega_0)$, we easily get an expression of the local Bartlett spectrum estimator based on this centered version of N_T, namely,

$$\widehat{\gamma}_{b_2,b_1}(u_0;\omega_0) = \widehat{E}\big[|\overline{N}_T(f;u_0)|^2; w_{b_1}\big] - \big|\widehat{E}[\overline{N}_T(f;u_0); w_{b_1}]\big|^2, \tag{35}$$

where analogously to (14), we denote, for the test function $f = S^{-Tu_0}q_{\omega_0,b_2}$,

$$\widehat{E}\big(|\overline{N}_T(f;u_0)|^2; w_{b_1}\big) := \int |\overline{N}_T\big(f(\cdot-t);u_0\big)|^2 w_{Tb_1}(t)\,dt, \tag{36}$$

$$\widehat{E}\big(\overline{N}_T(f;u_0); w_{b_1}\big) := \int \overline{N}_T\big(f(\cdot-t);u_0\big) w_{Tb_1}(t)\,dt. \tag{37}$$

In fact, using $\int w = 1$, f integrable and interchanging the order of integration, we immediately get the simplification

$$\widehat{E}\big(\overline{N}_T(f;u_0); w_{b_1}\big) = \overline{N}_T(f * w_{Tb_1};u_0), \tag{37}$$

where we used the convolution product $*$. The advantage of the new expression (35) in contrast to the original (13) is that now we expect the negated square modulus to be of negligible order. To see why, consider for instance the bias in (25), for the control of which we need to bound, as the second term of (35),

$$\mathbb{E}\big[|\widehat{E}\big(\overline{N}_T(f;u_0); w_{b_1}\big)|^2\big] = \mathrm{Var}\big(N_T(f*w_{Tb_1})\big) + \big|\mathbb{E}[\overline{N}_T(f*w_{Tb_1};u_0)]\big|^2. \tag{38}$$

(Here and in the following we repeatedly use the fact that N_T and $\overline{N}_T(\cdot;u_0)$ have the same variance). Finally, the control of the bias term in (25) now requires to evaluate

$$\mathbb{E}\big[\widehat{E}\big[|\overline{N}_T\big(S^{-Tu_0}q_{\omega_0,b_2};u_0\big)|^2; w_{b_1}\big]\big] - \gamma_{b_2}^{<\mathrm{LS}>}(u_0;\omega_0),$$

which is again decomposed as a main term

$$\int \big(\mathrm{Var}\big(N_T\big(S^{-Tu_0}q_{\omega_0,b_2}(\cdot-t)\big)\big) - \gamma_{b_2}^{<\mathrm{LS}>}(u_0;\omega_0)\big) w_{Tb_1}(t)\,dt, \tag{39}$$

added to a negligible (because involving a squared bias) term

$$\int \big|\mathbb{E}[\overline{N}_T\big(S^{-Tu_0}q_{\omega_0,b_2}(\cdot-t);u_0\big)]\big|^2 w_{Tb_1}(t)\,dt. \tag{40}$$

3.4.3. *Variance terms and exact centering*

The variance of the local mean density estimator directly requires to control the variance of $N_T(f)$ for given test functions f. This requires new deviation bounds for non-stationary Hawkes processes. By deviation bounds, we here mean that we bound the moments of

$$\overline{\overline{N}}_T(h) := N_T(h) - \mathbb{E}\big[N_T(h)\big], \tag{41}$$

where h is an appropriate test function. New results in this direction are gathered in Section 4, where the dependence structure of non-stationary Hawkes processes is investigated, leading to the appropriate control of such moments in Proposition 9. For instance, the moment of order 2 directly provides the adequate bound for the variance of the local mean density estimator

$$\text{Var}\big(\widehat{m}_{b_1}(u_0)\big) = \text{Var}\big(N_T\big(S^{-Tu_0}w_{Tb_1}\big)\big) = \big\|\overline{\overline{N}}_T(h)\big\|_2^2, \tag{42}$$

where we use the notation introduced in (3).

Now we turn our attention to the estimator of the local Bartlett spectrum. The control of the moments of $\overline{\overline{N}}_T$ will essentially be used to approximate $\widehat{\gamma}_{b_2,b_1}(u_0;\omega_0)$ by

$$\widetilde{\gamma}_{b_2,b_1}(u_0;\omega_0) := \int \big|\overline{\overline{N}}_T\big(S^{-Tu_0}q_{\omega_0,b_2}(\cdot - t)\big)\big|^2 w_{Tb_1}(t)\,\mathrm{d}t. \tag{43}$$

In contrast to the centering used in $\overline{N}_T(\cdot;u)$ for controlling the bias (a centering with respect to $\mathbb{E}[N(h;u)]$) for some absolute location u, here the term $\mathbb{E}[N_T(h)]$ is no longer invariant as h is shifted. This is why this centering cannot be used as a direct decomposition of estimator $\widehat{\gamma}_{b_2,b_1}(u_0;\omega_0)$ as in (35). Instead, we use a bound on the error of approximating $\widehat{\gamma}_{b_2,b_1}(u_0;\omega_0)$ by $\widetilde{\gamma}_{b_2,b_1}(u_0;\omega_0)$, see Lemma 15.

Finally the variance of the local Bartlett estimator is obtained by controlling the variance of $\widetilde{\gamma}_{b_2,b_1}(u_0;\omega_0)$ (Lemma 14), which in turn relies on a bound of

$$\text{Cov}\big(\big|\overline{\overline{N}}_T(h_1)\big|^2, \big|\overline{\overline{N}}_T(h_2)\big|^2\big)$$

for test functions h_1 and h_2, which is derived in Corollary 10.

4. Deviation bounds for non-stationary Hawkes processes

We now derive new results required for treating the variance of the studied estimators. In contrast to Poisson processes, we can not rely on the independence of the process on disjoint sets. To the best of our knowledge, the most advanced results on deviation bounds of Hawkes processes are to be found in Reynaud-Bouret and Roy [18] and only apply to stationary Hawkes processes with compactly supported fertility functions. Here we consider non-stationary Hawkes processes with exponentially decreasing local fertility functions. The generalization to non-stationary processes requires a specific approach rather than a mere adaption of Reynaud-Bouret and Roy [18].

Although we are here motivated by the study of the variance of the local mean density and Bartlett estimators, we believe that the results contained in this section are of broader interest, as they can serve more generally to understand the dependence structure of non-stationary Hawkes processes.

4.1. Setting

Recall that (5) is our minimal condition for defining the non-stationary Hawkes process N with immigrant intensity function $\lambda_c(\cdot)$ and varying fertility function $p(\cdot;\cdot)$. The cluster construction

of a Hawkes process relies on conditioning on the realization of a so-called *center process*, N_c a Poisson point process (PPP) with intensity λ_c on $\mathbb{R}$, which describes spontaneous appearing of the immigrants. At each *center point* t^c of N_c, a point process $N(\cdot|t^c)$ is generated as a branching process whose conditional distribution given N_c only depends on t^c and is entirely determined by the time varying fertility function $p(\cdot; \cdot)$. This distribution is described through its conditional Laplace functional in Roueff, von Sachs and Sansonnet [21], Section 6.1, under the additional assumption

$$\zeta_\infty < \infty \qquad \text{with } \zeta_q = \sup_{t \in \mathbb{R}} \big|p(\cdot; t)\big|_q \text{ for all } q \in [1, \infty]. \tag{44}$$

The following result directly follows from these derivations. A detailed proof is available in Roueff and von Sachs [20], Section 3, for sake of completeness.

Proposition 6. *Suppose that* (5) *and* (44) *hold and set*

$$r_1 = \frac{-\log \zeta_1}{2(1 - \zeta_1^{1/2} + \zeta_\infty \zeta_1^{-1/2})}. \tag{45}$$

Then, for all $h : \mathbb{R} \to \mathbb{R}$ *satisfying* $|h|_1 \leq 1$ *and* $|h|_\infty \leq 1$,

$$\mathbb{E}\big[e^{(1-\zeta_1^{1/2})r_1|N(h)|}\big] \leq \exp\big(|\lambda_c|_\infty \zeta_1^{-1/2} r_1\big). \tag{46}$$

Consequently, for all $q > 0$, *there exists a positive constant* B_q *only depending on* $|\lambda_c|_\infty$, ζ_1 *and* ζ_∞ *such that, for all* $h : \mathbb{R} \to \mathbb{R}$ *satisfying* $|h|_1 \leq 1$ *and* $|h|_\infty \leq 1$,

$$\big\|N(h)\big\|_q \leq B_q. \tag{47}$$

The moment bound (47) will be useful but far from sufficient to bound the variance of our estimators efficiently. To see why, let us suppose temporarily that N is a homogeneous Poisson process with unit rate and consider $h = n^{-1}\mathbb{1}_{[0,n]} - 1$ for some positive integer n. Then the above bound with $p = 2$ gives $\text{Var}(N(h)) \leq B_2^2$ although we know that in this very special case we have $\text{Var}(N(h)) = n^{-1}$, hence the bound is not sharp at all as $n \to \infty$. In the following we provide new deviation bounds applying to non-stationary Hawkes processes which allows one to recover the expected order of magnitude for the variance. To this end, we rely on stronger conditions than the ones used in Roueff, von Sachs and Sansonnet [21].

Define moreover, using the exponentially weighted norm notation in (2), for all $d \geq 0$, and $q \in [1, \infty]$,

$$\zeta_q(d) = \sup_{t \in \mathbb{R}} \big|p(\cdot; t)\big|_q.$$

We strengthen the basic conditions (6) and (44) into the following one.

(NS-1) We have $|\lambda_c|_\infty < \infty$. Moreover, for all $t \in \mathbb{R}$, $p(\cdot; t)$ is supported on $\mathbb{R}_+$ and there exists $d > 0$ such that $\zeta_1(d) < 1$ and $\zeta_\infty(d) < \infty$.

The version of (NS-1) in the locally stationary case is (LS-1) in the sense that the locally stationary Hawkes process $(N_T)_{T \geq 1}$ satisfies (LS-1) if and only if, for all $T \geq 1$, the non-stationary Hawkes process N_T satisfied (NS-1). Therefore all the results below relying on (NS-1) apply to the locally stationary Hawkes processes satisfying (LS-1). We recall that this assumption means that the local fertility functions satisfy some uniform exponential decreasing.

4.2. New deviation bounds

The deviations bounds are based on the following exponential bound control on component point processes $N(\cdot|t^c)$ defined above.

Proposition 7. *Suppose that* (NS-1) *holds for some $d > 0$. Then we have, for all $a \in (0, d)$, for all $t^c \in \mathbb{R}$,*

$$\mathbb{E}\left[\int_{[t,\infty)} e^{a(s-t)} N(ds|t^c)\right] \leq C_a \qquad \text{with } C_a := 1 + \frac{\zeta_\infty(d)}{(d-a)(1-\zeta_1(d))}. \tag{48}$$

Proof. Let $g : \mathbb{R} \to \mathbb{R}_+$ and define, for all $h : \mathbb{R} \to \mathbb{R}_+$,

$$\left[\mathcal{E}_g(h)\right](s) = g(s) + \int h(u) p(u - s; u) \, du.$$

Following Roueff, von Sachs and Sansonnet [21], Proposition 7 and Eq. (34), we have that, for all $t \in \mathbb{R}$,

$$\mathbb{E}\left[N\left(g|t^c\right)\right] = \lim_{n \to \infty} \left[\mathcal{E}_g^n(g)\right](t), \tag{49}$$

where $\mathcal{E}_g^n$ denotes the nth self-composition of $\mathcal{E}_g$. In the following, we take some $a \in (0, d)$ and set

$$g_t(s) = e^{a(s-t)} \mathbb{1}_{[t,\infty)}(s) = g_0(s - t).$$

Observe that, using that $p(; u)$ is supported on $[0, \infty)$ under (NS-1), for all $h : \mathbb{R} \to \mathbb{R}_+$,

$$\left[\mathcal{E}_{g_t}(h)\right](t) \leq g_t(t) + \int_t^\infty h(u) p(u - t; u) \, du$$

$$\leq 1 + \int_t^\infty h(u) e^{d(t-u)} e^{d(u-t)} p(u - t; u) \, du$$

$$\leq 1 + \int_t^\infty h(u) e^{d(t-u)} \zeta_\infty(d) \, du$$

$$= 1 + \zeta_\infty(d) \left|h(t + \cdot)\right|_{-d,1}.$$

Applying this to $h = \mathcal{E}_{g_t}^{n-1}(g_t)$ we get, for all $n \geq 1$,

$$\left[\mathcal{E}_{g_t}^n(g_t)\right](t) \leq 1 + \zeta_\infty(d) \left|\left[\mathcal{E}_{g_t}^{n-1}(g_t)\right](t + \cdot)\right|_{-d,1}. \tag{50}$$

Similarly, we also get that, for all $h : \mathbb{R} \to \mathbb{R}_+$,

$$\left|\left[\mathcal{E}_{g_t}(h)\right](t+\cdot)\right|_{-d,1} \leq |g_0|_{-d,1} + \zeta_1(d)\left|h(t+\cdot)\right|_{-d,1}.$$

Observing that $|g_0|_{-d,1} = (d-a)^{-1}$ and iterating the last inequality, we finally obtain that, for all $n \geq 1$,

$$\left|\left[\mathcal{E}^n_{g_t}(g_t)\right](t+\cdot)\right|_{-d,1} \leq \frac{1}{d-a}\left(1 + \zeta_1(d) + \cdots + \zeta_1(d)^n\right) \leq \frac{1}{(d-a)(1-\zeta_1(d))}.$$

Inserting this bound in (50) and letting n go to ∞ as in (49) with $g = g_t$, we get the claimed result. $\qquad\square$

Define the past and future σ-fields at time t respectively, as

$$\mathcal{F}_t = \sigma\left(N_c(A), N\left(B|t^c\right) : A \in \mathbb{B}\left((-\infty, t]\right), B \in \mathbb{B}(\mathbb{R}), t^c \leq t\right)$$
$$\supset \sigma\left(N(A) : A \in \mathbb{B}\left((-\infty, t]\right)\right)$$

and

$$\mathcal{G}_t = \sigma\left(N(A) : A \in \mathbb{B}\left((t, -\infty)\right)\right).$$

The following result provides a uniform exponential control on the time-dependence of N.

Proposition 8. *Suppose that* (NS-1) *holds for some $d > 0$ and that λ_c is bounded. Let $p \in [1, \infty)$, $t < u$ and Y be a centered L^1 $\mathcal{G}_u$-measurable random variable. Then, for all $a \in (0, d)$, for all $q \in (p, \infty]$, if Y is L^q,*

$$\left\|\mathbb{E}[Y|\mathcal{F}_t]\right\|_p \leq \|Y\|_q \left(|\lambda_c|_\infty C_a a^{-1} e^{-a(u-t)}\right)^{1/p - 1/q},$$

where C_a is defined in (48).

Proof. In the following, we denote by t^c_k the points of the Poisson process N_c, that is,

$$N_c = \sum_k \delta_{t^c_k}.$$

Define

$$\Delta_{t^c} = \inf\left\{\delta > 0 : N\left([t^c + \delta, \infty)|t^c\right) = 0\right\}.$$

In other words, Δ_{t^c} is the size of the cluster $N(\cdot|t^c)$, that is the distance between the most left point (which is t^c) and most right point. Since Δ_{t^c} is a point among the points of $N(\cdot|t^c)$), we have

$$e^{a\Delta_{t^c}} \leq \int_{[t,\infty)} e^{a(s-t)} N\left(ds|t^c\right),$$

and a direct consequence of Proposition 7 is that, for all $a \in (0, d)$,

$$\mathbb{E}\big[e^{a\Delta_{t^c}}\big] \leq C_a. \tag{51}$$

Now let us define the position of the last point generated by all clusters started before time t, namely,

$$\overline{\Delta}_t = \sup\big\{t_k^c + \Delta_{t_k^c} : t_k^c \leq t\big\}.$$

We observe that $\overline{\Delta}_t$ is $\mathcal{F}_t$-measurable. Moreover, if $t < u$ and Y is a centered L^1 $\mathcal{G}_u$-measurable random variable, then we have $\mathbb{E}[Y|\mathcal{F}_t] = 0$ on $\{\overline{\Delta}_t < u\}$. The Hölder inequality then yields for $1 \leq p < q \leq \infty$

$$\big\|\mathbb{E}[Y|\mathcal{F}_t]\big\|_p = \big\|\mathbb{E}[Y|\mathcal{F}_t]\mathbb{1}_{\{\overline{\Delta}_t \geq u\}}\big\|_p \leq \big\|\mathbb{E}[Y|\mathcal{F}_t]\big\|_q \big(\mathbb{P}(\overline{\Delta}_t \geq u)\big)^{1/p-1/q}.$$

Since $\|\mathbb{E}[Y|\mathcal{F}_t]\|_q \leq \|Y\|_q$, it only remains to prove that

$$\mathbb{P}(\overline{\Delta}_t \geq u) \leq C_0 e^{-\lambda_0(u-t)}. \tag{52}$$

Observe that $M = \sum_k \delta_{t_k^c, \Delta_{t_k^c}}$ is a marked Poisson point process such that, given N_c, the marks $\Delta_{t_k^c}$ are independent and for each k the conditional distribution of $\Delta_{t_k^c}$ only depends on t_k^c. Hence, M is a Poisson point process with points valued in $\mathbb{R} \times \mathbb{R}_+$ and

$$\{\overline{\Delta}_t \geq u\} = \big\{M\big(\{(t^c, \delta) \in (-\infty, t] \times \mathbb{R}_+ : t_c + \delta \geq u\}\big) > 0\big\}.$$

We thus get that

$$\mathbb{P}(\overline{\Delta}_t \geq u) = 1 - \exp\bigg(-\int_{-\infty}^t \mathbb{P}(t_c + \Delta_{t^c} \geq u)\lambda_c(t^c)\,dt^c\bigg)$$

$$\leq |\lambda_c|_\infty \int_{-\infty}^t \mathbb{P}(\Delta_{t^c} \geq u - t_c)\,dt^c,$$

where we used that $1 - e^{-x} \leq x$ for all $x \geq 0$. Using (51) and the exponential Markov inequality, it follows that

$$\mathbb{P}(\overline{\Delta}_t \geq u) \leq |\lambda_c|_\infty C_a \int_{-\infty}^t e^{a(t^c-u)}\,dt^c = |\lambda_c|_\infty C_a a^{-1} e^{a(t-u)}. \qquad \square$$

We can now derive a Burkhölder-type inequality.

Proposition 9. *Suppose that (NS-1) holds for some $d > 0$. Let $p \in [2, \infty)$. Then there exists a positive constant B_p such that, for all bounded functions h with support included in $[j, j+n]$ for some $j \in \mathbb{Z}$ and $n \in \mathbb{N}$,*

$$\big\|N(h) - \mathbb{E}\big[N(h)\big]\big\|_p \leq A\,|h|_\infty\,\sqrt{n}.$$

where A is a positive constant only depending on d, $|\lambda_c|_\infty$, ζ_1, ζ_∞, $\zeta_\infty(d)$ and $\zeta_1(d)$, for example, for any $a \in (0, d)$ and $q > p$,

$$A := (B_1 + B_p)(B_1 + B_q)\big(|\lambda_c|_\infty \, C_a a^{-1}\big)^{1/p-1/q} \, \frac{e^{-a(1/p-1/q)}}{1 - e^{-a(1/p-1/q)}},$$

where B_p is defined in Proposition 6 and C_a in (48).

Proof. We can assume $|h|_\infty = 1$ without loss of generality. We write

$$h = \sum_{i=1}^{n} h_i \qquad \text{with } h_i = h\mathbb{1}_{[j+i-1, j+i)}.$$

Then $|h_i|_\infty \le 1$ and $|h_i|_1 \le 1$ for all i and, defining $X_i = N(h_i) - \mathbb{E}[N(h_i)]$, from Proposition 6, we have, for all $q \ge 1$,

$$\|X_i\|_q \le B_q + B_1. \tag{53}$$

Then $N(h) - \mathbb{E}[N(h)] = \sum_{i=1}^{n} X_i$ and, applying Dedecker et al. [12], Proposition 5.4, page 123, we have

$$\big\|N(h) - \mathbb{E}\big[N(h)\big]\big\|_p \le \left(2p \sum_{i=1}^{n} b_{i,n}\right)^{1/2}, \tag{54}$$

where, denoting $\mathcal{M}_i = \mathcal{F}_{j+i}$,

$$b_{i,n} = \max_{1 \le \ell \le n} \left\| X_i \sum_{k=i}^{\ell} \mathbb{E}[X_k | \mathcal{M}_i] \right\|_{p/2}.$$

Observing that X_k is centered and $\mathcal{G}_{j+k-1}$-measurable, Proposition 8, gives that, for any $q > p$,

$$\big\|\mathbb{E}[X_k | \mathcal{M}_i]\big\|_p \le \|X_k\|_q \big(|\lambda_c|_\infty \, C_a a^{-1} e^{-a(k-i+1)}\big)^{1/p-1/q}.$$

The Hölder inequality, the last two displays and (53) yield, for all $q > p$,

$$b_{i,n} \le (B_1 + B_p)(B_1 + B_q)\big(|\lambda_c|_\infty \, C_a a^{-1}\big)^{1/p-1/q} \, \frac{e^{-a(1/p-1/q)}}{1 - e^{-a(1/p-1/q)}}.$$

Applying this in (54), we get the claimed bound. $\square$

Another consequence of Proposition 8 is the following useful covariance bound.

Corollary 10. *Suppose that* (NS-1) *holds for some $d > 0$ and that λ_c is bounded. Let h_1 and h_2 be two bounded integrable functions. Let γ satisfy one of the following assertions.*

  (i) *There exist $t \in \mathbb{R}$ such that $\mathrm{Supp}(h_1) \subset (-\infty, t]$ and $\mathrm{Supp}(h_2) \subset [t + \gamma, \infty)$.*
 (ii) *There exist $t \in \mathbb{R}$ such that $\mathrm{Supp}(h_2) \subset (-\infty, t]$ and $\mathrm{Supp}(h_1) \subset [t + \gamma, \infty)$.*
(iii) *$\gamma = 0$.*

Then for all $q > 4$, there exists $C_q > 0$ and $\alpha_q > 0$ both only depending on q, $|\lambda_c|_\infty$, and a and C_a in Proposition 8 such that

$$\left|\mathrm{Cov}\left(\left|\overline{\overline{N}}(h_1)\right|^2, \left|\overline{\overline{N}}(h_2)\right|^2\right)\right| \le C_q \left\|\overline{\overline{N}}(h_1)\right\|_q^2 \left\|\overline{\overline{N}}(h_2)\right\|_q^2 \mathrm{e}^{-\alpha_q \gamma}, \tag{55}$$

where $\overline{\overline{N}}(h) = N(h) - \mathbb{E}[N(h)]$.

Proof. In the case (iii), the bound (55) actually holds with $q = 4$ by the Cauchy–Schwarz inequality and thus also holds with $q > 4$ by Jensen's inequality.

We now consider the case (i) (the last one (ii) being obtained by inverting h_1 and h_2). We have in this case, denoting $Y = |\overline{\overline{N}}(h_2)|^2 - \mathbb{E}[|\overline{\overline{N}}(h_2)|^2]$,

$$\begin{aligned}
\left|\mathrm{Cov}\left(\left|\overline{\overline{N}}(h_1)\right|^2, \left|\overline{\overline{N}}(h_2)\right|^2\right)\right| &= \mathbb{E}\left[\left|\overline{\overline{N}}(h_1)\right|^2 \left(\mathbb{E}[Y|\mathcal{F}_t]\right)\right] \\
&\le \left\|\overline{\overline{N}}(h_1)\right\|_4^2 \left\|\mathbb{E}[Y|\mathcal{F}_t]\right\|_4^2.
\end{aligned}$$

The Jensen Inequality and $q > 4$ give that $\|\overline{\overline{N}}(h_1)\|_4 \le \|\overline{\overline{N}}(h_1)\|_q$ and the proof is concluded by using Proposition 8 to bound $\|\mathbb{E}[Y|\mathcal{F}_t]\|_4$. $\qquad\square$

Appendix A: Useful lemmas

In the following lemmas, we have gathered some simple bounds that we will repeatedly use in the sequel.

Lemma 11. *Let $a, \beta > 0$ and $p \in [1, \infty]$. For any function g and any $t \in \mathbb{R}$, we have*

$$\left|g(\cdot - t)\right|_{(\beta), p} \le 2^{(\beta - 1)+} \left(|g|_{(\beta), p} + |t|^\beta |g|_p\right), \tag{56}$$

$$\left|g(\cdot - t)\right|_{a, 1} \le \mathrm{e}^{a|t|} |g|_{a, 1}. \tag{57}$$

Let $q_{\omega_0, b_2} = b_2^{1/2} \mathrm{e}^{i\omega_0 t} q(b_2 t)$ and $w_{Tb_1} = (Tb_1)^{-1} w(u/(Tb_1))$, where the kernels w and q satisfy $|w|_1 = |q|_2 = 1$. Then, we have, for all $b_1, b_2 \in (0, 1]$ and $T > 0$,

$$|q_{\omega_0, b_2}|_{(\beta), 1} = b_2^{-1/2 - \beta} |q|_{(\beta), 1}, \tag{58}$$

$$|q_{\omega_0, b_2}|_1 = b_2^{-1/2} |q|_1, \tag{59}$$

$$|q_{\omega_0, b_2}|_2 = 1, \tag{60}$$

$$|q_{\omega_0, b_2}|_\infty = b_2^{1/2} |q|_\infty, \tag{61}$$

$$|q_{\omega_0, b_2}|_{(\beta), 2} = b_2^{-\beta} |q|_{(\beta), 2}, \tag{62}$$

$$|w_{Tb_1}|_1 = 1, \tag{63}$$

$$|w_{Tb_1}|_\infty = (b_1 T)^{-1} |w|_\infty, \tag{64}$$

$$|w_{Tb_1}|_{(\beta),1} = (b_1 T)^\beta |w|_{(\beta),1}, \tag{65}$$

$$|q_{\omega_0,b_2} * w_{Tb_1}|_1 \le b_2^{-1/2} |q|_1, \tag{66}$$

$$|q_{\omega_0,b_2} * w_{Tb_1}|_2 \le |w_{Tb_1}|_2 |q_{\omega_0,b_2}| \le b_2^{-1/2}(b_1 T)^{-1/2} |w|_2 |q|, \tag{67}$$

$$|q_{\omega_0,b_2} * w_{Tb_1}|_\infty \le (b_1 T)^{-1} b_2^{-1/2}|w|_\infty|q|, \tag{68}$$

$$|q_{\omega_0,b_2} * w_{Tb_1}|_{(\beta),1} \le 2^{(\beta-1)_+}\big(b_2^{-1/2}(b_1 T)^\beta |q|_1 |w|_{(\beta),1} + b_2^{-1/2-\beta} |q|_{(\beta),1}\big). \tag{69}$$

If q and w have compact supports both included in $[-\tilde{a}, \tilde{a}]$ for some $\tilde{a} > 0$, we have

$$|w_{Tb_1}|_{a,1} = |w|_{a Tb_1,1} \le e^{a\tilde{a} Tb_1} |w|_1, \tag{70}$$

$$|q_{\omega_0,b_2}|_{a,1} = b_2^{-1/2} |q|_{a/b_2,1} \le b_2^{-1/2} e^{a\tilde{a} b_2^{-1}} |q|_1, \tag{71}$$

$$|q_{\omega_0,b_2} * w_{Tb_1}|_{a,1} \le |w_{Tb_1}|_{a,1} q_{\omega_0,b_2}|_{a,1} \le b_2^{-1/2} e^{a\tilde{a}(Tb_1+b_2^{-1})} |q|_1. \tag{72}$$

Proof. All these bounds are straightforward. We use the usual L^p bounds for convolution $|h \star g|_1 \le |h|_1 |g|_1$ and $|h \star g|_2 \le |h|_2 |g|_1$. When necessary, the weights are handled by using

$$|s|^\beta \le 2^{(\beta-1)_+}\big(|s - t|^\beta + |t|^\beta\big) \quad \text{and} \quad e^{a|s|} \le e^{a|t|}e^{a|s-t|}. \tag{73}$$

$$\square$$

Lemma 12. *Let T, b_1, b_2 satisfy* (21) *and* (22). *Then for all $a_1, a_2 > 0$, we have*

$$\exp\big(a_2\big(Tb_1 + b_2^{-1}\big) - a_1 T\big) \le \max\big(1, \exp\big(2a_2 e^{1/a_1}\big)\big).$$

In particular, under Convention 2, *we have* $\exp(A_2(Tb_1 + b_2^{-1}) - A_1 T) \lesssim 1$.

Proof. First note that, by (21), $b_2^{-1} \le Tb_1$ and thus

$$\exp\big(a_2\big(Tb_1 + b_2^{-1}\big) - a_1 T\big) \le \exp\big(2a_2 T(b_1 - a_1)\big).$$

If $b_1 \le a_1$ this upper bound is at most 1 and otherwise, using that $b_1 \le 1$ and then (22), we have that $T(b_1 - a_1) \le T \le e^{1/b_1} \le e^{1/a_1}$. $\square$

Appendix B: Proof of main results

B.1. Proof of Theorem 1 (local mean density estimation)

For treating the bias, expressed as (34), we apply (29) in Theorem 4 with $g = w_{Tb_1}$. Using the norm estimates of equations (63) and (65), we immediately get

$$\mathbb{E}\big[\widehat{m}_{b_1}(u_0)\big] - m_1(u_0) \lesssim T^{-\beta}\big(1 + |w_{Tb_1}|_{(\beta),1}\big) \lesssim b_1^\beta + T^{-\beta}.$$

For treating the variance, expressed as (42), we use Proposition 9 with $h(\cdot) = w_{Tb_1}(\cdot - Tu_0)$ along with $|h|_\infty = |w_{Tb_1}|_\infty = (Tb_1)^{-1}$ by (64) and the obvious bound on the support of the kernel, $\mathrm{Supp}(w_{Tb_1}) \lesssim Tb_1$. We immediately get (24), which concludes the proof.

B.2. Proof of Theorem 2 (bias of spectral estimator)

The proof of this theorem requires to show two bounds, namely, the bound of the bias in time direction, (25), and the bound of the bias in frequency direction, (27). These two bounds are proved quite independently.

Proof of (25). The derivations of Section 3.4.2, namely (35), (36) and (37), show that we can decompose $\mathbb{E}[\widehat{\gamma}_{b_2,b_1}(u_0; \omega_0) - \gamma_{b_2}^{<\mathrm{LS}>}(u_0; \omega_0)]$ as (I) + (II) − (III), where

$$(\mathrm{I}) = \int \left(\mathrm{Var}\big(N_T\big(S^{-Tu_0} q_{\omega_0,b_2}(\cdot - t)\big)\big) - \gamma_{b_2}^{<\mathrm{LS}>}(u_0; \omega_0)\right) w_{Tb_1}(t)\, \mathrm{d}t$$

$$(\mathrm{II}) = \int \big| \mathbb{E}\big[\overline{N}_T\big(S^{-Tu_0} q_{\omega_0,b_2}(\cdot - t); u_0\big)\big]\big|^2 w_{Tb_1}(t)\, \mathrm{d}t$$

$$(\mathrm{III}) = \underbrace{\mathrm{Var}\big(N_T\big(S^{-Tu_0} q_{\omega_0,b_2} * w_{Tb_1}\big)\big)}_{(\mathrm{IIIa})} + \underbrace{\big| \mathbb{E}\big[\overline{N}_T\big(S^{-Tu_0} q_{\omega_0,b_2} * w_{Tb_1}; u_0\big)\big]\big|^2}_{(\mathrm{IIIb})}$$

correspond to (39), (40) and (38), respectively. We will show now that

  (i) the term (I) is of order b_1^β;
  (ii) the terms (II) and (IIIb) are of order $b_1^{2\beta} b_2^{-1}$;
  (iii) the term (IIIa) is of order $(Tb_1b_2)^{-1}$;

which will conclude the proof of (25).

Term (I): By (32) in Theorem 5 with $g = q_{\omega_0,b_2}(\cdot - t)$, and recalling from equations (60), (62), (71), (56) and (57) that $|g|_2 = 1$, $|g|_{(\beta),2} \lesssim b_2^{-\beta} + |t|^\beta$ and

$$|g|_{d,1} \le \mathrm{e}^{d|t|} |q_{\omega_0,b_2}|_{d,1} \lesssim \mathrm{e}^{d|t|} b_2^{-1/2} \mathrm{e}^{A_2 b_2^{-1}} \lesssim \mathrm{e}^{d|t|} \mathrm{e}^{A_2 b_2^{-1}},$$

we have for all $t \in \mathbb{R}$,

$$\big| \mathrm{Var}\big(N_T\big(S^{-Tu_0} q_{\omega_0,b_2}(\cdot - t)\big)\big) - \mathrm{Var}\big(N(q_{\omega_0,b_2}; u_0)\big)\big| \lesssim T^{-\beta}\big(b_2^{-\beta} + |t|^\beta + \mathrm{e}^{A_2|t|} \mathrm{e}^{A_2 b_2^{-1} - A_1 T}\big),$$

where we have used that $1 \lesssim b_2^{-\beta} \lesssim \mathrm{e}^{b_2^{-1}}$ and that $|t|^\beta \lesssim \mathrm{e}^{|t|}$.

By (12), we can use this to bound the integrand in the definition of (I) and thus get, using (63), (65) and (70),

$$(\mathrm{I}) \lesssim T^{-\beta}\big(b_2^{-\beta} + (Tb_1)^\beta + \mathrm{e}^{-A_1 T + A_2(Tb_1 + b_2^{-1})}\big).$$

By (21) and applying Lemma 12, we have that the main term between the parentheses is the second one, hence we get (i).

Term (II): Applying (29) in Theorem 4 with $g = q_{\omega_0, b_2}(\cdot - t)$ and using (58), (59) and (56), we get

$$
\begin{aligned}
\left| \mathbb{E}\left[\overline{N}_T \left(S^{-Tu_0} q_{\omega_0, b_2}(\cdot - t); u_0 \right) \right] \right| &\lesssim T^{-\beta} \left(b_2^{-1/2} \left(1 + |t|^\beta \right) + b_2^{-1/2 - \beta} \right) \\
&\lesssim T^{-\beta} \left(b_2^{-1/2} |t|^\beta + b_2^{-1/2 - \beta} \right).
\end{aligned}
\tag{74}
$$

(Since $b_2 \le 1$.) Taking the square and integrating this with respect to $w_{Tb_1}(t)\,\mathrm{d}t$, and using (65), (63) and $T \ge 1$, we get (II) $\lesssim b_1^{2\beta} b_2^{-1}$ as claimed in (ii).

Term (IIIb): This term is treated similarly as (II) but using directly the bounds (66) and (69) inserted into (29) and taking the square. The same order is obtained and (ii) follows.

Term (IIIa): Applying the bound (31) of Theorem 5, we get that, setting here $g = q_{\omega_0, b_2} * w_{Tb_1}$,

$$
\mathrm{Var}\left(N_T(g) \right) \lesssim |g|_2^2 + \mathrm{e}^{-aT} |g|_{d,1}^2.
$$

Now, from (67) and (72), this bound reads

$$
\mathrm{Var}\left(N_T(g) \right) \lesssim b_2^{-1} (Tb_1)^{-1} + \mathrm{e}^{-A_1 T + A_2 (Tb_1 + b_2^{-1})}.
$$

By Lemma 12 and since $Tb_1 b_2 \ge 1$, the first term dominates and we get (iii), which concludes the proof. $\qquad\qquad\square$

We now provide a proof for the second part of the theorem controlling the bias.

Proof of (27). This bound requires the usual control of the kernel-regularization of a smooth function as can be seen from (11). Namely, the function to consider is $\omega \mapsto \gamma^{<\mathrm{LS}>}(\omega; u_0)$ and the kernel is $\omega \mapsto b_2^{-1} |Q((\omega - \omega_0)/b_2)|^2$ which integrates to 1 since $|Q|_2 = |q|_2 / \sqrt{2\pi} = 1$ by (K-2) and the Parseval theorem. Using the formula (8) to express $\omega \mapsto \gamma^{<\mathrm{LS}>}(\omega; u_0)$ and the usual conditions on the kernel (26), it is thus sufficient to prove that, for all $\omega \in \mathbb{R}$,

$$
\frac{m_1^{<\mathrm{LS}>}(u_0)}{2\pi} \left| \left| 1 - P^{<\mathrm{LS}>}(\omega; u_0) \right|^{-2} - \left| 1 - P^{<\mathrm{LS}>}(\omega_0; u_0) \right|^{-2} + C(\omega - \omega_0) \right| \lesssim (\omega - \omega_0)^2,
$$

where C is any constant (possibly depending on ω_0 and u_0 but not on depending on ω). As already seen, we have

$$
m_1^{<\mathrm{LS}>}(u_0) \le \left| \lambda_c^{<\mathrm{LS}>} \right|_\infty / \left(1 - \zeta_1^{<\mathrm{LS}>} \right) \lesssim 1,
$$

so that, we can consider the ratio $m_1^{<\mathrm{LS}>}(u_0)/(2\pi)$ as a constant. For the remaining term involving the function $\omega \mapsto |1 - P^{<\mathrm{LS}>}(\omega; u_0)|^{-2}$, we first observe that

$$
\begin{aligned}
&P^{<\mathrm{LS}>}(\omega; u_0) - P^{<\mathrm{LS}>}(\omega_0; u_0) \\
&= \int p^{<\mathrm{LS}>}(t; u_0) \left(\mathrm{e}^{-i\omega t} - \mathrm{e}^{-i\omega_0 t} \right) \mathrm{d}t \\
&= \tilde{p}(\omega_0; u_0)(\omega - \omega_0) + \int p^{<\mathrm{LS}>}(t; u_0) \mathrm{e}^{-i\omega_0 t} \left(\mathrm{e}^{-i(\omega - \omega_0)t} - 1 - i(\omega - \omega_0)t \right) \mathrm{d}t
\end{aligned}
$$

with $\tilde{p}(\omega_0; u_0) := \int \mathrm{i} t p^{<\mathrm{LS}>}(t, u_0) \mathrm{e}^{-\mathrm{i}\omega_0 t} \, \mathrm{d}t$. In the latter display, the first term is of the form $C(\omega - \omega_0)$ with $|C| \lesssim 1$ and the second term is of order $(\omega - \omega_0)^2$. This comes respectively from

$$\left| \tilde{p}(\omega_0; u_0) \right| \leq \int |t| p^{<\mathrm{LS}>}(t; u_0) \, \mathrm{d}t \leq \zeta_1^{<\mathrm{LS}>} + \zeta_{(2)}^{<\mathrm{LS}>}$$

and

$$\left| \int p^{<\mathrm{LS}>}(t, u_0) \mathrm{e}^{-\mathrm{i}\omega_0 t} \left(\mathrm{e}^{-\mathrm{i}(\omega - \omega_0)t} - 1 - \mathrm{i}(\omega - \omega_0)t \right) \mathrm{d}t \right| \leq \zeta_{(2)}^{<\mathrm{LS}>} (\omega - \omega_0)^2.$$

To conclude the proof, we argue that the form $C(\omega - \omega_0) + R(\omega)$ with $C \lesssim 1$ and $R(\omega) \lesssim (\omega - \omega_0)^2$ satisfied by $P^{<\mathrm{LS}>}(\omega; u_0) - P^{<\mathrm{LS}>}(\omega_0; u_0)$ is inherited by $|1 - P^{<\mathrm{LS}>}(\omega; u_0)|^{-2} - |1 - P^{<\mathrm{LS}>}(\omega_0; u_0)|^{-2}$. This follows from

$$\left| 1 - P^{<\mathrm{LS}>}(\omega; u_0) \right|^{-1} \leq \left(1 - \zeta_1^{<\mathrm{LS}>} \right)^{-1} \in (1, \infty)$$

(using (6)) and from the identity valid for all complex numbers z, z' inside the unit disk

$$\frac{1}{|1-z|^2} - \frac{1}{|1-z'|^2} = \frac{|1-z|^2 - |1-z'|^2}{|1-z'|^4} + \frac{(|1-z|^2 - |1-z'|^2)^2}{|1-z'|^4 |1-z|^2}.$$

The numerator of the first term of this sum can be expressed as a sum of terms depending on the difference $|z - z'|$, and this applies to the numerator of the second term, too (as it is the square of the first numerator):

$$|1-z|^2 - |1-z'|^2 = 2\Re(z - z')\Re(1 - z') + 2\Im(z - z')\Im(1 - z') + |z - z'|^2. \qquad \square$$

B.3. Additional lemmas

As explained in Section 3.4.3, the treatment of the variance is done via the introduction of the "truly" centered process $\overline{\overline{N}}_T$. Here we provide three lemmas, two concerned with useful bounds for this centered process and the third one which controls the quality of the approximation of the estimator based on the centered process.

Lemma 13. *Let $p \geq 1$. Under the conditions of Theorem 3, we have, for all $u_0 \in \mathbb{R}$, $\omega_0 \in \mathbb{R}$, and b_1, b_2, T as in (21),*

$$\left\| \overline{\overline{N}}_T \left(S^{-Tu_0} q_{\omega_0, b_2} * w_{Tb_1} \right) \right\|_p \lesssim (Tb_1b_2)^{-1/2}, \tag{75}$$

where $\overline{\overline{N}}_T$ is defined in (41). Let moreover $h : \mathbb{R} \to \mathbb{C}$ be such that, for all $t \in \mathbb{R}$, we have

$$\left| h(t) \right| \leq \left(a_T + b_T |t|^\beta \right) w_{Tb_1}(t)$$

for two positive constants a_T and b_T (possibly depending on T, b_1 and b_2). Then we have, for all $u_0 \in \mathbb{R}$, $\omega_0 \in \mathbb{R}$, and b_1, b_2, T as in (21),

$$\left\| \overline{\overline{N}}_T \left(S^{-Tu_0} q_{\omega_0,b_2} * h \right) \right\|_p \lesssim \left(a_T + b_T (b_1 T)^\beta \right) (T b_1 b_2)^{-1/2}. \tag{76}$$

Proof. We apply Proposition 9, and get

$$\left\| \overline{\overline{N}}_T \left(S^{-Tu_0} q_{\omega_0,b_2} * w_{Tb_1} \right) \right\|_p \leq A |q_{\omega_0,b_2} * w_{Tb_1}|_\infty \sqrt{n}$$

for some generic constant A and with a positive integer upper bound n on the length of the support of $S^{-Tu_0} q_{\omega_0,b_2} * w_{Tb_1}$, denoted by $\mathrm{Supp}(S^{-Tu_0} q_{\omega_0,b_2} * w_{Tb_1})$. Observing that $\mathrm{Supp}(S^{-Tu_0} q_{\omega_0,b_2} * w_{Tb_1}) \subset \mathrm{Supp}(S^{-Tu_0} q_{\omega_0,b_2}) + \mathrm{Supp}(w_{Tb_1})$, that the length of $\mathrm{Supp}(q_{\omega_0,b_2})$ is of order b_2^{-1} and that the length of $\mathrm{Supp}(w_{Tb_1})$ is of order $T b_1$, we have $n \lesssim b_2^{-1} + T b_1 \lesssim T b_1$. We thus obtain the bound (75) with (68).

The bound (76) is obtained similarly but this time we rely on the bound

$$|q_{\omega_0,b_2} * h|_\infty \leq a_T |q_{\omega_0,b_2} * w_{Tb_1}|_\infty + b_T |q_{\omega_0,b_2}|_1 |w_{Tb_1}|_\infty n^\beta.$$

(Recall that n is length of the support of w_{Tb_1}.) With (68), $n \lesssim T b_1$, (61) and (65) we get $\left| q_{\omega_0,b_2} * h \right|_\infty \lesssim b_2^{-1/2} (b_1 T)^{-1} (a_T + b_T (b_1 T)^\beta)$, which yields (76). $\square$

Lemma 14. *Under the conditions of Theorem 3, we have, for all $u_0 \in \mathbb{R}$, $\omega_0 \in \mathbb{R}$, and b_1, b_2, T as in (21),*

$$\mathrm{Var}\left(\int \left| \overline{\overline{N}}_T \left(S^{-Tu_0} q_{\omega_0,b_2} (\cdot - t) \right) \right|^2 w_{Tb_1}(t) \, \mathrm{d}t \right) \lesssim (T b_1 b_2)^{-1}. \tag{77}$$

Proof. We can write the left-hand side of (77) as

$$\iint \mathrm{Cov}\left(\left| \overline{\overline{N}}_T \left(f(\cdot - t) \right) \right|^2, \left| \overline{\overline{N}}_T \left(f(\cdot - t') \right) \right|^2 \right) w_{Tb_1}(t) w_{Tb_1}(t') \, \mathrm{d}t \, \mathrm{d}t'.$$

Let ℓ denote the length of the support of q_{ω_0,b_2}, which clearly satisfies for $b_2 \in (0, 1]$,

$$\ell \lesssim b_2^{-1}. \tag{78}$$

Then with $h_1 = q_{\omega_0,b_2}(\cdot - t)$ and $h_2 = q_{\omega_0,b_2}(\cdot - t')$, setting $\gamma := (|t - t'| - \ell)_+$, we have one of the assertions (i), (ii) or (iii) which is satisfied. Hence Corollary 10 gives that, for some $q > 4$,

$$\left| \mathrm{Cov}\left(\left| \overline{\overline{N}}_T(h_1) \right|^2, \left| \overline{\overline{N}}_T(h_2) \right|^2 \right) \right| \leq C_q \left\| \overline{\overline{N}}_T(h_1) \right\|_q^2 \left\| \overline{\overline{N}}_T(h_2) \right\|_q^2 \mathrm{e}^{-\alpha_q (|t - t'| - \ell)_+}. \tag{79}$$

Further we apply Proposition 9 with (61) and (78) and get, for $i = 1, 2$, $\|\overline{\overline{N}}_T(h_i)\|_q^2 \lesssim (|q_{\omega_0, b_2}|_\infty)^2 \ell \lesssim 1$. Hence, we finally get

$$\mathrm{Var}\left(\int \left|\overline{\overline{N}}_T\left(S^{-Tu_0}q_{\omega_0,b_2}(\cdot - t)\right)\right|^2 w_{Tb_1}(t)\,\mathrm{d}t\right) \lesssim \iint \mathrm{e}^{-\alpha_q(|t-t'|-\ell)_+} w_{Tb_1}(t) w_{Tb_1}(t')\,\mathrm{d}t\,\mathrm{d}t'$$

$$\leq |w_{Tb_1} * w_{Tb_1}|_\infty \int \mathrm{e}^{-\alpha_q(|u|-\ell)_+}\,\mathrm{d}u.$$

Now, by (78) we have $\int \mathrm{e}^{-\alpha_q(|u|-\ell)_+}\,\mathrm{d}u \lesssim b_2^{-1}$ and by (63) and (64), we have $\left|w_{Tb_1} * w_{Tb_1}\right|_\infty \lesssim (b_1 T)^{-1}$. Hence, we get (77) and the proof is concluded. $\qquad\square$

Lemma 15. *Under the conditions of Theorem 3, we have, for all $u_0 \in \mathbb{R}$, $\omega_0 \in \mathbb{R}$, and b_1, b_2, T as in (21),*

$$\left\|\widehat{\gamma}_{b_2,b_1}(u_0; \omega_0) - \widetilde{\gamma}_{b_2,b_1}(u_0; \omega_0)\right\|_2 \lesssim b_1^{2\beta} b_2^{-1} + (Tb_1b_2)^{-1}, \tag{80}$$

where $\widehat{\gamma}_{b_2,b_1}(u_0; \omega_0)$ and $\widetilde{\gamma}_{b_2,b_1}(u_0; \omega_0)$ are respectively defined by (13) and (43).

Proof. By definitions (33) and (41), we have, for any integrable test function f, $\overline{\overline{N}}_T(f) = \overline{N}_T(f; u_0) - \mathbb{E}[\overline{N}_T(f; u_0)]$. Thus, (35) and (37) yield the following expression for $\widehat{\gamma}_{b_2,b_1}(u_0; \omega_0)$

$$\int \left|\overline{\overline{N}}_T\left(f(\cdot - t)\right) + \mathbb{E}[\overline{N}_T\left(f(\cdot - t); u_0\right)]\right|^2 w_{Tb_1}(t)\,\mathrm{d}t - \left|\overline{\overline{N}}_T(f * w_{Tb_1}) + \mathbb{E}[\overline{N}_T(f * w_{Tb_1}; u_0)]\right|^2,$$

where we used the test function $f = S^{-Tu_0}q_{\omega_0,b_2}$. Developing the first square modulus and using for the second that $|a + b|^2 \leq 2|a|^2 + 2|b|^2$, and since $\int w_{Tb_1} = 1$, we obtain

$$\left|\widehat{\gamma}_{b_2,b_1}(u_0; \omega_0) - \widetilde{\gamma}_{b_2,b_1}(u_0; \omega_0) - 2\Re(B_T)\right| \leq A_T + 2C_T + 2D_T, \tag{81}$$

where we set, denoting by f^* the complex conjugate of f,

$$A_T := \int \left|\mathbb{E}[\overline{N}_T\left(f(\cdot - t); u_0\right)]\right|^2 w_{Tb_1}(t)\,\mathrm{d}t,$$

$$B_T := \int \overline{\overline{N}}_T\left(f(\cdot - t)\right)\mathbb{E}[\overline{N}_T\left(f^*(\cdot - t); u_0\right)] w_{Tb_1}(t)\,\mathrm{d}t,$$

$$C_T := \left|\overline{\overline{N}}_T(f * w_{Tb_1})\right|^2 \quad \text{and} \quad D_T := \left|\mathbb{E}[\overline{N}_T(f * w_{Tb_1}; u_0)]\right|^2.$$

Note that A_T and D_T have been treated in the proof of Theorem 2 as the (deterministic) terms (II) and (IIIb). The assumptions in Theorem 2 are weaker than that of this lemma. Hence, we can directly use (ii) of the proof of Theorem 2 and obtain that

$$A_T, D_T \lesssim b_1^{2\beta} b_2^{-1}.$$

The bound (75) in Lemma 13 immediately gives

$$\|C_T\|_2 = \left\|\overline{\overline{N}}_T\big(S^{-Tu_0}q_{\omega_0,b_2} * w_{Tb_1}\big)\right\|_4^2 \lesssim (Tb_1b_2)^{-1}$$

and we are left with treating B_T. Note that we can rewrite B_T as $B_T = \overline{\overline{N}}_T(f * h)$, where h is the (deterministic) function $t \mapsto \mathbb{E}[\overline{N}_T(f^*(\cdot - t); u_0)]w_{Tb_1}(t)$. Now, by (74), we have, for all $t \in \mathbb{R}$,

$$\big|h(t)\big| \lesssim b_2^{-1/2}T^{-\beta}\big(b_2^{-\beta} + |t|^\beta\big)w_{Tb_1}(t).$$

Applying (76) with $a_T = b_2^{-1/2}(b_2T)^{-\beta}$ and $b_T = b_2^{-1/2}T^{-\beta}$ in Lemma 13, it follows that

$$\|B_T\|_2 \lesssim b_2^{-1/2}T^{-\beta}\big(b_2^{-\beta} + (b_1T)^\beta\big)(Tb_1b_2)^{-1/2} \lesssim b_2^{-1/2}b_1^\beta(Tb_1b_2)^{-1/2}$$

for $Tb_1b_2 \geq 1$.

Inserting the previous bounds on A_T, B_T, C_T and D_T in (81) yields

$$\left\|\widehat{\gamma}_{b_2,b_1}(u_0; \omega_0) - \widetilde{\gamma}_{b_2,b_1}(u_0; \omega_0)\right\|_2 \lesssim b_2^{-1/2}b_1^\beta(Tb_1b_2)^{-1/2} + b_1^{2\beta}b_2^{-1} + (Tb_1b_2)^{-1}.$$

Using that $2b_2^{-1/2}b_1^\beta(Tb_1b_2)^{-1/2} \leq b_1^{2\beta}b_2^{-1} + (Tb_1b_2)^{-1}$, we get (80). $\qquad\square$

B.4. Proof of Theorem 3 (variance of spectral estimator)

Lemmas 14 and 15, together with the definition of $\widetilde{\gamma}_{b_2,b_1}(u_0; \omega_0)$ in (43), directly give that

$$\mathrm{Var}\big(\widehat{\gamma}_{b_2,b_1}(u_0; \omega_0)\big) \lesssim \frac{1}{Tb_1b_2} + \big(b_1^{2\beta}b_2^{-1} + (Tb_1b_2)^{-1}\big)^2.$$

Since $Tb_1b_2 \geq 1$ in (21), the second term within the squared parentheses is at most of the same order as the term outside the squared parentheses and can thus be discarded. Hence, we obtain Theorem 3.

Acknowledgements

Supplementary Material

Supplement to "Time-frequency analysis of locally stationary Hawkes processes" (DOI: 10.3150/18-BEJ1023SUPP; .pdf). We provide some numerical experiments on transaction data

illustrating our time frequency analysis for point processes. Moreover, we give detailed proofs of the uniform exponential moment bound for stable non-stationary Hawkes processes as stated in Proposition 6 and of the uniform approximation of the variance of a locally stationary Hawkes process as stated in Theorem 5.

References

[1] Aalen, O.O. (1975). *Statistical inference for a family of counting processes*. Ann Arbor, MI: ProQuest LLC. Ph.D. thesis, Univ. California, Berkeley. MR2625917

[2] Andersen, P.K. and Borgan, Ø. (1985). Counting process models for life history data: A review. *Scand. J. Stat*. **12** 97–158. With discussion and a reply by the authors. MR0808151

[3] Bacry, E., Delattre, S., Hoffmann, M. and Muzy, J.F. (2013). Modelling microstructure noise with mutually exciting point processes. *Quant. Finance* **13** 65–77. MR3005350

[4] Birr, S., Volgushev, S., Kley, T., Dette, H. and Hallin, M. (2017). Quantile spectral analysis for locally stationary time series. *J. R. Stat. Soc. Ser. B. Stat. Methodol*. **79** 1619–1643. MR3731679

[5] Chen, F. and Hall, P. (2013). Inference for a nonstationary self-exciting point process with an application in ultra-high frequency financial data modeling. *J. Appl. Probab.* **50** 1006–1024. MR3161370

[6] Chen, F. and Hall, P. (2016). Nonparametric estimation for self-exciting point processes – A parsimonious approach. *J. Comput. Graph. Statist*. **25** 209–224. MR3474044

[7] Dahlhaus, R. (1996). On the Kullback–Leibler information divergence of locally stationary processes. *Stochastic Process*. *Appl*. **62** 139–168. MR1388767

[8] Dahlhaus, R. (1996). Asymptotic statistical inference for nonstationary processes with evolutionary spectra. In *Athens Conference on Applied Probability and Time Series Analysis*, *Vol*. *II* (1995). *Lecture Notes in Statistics* **115** 145–159. New York: Springer. MR1466743

[9] Dahlhaus, R. (2000). A likelihood approximation for locally stationary processes. *Ann*. *Statist*. **28** 1762–1794. MR1835040

[10] Dahlhaus, R. (2009). Local inference for locally stationary time series based on the empirical spectral measure. *J. Econometrics* **151** 101–112. MR2559818

[11] Daley, D.J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes*, *Vol*. *I*: *Elementary Theory and Methods*, 2nd ed. *Probability and Its Applications* (*New York*). New York: Springer. MR1950431

[12] Dedecker, J., Doukhan, P., Lang, G., León, J.R., Louhichi, S. and Prieur, C. (2007). *Weak Dependence*: *With Examples and Applications*. *Lecture Notes in Statistics* **190**. New York: Springer. MR2338725

[13] Giraud, C., Roueff, F. and Sanchez-Perez, A. (2015). Aggregation of predictors for nonstationary sublinear processes and online adaptive forecasting of time varying autoregressive processes. *Ann. Statist*. **43** 2412–2450. MR3405599

[14] Hansen, N.R., Reynaud-Bouret, P. and Rivoirard, V. (2015). Lasso and probabilistic inequalities for multivariate point processes. *Bernoulli* **21** 83–143. MR3322314

[15] Mammen, E. (2017). Nonparametric estimation of locally stationary Hawkes processe. arXiv e-prints.

[16] Neumann, M.H. and von Sachs, R. (1997). Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra. *Ann. Statist*. **25** 38–76. MR1429917

[17] Ramlau-Hansen, H. (1983). Smoothing counting process intensities by means of kernel functions. *Ann. Statist*. **11** 453–466. MR0696058

[18] Reynaud-Bouret, P. and Roy, E. (2006). Some non asymptotic tail estimates for Hawkes processes. *Bull. Belg. Math. Soc. Simon Stevin* **13** 883–896. MR2293215

[19] Richter, S. and Dahlhaus, R. (2017). Cross validation for locally stationary processes. arXiv e-prints.

[20] Roueff, F. and von Sachs, R. (2019). Supplement to "Time-frequency analysis of locally stationary Hawkes processes". DOI:10.3150/18-BEJ1023SUPP.

[21] Roueff, F., von Sachs, R. and Sansonnet, L. (2016). Locally stationary Hawkes processes. *Stochastic Process. Appl.* **126** 1710–1743. MR3483734

[22] van Delft, A. and Eichler, M. (2015). Data-adaptive estimation of time-varying spectral densities. arXiv e-prints.

[23] Zheng, B., Roueff, F. and Abergel, F. (2014). Modelling bid and ask prices using constrained Hawkes processes: Ergodicity and scaling limit. *SIAM J. Financial Math.* **5** 99–136. MR3164121

[24] Zhou, Z. and Wu, W.B. (2009). Local linear quantile estimation for nonstationary time series. *Ann. Statist.* **37** 2696–2729. MR2541444