Mutually exciting point processes with latency*

Yoann Potiron

Faculty of Business and Commerce, Keio University, Tokyo

Vladimir Volkov

School of Business and Economics, University of Tasmania, Australia

Abstract

A novel statistical approach to estimating latency, defined as the time it takes to learn about an event and generate response to this event, is proposed. Our approach only requires a multidimensional point process describing event times, which circumvents the use of more detailed datasets which may not even be available. We consider the class of parametric Hawkes models capturing clustering effects and define latency as a known function of kernel parameters, typically the mode of kernel function. Since latency is not well-defined when the kernel is exponential, we consider maximum likelihood estimation in the mixture of generalized gamma kernels case and derive the feasible central limit theory with in-fill asymptotics. As a byproduct, central limit theory for a latency estimator and related tests are provided. Our numerical study corroborates the theory. An empirical application on high frequency data transactions from the New York Stock Exchange and Toronto Stock Exchange shows that latency estimates for the US and Canadian stock exchanges vary between 1 and 6 milliseconds from 2020 to 2021.

Keywords

Latency, parametric Hawkes processes, mixture of generalized gamma kernels, maximum likelihood estimation, in-fill asymptotics

*CONTACT Yoann Potiron Faculty of Business and Commerce, 2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan. Email: yoann.potiron@fbc.keio.ac.jp Web: https://www.fbc.keio.ac.jp/~potiron

1 Introduction

Over the last decade, financial markets have undergone revolutionary institutional and technological changes. Massive increases in computer power have led to algorithmic trading, explosions in sub-second orders, and large increases in trading volume. These changes transformed the financial markets and made latency an inherent part of investment process. Low latency, or simply latency, can be broadly defined following Hasbrouck and Saar (2013) as the time it takes to learn about an event and generate response to this event.

Although the term latency is widely used in finance, it is also related to delay, a more common term in statistics. In fact, delay is present in datasets related to seismology, insurance, criminology, sociology and medicine as in e.g., Harris (1990). Indeed, just like a time lag before a trading event is revealed to market participants, there is also a time lag before a tweet post becomes available on X, or a registration of medical incidents. Despite the fact that we mainly use the term latency and our empirical application focuses on finance, a statistician should keep this parallel in mind when reading the manuscript.

In the finance literature an approach to obtaining latency is heavily dependent on datasets that are not available in many cases. Hasbrouck and Saar (2013) propose a widely used low latency measure based on strategic runs representing series of submissions, cancellations, and executions that are linked by direction, size, and timing, and which are likely to arise from a single algorithm. However, as the proposed approach requires detailed information about the timing and other characteristics of cancellations - which is not available for many markets - a unified statistical framework for accurate estimation of latency is a missing link in the literature.

An alternative method, proposed in this paper, is to make use of a statistical model relying solely on multidimensional event times. These time series are normally available to a statistician, making our approach widely suitable for applications. Building upon the stylized fact that arrival times are not deterministic, an obvious choice is to use tractable Poisson processes, in which inter-arrival times are IID. Yet, Poisson processes are not well suited for modelling the arrival times as the empirical literature on inter-arrival durations points out that trades tend to cluster together over time. Accordingly, the class of autoregressive conditional duration (ACD) models is introduced in Engle and Russell (1998). ACD models are closely aligned with GARCH models and there are many multidimensional extensions of GARCH models. Most of these extensions are also applicable to ACD models, and copulas can be also used to provide appropriate solutions to such extensions, see Heinen and Rengifo (2007), Koopman et al. (2018), Barra et al. (2018) and the references therein. However, these models are hard to generalize to a set-up with asynchronous times.

This motivates using a more suitable class of multidimensional models, the so-called mutually exciting processes such that the occurrence of any event fuels the probability of the next events occurring. The d-dimensional intensity, which can be interpreted as the instantaneous expected number of events, is defined as

$$\lambda(t) = \nu + \int_0^t h(t-s)\, dN_s, \tag{1}$$

where ν is a d-dimensional Poisson baseline, $h(t)$ is a $d \times d$-dimensional kernel matrix whose diagonal components $h^{(i,i)}$ are self-exciting terms for the related i-th process and non-diagonal components $h^{(i,j)}$ are cross-exciting terms made by event from the j-th process to the i-th process.

Considering a classical parametric specification of model (1), the novelty in our model is that we define latency as a known function of parameters, typically the time corresponding to the peak of the kernel function, i.e. the mode. We assume that the latency is a $d \times d$-dimensional matrix. We insist on the fact that latency is not well-defined when the kernel is exponential as the mode is always equal to 0 in that case. Thus, we introduce a mixture of generalized gamma kernels in which latency is well-defined.

In this paper, we focus on in-fill asymptotics, i.e. when T is fixed, since it is well-known that latency changes empirically across different days, and this is also corroborated by the findings in our empirical application. In the absence of latency, there already exists successful attempts to accommodate for in-fill asymptotics in Hawkes processes. Chen and Hall (2013) allow for a nonrandom parametric time-varying baseline. Their in-fill asymptotic results are based on random observation times of order n within the fixed time interval $[0, T]$. A single boosting of the baseline, i.e. $\lambda(t) = \alpha_n \nu_t + \int_0^t h(t-s)\, dN_s$, is considered where $\alpha_n \to \infty$ is a scaling sequence when $n \to \infty$. Chen and Hall (2013) derive a central limit theorem (CLT) for MLE parameters related to the baseline and kernel. Clinet and Potiron (2018) consider stochastic time-varying baseline and kernel parameters in the exponential kernel case, and introduce a joint boosting of the baseline and the kernel, i.e. $\lambda(t) = n\nu_t + \int_0^t n a_s \exp(-n b_s (t-s))\, dN_s$ to derive CLTs on integrated baseline and parameters with local MLE. Kwan et al. (2023) revisit Chen and Hall (2013) in the exponential kernel case and with the same in-fill asymptotics as in Clinet and Potiron (2018), i.e. $\lambda(t) = n\nu_t + \int_0^t n a \exp(-n b(t-s))\, dN_s$. Kwan (2023) considers the non-exponential kernel case and advocates the use of in-fill asymptotics for statistical inference to better match high-frequency data. An example of Aït-Sahalia and Jacod (2014) for financial applications also confirms the feasibility of in-fill asymptotics for financial data.

In the absence of latency, the parametric Hawkes literature provides the results of large-T asymptotics, i.e. when the horizon time $T \to \infty$. Maximum likelihood estimation (MLE) is employed in the seminal paper of Ogata (1978), which shows the asymptotic normality of the MLE for an ergodic stationary point process. However, the definition of ergodicity is vague in that paper and most of the papers on parametric Hawkes models (e.g. Cavaliere et al. (2023), Assumption 1(b) and Remark 2.1) make this ergodicity assumption and point out this assumption is satisfied for Hawkes processes, whereas in fact this is hard to establish. As far as we know, there are only two results in the literature showing rigorously the ergodicity of Hawkes processes. Clinet and Yoshida (2017) provide a general point process framework where they obtain MLE based CLTs in Theorem 3.11 (p. 1809) when assuming ergodicity of the couplet of intensity process and an intensity process derivative. Their general machinery is verified in the case of a Hawkes process with exponential kernel in Theorem 4.6 (p. 1821) by proving first that the couplet of intensity process and the intensity process derivative is mixing and stable, and then ergodicity is implied. With in-fill asymptotics, i.e., when T is fixed based on random observation times of order n and by exploiting a joint boosting of the baseline and the kernel, Kwan (2023) considers the non-exponential kernel case but the author mentions that such a setup is challenging since the resulting intensity process is non-Markovian, thus rendering standard techniques for asymptotic inference of Markov processes futile. Consequently, the author can only show the ergodicity for the intensity procecss itself but not for the couplet of intensity process and intensity process derivative, and only the consistency of the MLE in Theorem 3.4.3 (p. 73) of this paper is shown.

These two results are useful, but Clinet and Yoshida (2017) (Theorem 4.6) is restricted to the exponential kernel case and Kwan (2023) (Theorem 3.4.3) only obtains consistency of the MLE. Thus, no feasible CLTs are available when the kernel is not exponential, and tests on latency cannot be directly inferred from these two results. In our Theorem 1, we consider MLE in the mixture of generalized gamma kernels case and derive the feasible CLT with in-fill asymptotics. Our proof strategy builds on the general machinery of Clinet and Yoshida (2017) by proving first that the couplet of intensity process and intensity process derivative is mixing and stable, and then ergodicity is implied. The novelty in the proofs is in establishing the couplet is mixing and ergodic (Proposition C1). Consistent estimators of the asymptotic variance and feasible normalized CLTs are also provided.

Our latency estimator is defined as the known function of MLE kernel parameters. As a byproduct of Theorem 4.6, we obtain a feasible CLT for our latency estimator (Proposition 2 and Corollary 3). We also construct three Wald tests for latency. We first develop a test for equality between a latency matrix value for one particular index and a fixed value which can in particular be used to test for the inexistence of latency. Our second test focuses on equality between two latency matrix values. Our third test considers multidimensional linear hypotheses on the latency vector. The limit theory of the three tests is established (Corollaries 4, 5 and 6). A complementary approach with nonparametric Hawkes processes and applications to Covid-19 pandemic in France is given in Gámiz et al. (2022) and Gámiz et al. (2023) for the time-varying case.

Our newly developed model contributes to the literature where a compensator of intensity is interpreted as business time or economic time, see Engle (2000). These so-called models of time deformation deal with the relevant time scale as 'economic time' rather than 'calendar time'. Intuitively, economic time measures the arrival rate of new information that influences trading intensity. The joint analysis of transaction times and latency facilitates analysing an impact of trading characteristics observed at ultra-high frequency on the complex interaction between financial markets.

In our empirical analysis we focus on two major stock exchanges, the NYSE (US) and the TSX (Canada). These exchanges have simultaneous trading sessions and are directly comparable due to similarity in trading environment, which creates opportunities for investors to exploit price inefficiencies across venues. Gagnon and Karolyi (2010) show that deviations from price parity are economically small but volatile and can reach large extremes. They report that price parity deviations relate positively to proxies for holding costs that can limit arbitrage. Moreover, the cross-exchange interactions may happen due to this excessive level of volatility.

In both markets latency varies between 1 and 6 milliseconds in 2020 and 2021 with the traders in Canada being overall faster. The presence of interaction between trades and quotes implies that information in trading (quoting) events can be absorbed with delay in response to quoting (trading) events, which is associated with a phenomenon of co-latency. Our findings indicate the existence of co-latency channel working in both directions between trades and quotes in the NYSE and the TSX. We observe a faster reaction of trading co-latency on the response to quotes in both exchanges. This corresponds to Hoffmann (2014) where an ability of fast traders to revise their quotes quickly after news arrivals helps reducing market risks in some markets.

The rest of this paper is organized as follows. The model is introduced in Section 2. Estimation and tests are given in Section 3. The theory is developed in Section 4. Our empirical application is provided in Section 5. Our numerical study is carried over in the supplementary material Appendix A. Examples of kernels, that meet the assumptions of the proposed framework, are given in the supplementary material Appendix B. All proofs of the theory are shown in the supplementary material Appendix C. Additional empirical results belong to the supplementary material Appendix D.

2 A parametric Hawkes model accommodating for latency

We start this section from a literature review. Then, we recall definitions of a point process and a classic parametric Hawkes model, see Ogata (1978), Embrechts et al. (2011), Clinet and Yoshida (2017), Cavaliere et al. (2023) and Kwan (2023). Finally, we introduce a definition of latency, which is novel in the point processes literature.

2.1 Literature review

Hawkes (1971b) and Hawkes (1971a) introduces a family of models for point processes with stochastic intensity called "self-exciting and mutually exciting point processes" such that the occurrence of any event fuels the probability of the next occurring events. Importantly, these papers provide the Bartlett spectrum and the corresponding covariance density function, useful tools for analyzing point process models. Details about these models are discussed in Liniger (2009) and applications in finance are shown in Hawkes (2018) with the references therein.

Over the last few decades Hawkes processes have been widely used in the context of seismology. The classical MLE for point processes is originally described in Rubin (1972), and applied to Hawkes processes in Vere-Jones (1978) and Ozaki (1979). Vere-Jones and Ozaki (1982) rely on the MLE and provide applications to earthquake data. Ogata (1978) shows the asymptotic normality of the MLE for an ergodic stationary point process with large-T asymptotics.

Applications of Hawkes processes in finance have been evolving over the last two decades. Bowsher (2007) considers a two-dimensional Hawkes process model of the timing of trades and mid-quote price changes. Chavez-Demoulin et al. (2005) introduce a marked Hawkes process to model extreme returns. A ten-dimensional Hawkes process model is used by Large (2007). Embrechts et al. (2011) consider the application of Hawkes processes with marks using MLE. Bacry et al. (2013) provide a CLT for the multidimensional Hawkes point process with large-T asymptotics. Aït-Sahalia et al. (2014) model self- and cross-excitation shocks in CDS markets for several European countries using a standard multidimensional Hawkes process with exponential kernels. Aït-Sahalia et al. (2015) study Hawkes jump-diffusion processes in different stock markets and use a parametric moment-based estimation.

The most recent use of point processes is also widespread. Corradi et al. (2020) develop a test for conditional independence in quadratic variation jump components. Ikefuji et al. (2022) analyze the impact of earthquake risk on real estate prices with the use of ETAS Hawkes-based model. A bootstrap approach for Hawkes and more general point processes is developed in Cavaliere et al. (2023). In Karim et al. (2021), the authors provide an analysis of the probabilistic behavior of the couplet of point process and intensity process. Kernel-based estimation of intensity with in-fill asymptotics is presented in van Lieshout (2021). Bennedsen et al. (2023) develop likelihood-based methods for estimation of continuous-time integer-valued trawl processes.

None of these strands of literature provide a formal definition of latency using a point process framework.

2.2 Parametric Hawkes model

Let T stand for the horizon time. A d-dimensional point process

$$(N_t)_{0\leq t\leq T} := (N_t^{(1)}, \dots, N_t^{(d)})_{0\leq t\leq T},$$

corresponds to the accumulated number of market events at time t. In other words, the i-th component, which corresponds to the i-th event type of the point process, is formally defined as

$$\begin{aligned} dN_t^{(i)} := N_t^{(i)} - N_{t^-}^{(i)} &= 1 \text{ if there is an event at time t}, \\ &= 0 \text{ otherwise.} \end{aligned}$$

We will refer to $(T_1^{(i)}, \dots, T_{N^{(i)}}^{(i)})$ for the event times. A point process is driven by its d-dimensional intensity $\lambda(t)$, which can be interpreted as the instantaneous expected number of events since

$$\lambda(t) = \lim_{u\to 0} \mathbb{E}\left[\frac{N_{t+u} - N_t}{u} | \mathcal{F}_t^N\right],$$

where $\mathcal{F}_t^N = \sigma\{N_s, 0 \leq s \leq t\}$ is defined as the canonical filtration generated by N_t. For formal definitions related to the theory of point processes, see Daley and Vere-Jones (2003), Daley and Vere-Jones (2008), and more generally Jacod and Shiryaev (2013).

The parametric mutually exciting processes have a d-dimensional intensity defined as

$$\lambda(t) = \nu^* + \int_0^t h(t-s, \theta_{ker}^*)\, dN_s, \tag{2}$$

where ν^* is a d-dimensional Poisson baseline, $h(t, \theta_{ker}^*)$ is a $d \times d$-dimensional kernel matrix whose diagonal components $h^{(i,i)}$ are self-exciting terms for the related i-th process and non-diagonal components $h^{(i,j)}$ are cross-exciting terms made by event from the j-th process to the i-th process.

2.3 Latency

The latency is defined as a $d \times d$-dimensional matrix which is a known function of the kernel parameter θ^*_{ker}, i.e. we assume that

$$L = F(\theta^*_{ker}). \tag{3}$$

If $L^{(i,j)} > 0$, a latency between an event in process j and its impact on process i is introduced. Typically, we set F such that the latency $L^{(i,j)}$ is specified as the time it takes before reaching the pick, i.e. the mode, of the kernel $h^{(i,j)}(t, \theta^*_{ker})$. This definition of latency is in agreement with the finance literature, e.g. Hasbrouck and Saar (2013), defining it as the time it takes to learn and generate response to a trading event. An advantage of this definition is that latency can be characterized by parameters $\theta^{(i,j)}_{ker}$ associated with factors affecting latency. Such a structural approach permits identifying different aspects of latency. As we show in Section 3.2, sub-parameters of θ, i.e. D, are interpreted as delay measures. This component of latency D identifies the time of learning about a trading event, which is critical for financial applications.

3 Estimation and tests

In this section, we first introduce MLE for the parametric Hawkes model and discuss the in-fill asymptotics used for theoretical analysis. Then, we introduce a mixture of generalized gamma kernels in which latency is well-defined highlighting that latency is not well-defined when the kernel is exponential. Finally, we introduce latency estimation and tests related to latency.

3.1 MLE

We assume that a stochastic basis $\mathcal{B}_n = (\Omega, \mathcal{F}, \mathbf{F}_n, \mathbb{P})$ is given, where the filtration is defined as $\mathbf{F}_n = (\mathcal{F}_t)_{t\in[0,T]}$, where T is the horizon time, i.e. 1 trading interval. The filtration contains all the necessary information to the statistician. We implicitly assume that the defined quantities depend on n, but we do not write explicitly such a dependence when it is clear from the context. Furthermore, we also assume that all the stochastic processes defined in the following are $\mathbf{F}_n$-adapted processes. In particular, this implies that $\mathcal{F}_t^N \subset \mathcal{F}_t$ for any $t \in [0, T]$.

For inference purposes, we consider in-fill asymptotics with joint boosting of the baseline and the kernel. Relying on a parametric approach we assume the existence of an unknown true value $\theta^* = (\nu^*, \theta^*_{ker})$ such that for $i = 1, \dots, d$ we have that the ith component of the $\mathbf{F}$-intensity is equal to

$$\lambda^{(i)}(t, \theta^*) = n\nu^{*,(i)} + \sum_{j=1}^{d} \int_0^t nh^{(i,j)}(n(t-s), \theta^{*,(i,j)}_{ker})dN_s^{(j)}. \tag{4}$$

We define the parameter space as Θ, i.e. $\theta^* \in \Theta$. In Equation (4), in-fill asymptotics are based on random observation times of order n within the time interval $[0, T]$ for a finite horizon time T. Kwan (2023) extends the asymptotic analysis of Clinet and Potiron (2018) and Kwan et al. (2023), also based on joint boosting, by not imposing an exponential kernel. Our case is different from in-fill asymptotics of Chen and Hall (2013) who consider no boosting of the kernel. For any space S such that $0 \in S$, we define the space without zero as S^*. We assume that $m \in \mathbb{N}^*$ is such that $m \geq 2d$. We also assume that $\Theta \subset (\mathbb{R}_+^*)^d \times \mathbb{R}^{m-d}$ is a convex, bounded and open space which satisfies the assumptions from the Sobolev embedding Theorem (see Theorem 4.12 (p. 85) in Adams and Fournier (2003)). We rely on the log likelihood process, see Ogata (1978) and Daley and Vere-Jones (2003),

$$l_T(\theta) = \sum_{i=1}^{d} \int_0^T \log(\lambda^{(i)}(t, \theta))dN_t - \sum_{i=1}^{d} \int_0^T \lambda^{(i)}(t, \theta)dt,$$

i.e. the MLE is defined as $\widehat{\theta}_T \in \operatorname{argmax}_{\theta\in\Theta} l_T(\theta)$.

3.2 Mixture of generalized gamma kernels

For any $i = 1, \ldots, d$ and $j = 1, \ldots, d$ the mixture of generalized gamma kernels is defined as

$$h^{(i,j)}(t, \theta_{ker}^{(i,j)}) = \sum_{k=1}^{K^{(i,j)}} \alpha_k^{(i,j)} \frac{p_k^{(i,j)} t^{(D_k^{(i,j)}-1)} \exp(-(t/\beta_k^{(i,j)})^{p_k^{(i,j)}})}{(\beta_k^{(i,j)})^{D_k^{(i,j)}} \Gamma(D_k^{(i,j)}/p_k^{(i,j)})}, \tag{5}$$

in which $\Gamma(\cdot)$ is the gamma function, $\alpha_k^{(i,j)} \in \mathbb{R}_+^*$ is the size of the jump, $\beta_k^{(i,j)} \in \mathbb{R}_+^*$ is the scale parameter, $D_k^{(i,j)} \in \mathbb{R}_+^*$ and $p_k^{(i,j)} \in \mathbb{R}_+^*$ are shape parameters. In Equation (5) the number of terms in the sum corresponding to the cross excitation between the ith and the jth market $K^{(i,j)}$ is fixed by the statistician, so they are not parameters to be estimated. We assume that the parameter related to the kernel is of the form

$$\begin{aligned} \theta_{ker} &= (\theta_{ker}^{(i,j)})_{1 \leq i,j \leq d} = (\theta_{ker}^{(1,1)}, \theta_{ker}^{(1,2)}, \cdots, \theta_{ker}^{(d,d-1)}, \theta_{ker}^{(d,d)}) \\ \theta_{ker}^{(i,j)} &= (\alpha^{(i,j)}, \beta^{(i,j)}, D^{(i,j)}, p^{(i,j)}) \in (\mathbb{R}_+^*)^{K^{(i,j)}} \times (\mathbb{R}_+^*)^{K^{(i,j)}} \times (\mathbb{R}_+^*)^{K^{(i,j)}} \times (\mathbb{R}_+^*)^{K^{(i,j)}}. \end{aligned} \tag{6}$$

For simplicity of exposition, we assume that each term in the sum of Equation (5) is generalized gamma kernel. However, all the theory of this paper also holds when some of parameters $\theta_{ker}^{(i,j)}$ are fixed to a value or equal to each other. In particular, the kernel can be exponential, gamma or Weibull. Several examples covered by this framework are discussed in Appendix B from the supplementary material.

3.3 Latency estimation

We recall from Equation (3) that latency is defined as a $d \times d$-dimensional matrix which is a known function of the kernel parameter θ_{ker}^*, i.e. we have

$$L = F(\theta_{ker}^*) \in \mathbb{R}_+^{d \times d}. \tag{7}$$

The latency estimator is naturally defined as

$$\widehat{L}_T = F(\widehat{\theta}_T). \tag{8}$$

3.4 Tests related to latency

We consider three Wald tests associated with latency. We first provide a test for equality between a latency value $L^{(i,j)}$ for an index $(i,j) \in \{1,\dots,d\}^2$ and a latency value $\widetilde{L} \in [0,T]$, i.e. we define the null hypothesis as $H_0(\widetilde{L}) : \{L^{(i,j)} = \widetilde{L}\}$ and the alternative hypothesis as $H_1(\widetilde{L}) : \{L^{(i,j)} \neq \widetilde{L}\}$. We let our first test statistic be

$$W(\widetilde{L}) = nT \frac{(\widehat{L}_T^{(i,j)} - \widetilde{L})^2}{\widehat{\mathrm{Var}\left[\eta^{(i,j)}\right]}}, \tag{9}$$

where the variance estimator used in the denominator will be defined in Equation (23). In the particular case when $\widetilde{L} = 0$, it provides a test for the absence against the presence of latency, i.e. $H_0(0) : \{L^{(i,j)} = 0\}$ against $H_1(0) : \{L^{(i,j)} > 0\}$.

We secondly propose a test for equality between two latency values $L^{(i,j)}$ and $L^{(k,l)}$ for two indices $(i,j) \in \{1,\dots,d\}^2$ and $(k,l) \in \{1,\dots,d\}^2$, i.e. we define the null hypothesis as $H_0' : \{L^{(i,j)} = L^{(k,l)}\}$ and the alternative hypothesis as $H_1' : \{L^{(i,j)} \neq L^{(k,l)}\}$. We let our second test statistic be

$$W' = nT \frac{(\widehat{L}_T^{(i,j)} - \widehat{L}_T^{(k,l)})^2}{\widehat{\mathrm{Var}\left[\eta^{(i,j)}\right]} + \widehat{\mathrm{Var}\left[\eta^{(k,l)}\right]} - 2\widehat{\mathrm{Cov}\left[\eta^{(i,j)}, \eta^{(k,l)}\right]}}. \tag{10}$$

where the variance and covariance estimators used in the denominator will be defined in Equation (23).

For convenience we rewrite the $d \times d$-dimensional matrix of latencies $(L^{(i,j)})_{i=1,\dots,d}^{j=1,\dots,d}$ as a d^2-dimensional vector of latencies

$$\overline{L} = (L^{(i,j)})_{i,j=1,\dots,d} = (L^{(1,1)}, L^{(1,2)}, \cdots, L^{(d,d)})^T.$$

We thirdly introduce a test of q linear hypotheses on the d^2 latency vector which is expressed with the $q \times d^2$-dimensional matrix R, i.e. we define the null hypothesis as $\overline{H}_0(r) : \{R\overline{L} = r\}$ and the alternative hypothesis as $\overline{H}_1(r) : \{R\overline{L} \neq r\}$ for $r \in \mathbb{R}$. We let our third test statistic be

$$\overline{W}(r) = nT\big(R\widehat{L}_T - r\big)^T\big(R\widehat{\overline{\Gamma}}_T R^T\big)^{-1}\big(R\widehat{L}_T - r\big), \tag{11}$$

where the $d^2 \times d^2$-dimensional covariance matrix estimator used in the denominator will be defined in Equation (34).

4 Theory

In this section, we first derive the feasible CLT of the MLE in the mixture of generalized gamma kernels case with in-fill asymptotics. This extends Clinet and Yoshida (2017) (Theorem 4.6) which is restricted to the exponential kernel case and Kwan (2023) (Theorem 3.4.3) which only obtains consistency of the MLE. Then, we derive the feasible CLT for the latency estimator. Finally, we obtain the limit theory for the tests related to latency.

4.1 CLT with mixture of generalized gamma kernels

We make the following assumptions for the CLT. Following the conventional MLE set-up, we require uniform boundedness of the parameter space.

[A**]** (i) There exists $\nu_- \in \mathbb{R}_+^*$ such that for any $\theta = (\nu, \theta_{ker}) \in \Theta$ we have that

$$\nu^{(i)} > \nu_-. \tag{12}$$

for any $i = 1, \ldots, d$.

(ii) For any $\theta = (\nu, \theta_{ker}) \in \Theta$, we have that the kernel parameter θ_{ker} is of the form (6) and the kernel $h(t, \theta_{ker})$ is of the form (5).

(iii) There exists $p_- \in \mathbb{R}_+^*$ such that for any $i = 1, \ldots, d$, any $j = 1, \ldots, d$ and any $k = 1, \ldots, K^{(i,j)}$ we have that

$$p_k^{(i,j)} > p_-. \tag{13}$$

(iv) There exists $D_- \in \mathbb{R}_+^*$ such that for any $i = 1, \dots, d$, any $j = 1, \dots, d$ and any $k = 1, \dots, K^{(i,j)}$ we have that

$$D_k^{(i,j)} > D_-. \tag{14}$$

(v) Let us define the matrix $\phi(\theta_{ker}) = (\phi^{(i,j)}(\theta_{ker}^{(i,j)}))_{i=1,\dots,d}^{j=1,\dots,d}$ where

$$\phi^{(i,j)}(\theta_{ker}^{(i,j)}) = \int_0^\infty h^{(i,j)}(s, \theta_{ker}^{(i,j)}) ds,$$

and write $\rho(\phi(\theta_{ker}))$ being a spectral radius. There exists $0 < h_+ < 1$ such that for any $\theta \in \Theta$ we have that

$$\rho(\phi(\theta_{ker})) \leq h_+. \tag{15}$$

Condition **[A] (i)**, i.e. the positivity of the baseline, is well-known for Hawkes processes being well-defined. Condition **[A] (ii)** restricts to Hawkes processes with mixture of generalized gamma kernels. Condition **[A] (v)** is also common for Hawkes processes being well-defined, as in the large-T asymptotics where N_t starts from $-\infty$, this assumption is already required to establish the existence of a stationary version of N_t on the same probability space and that N_t tends in distribution to this stationary process for a certain topology, see Theorem 7 in Brémaud and Massoulié (1996); Proposition 4.4 in Clinet and Yoshida (2017)).

Now we derive an important practical implication of Condition **[A] (v)**. We obtain

$$\begin{aligned}
\int_0^\infty h^{(i,j)}(t, \theta_{ker}^{(i,j)}) dt &= \int_0^\infty \sum_{k=1}^{K^{(i,j)}} \alpha_k^{(i,j)} \frac{p_k^{(i,j)} t^{(D_k^{(i,j)}-1)} \exp(-(t/\beta_k^{(i,j)})^{p_k^{(i,j)}})}{(\beta_k^{(i,j)})^{D_k^{(i,j)}} \Gamma(D_k^{(i,j)}/p_k^{(i,j)})} dt \\
&= \sum_{k=1}^{K^{(i,j)}} \alpha_k^{(i,j)} \int_0^\infty \frac{p_k^{(i,j)} t^{(D_k^{(i,j)}-1)} \exp(-(t/\beta_k^{(i,j)})^{p_k^{(i,j)}})}{(\beta_k^{(i,j)})^{D_k^{(i,j)}} \Gamma(D_k^{(i,j)}/p_k^{(i,j)})} dt \\
&= \sum_{k=1}^{K^{(i,j)}} \alpha_k^{(i,j)},
\end{aligned}$$

where we use Definition (5) in the first equality, Tonelli's theorem in the second equality, and the fact that the integral of any distribution is equal to 1. Furthermore, we can deduce that

$$\int_0^\infty h^{(i,j)}(t, \theta_{ker}^{(i,j)})dt \quad \leq \quad \rho(\phi(\theta_{ker}))$$

by a spectral norm definition. Thus, Condition **[A] (v)** implies that $\sum_{k=1}^{K^{(i,j)}} \alpha_k^{(i,j)} \leq h_+$

We now provide the feasible CLT of the MLE in the mixture of generalized gamma kernels case with in-fill asymptotics. This extends Clinet and Yoshida (2017) (Theorem 4.6) which is restricted to the exponential kernel case and Kwan (2023) (Theorem 3.4.3) which only shows consistency of the MLE. We define the space E as $E = \mathbb{R}_+^* \times \mathbb{R}_+^* \times \mathbb{R}^m$. We also define as $C_\uparrow(E, \mathbb{R})$ the set of continuous functions $\psi : (u, v, w) \to \psi(u, v, w)$ from E to $\mathbb{R}$ that satisfy ψ is of polynomial growth in (u, v, w) and in $(\frac{1}{u}, \frac{1}{v}, w)$. For any $i = 1, \ldots, d$ and any $\theta \in \Theta$, we also define the rescaled time-transformed intensity process as $\overline{\lambda}^{(i)}(t, \theta) = \frac{\lambda^{(i)}(\frac{t}{n}, \theta)}{n}$, and the triplet as $X_t^{(i)} = (\overline{\lambda}^{(i)}(t, \theta^*), \overline{\lambda}^{(i)}(t, \theta), \partial_\theta \overline{\lambda}^{(i)}(t, \theta))$. Propositions C1 and C2 from the supplementary material state that $X_t^{(i)}$ is stable, i.e. there exists an $\mathbb{R}_+^*$-valued random variable $\overline{\lambda}_{lim}^{(i)}$ such that $X_{nT}^{(i)} \to^{\mathcal{D}} (\overline{\lambda}_{lim}^{(i)}(\theta^*), \overline{\lambda}_{lim}^{(i)}(\theta), \partial_\theta \overline{\lambda}_{lim}^{(i)}(\theta))$. They also state that the triplet is ergodic, i.e. there exists a mapping $\pi_{\theta^*}^{(i)} : C_\uparrow(E, \mathbb{R}) \times \Theta \to \mathbb{R}$ such that for any $(\psi, \theta) \in C_\uparrow(E, \mathbb{R}) \times \Theta$ we have $\frac{1}{nT} \int_0^{nT} \psi(X_{s,n}^{(i)}) ds \to^{\mathbb{P}} \pi^{(i)}(\psi, \theta)$, where $\pi_{\theta^*}^{(i)}(\psi, \theta) = \mathbb{E}[\psi(\overline{\lambda}_{lim}^{(i)}(\theta^*), \overline{\lambda}_{lim}^{(i)}(\theta), \partial_\theta \overline{\lambda}_{lim}^{(i)}(\theta))]$. Finally, they state that there exists a probability measure $\Pi_{\theta^*}^{(i)}$ on $(E, \mathbf{B}(E))$ such that for any $\psi \in C_\uparrow(E, \mathbb{R})$, we have $\pi_{\theta^*}^{(i)}(\psi, \theta) = \int_E \psi(u, v, w) \Pi_{\theta^*}^{(i)}(du, dv, dw)$. If we consider a vector $z \in \mathbb{R}^m$, we define the tensor product as $z^{\otimes 2} = z \times z^T \in \mathbb{R}^{m \times m}$. We define the $m \times m$-dimensional Fisher information matrix Γ as

$$\Gamma = \sum_{i=1}^d \int_E w^{\otimes 2} \frac{1}{u} \Pi_{\theta^*}^{(i)}(du, dv, dw). \tag{16}$$

In other words Γ^{-1} corresponds to the asymptotic covariance matrix. The Fisher information matrix is estimated from

$$\widehat{\Gamma}_T = -\left[\partial_\theta^2 \overline{l}_T(\widehat{\theta}_T)\right]^{-1}, \tag{17}$$

where $\partial_\theta^2 \bar{l}_T(\theta)$ is the $m \times m$-dimensional Hessian matrix of the time-transformed likelihood defined as $\bar{l}_T(\theta) = \sum_{i=1}^d \int_0^{Tn} \log(\overline{\lambda}^{(i)}(t,\theta)) d\overline{N}_t^{(i)} - \sum_{i=1}^d \int_0^{Tn} \overline{\lambda}^{(i)}(t,\theta) dt$ with $\overline{N}_t^{(i)} = N_{\frac{t}{n}}^{(i)}$. Also ξ is defined as a m-dimensional standard normal vector.

Theorem 1. *We assume that Condition **[A]** holds. As $n \to +\infty$, we have the consistency*

$$\widehat{\theta}_T \to^{\mathbb{P}} \theta^* \tag{18}$$

and the CLT

$$\sqrt{n}(\widehat{\theta}_T - \theta^*) \to^{\mathcal{D}} \sqrt{T}\Gamma^{-1/2}\xi. \tag{19}$$

We show the consistency of the Fisher information matrix estimator

$$\widehat{\Gamma}_T \to^{\mathbb{P}} \Gamma. \tag{20}$$

Moreover, we show the feasible normalized CLT

$$\widehat{\Gamma}_T^{1/2} \sqrt{nT}(\widehat{\theta}_T - \theta^*) \to^{\mathcal{D}} \xi. \tag{21}$$

4.2 CLT for latency

We make the following assumption for deriving the CLT.

[B**]** We assume that $F : \Theta \to \mathbb{R}_+^{d \times d}$ is twice continuously differentiable.

For any $i = 1, \ldots, d$ and $j = 1, \ldots, d$ we define the m-dimensional differential vector corresponding to the (i,j)-index of F at point θ as

$$dF^{(i,j)}(\theta) := \Big(\frac{\partial F^{(i,j)}}{\partial \theta^{(1)}}(\theta), \ldots, \frac{\partial F^{(i,j)}}{\partial \theta^{(m)}}(\theta)\Big).$$

We then define the asymptotic matrix as

$$(\eta^{(i,j)})_{i=1,\ldots,d}^{j=1,\ldots,d} = (dF^{(i,j)}(\theta^*)\Gamma^{-1/2}\xi)_{i=1,\ldots,d}^{j=1,\ldots,d}. \tag{22}$$

We estimate the covariance between $\eta^{(i,j)}$ and $\eta^{(k,l)}$ as

$$\widehat{\mathrm{Cov}\left[\eta^{(i,j)}, \eta^{(k,l)}\right]} = \sum_{q=1}^{m} \Big(\sum_{r=1}^{m} dF^{(i,j,r)}(\widehat{\theta}_T)\big(\widehat{\Gamma}_T^{-1/2}\big)^{(r,q)} \Big) \Big(\sum_{r=1}^{m} dF^{(k,l,r)}(\widehat{\theta}_T)\big(\widehat{\Gamma}_T^{-1/2}\big)^{(r,q)} \Big) \tag{23}$$

and the correlation between the normalized asymptotic covariance matrix (i,j)-index and (k,l)-index as

$$\widehat{\mathrm{Cor}\left[\widetilde{\xi}^{(i,j)}, \widetilde{\xi}^{(k,l)}\right]} = \frac{\widehat{\mathrm{Cov}[\eta^{(i,j)}, \eta^{(k,l)}]}}{\sqrt{\widehat{\mathrm{Var}\left[\eta^{(i,j)}\right]}\,\widehat{\mathrm{Var}\left[\eta^{(k,l)}\right]}}}, \tag{24}$$

Now we provide the feasible CLT for the latency estimator with in-fill asymptotics. This complements the related approach on nonparametric Hawkes processes and applications to Covid-19 pandemic in France given in Gámiz et al. (2022) and Gámiz et al. (2023) for the time-varying case.

Proposition 2. *We assume that Condition **[A]** and Condition **[B]** hold. As $n \to +\infty$, we have the consistency*

$$\widehat{L}_T \to^{\mathbb{P}} L \tag{25}$$

and the CLT

$$\sqrt{n}(\widehat{L}_T^{(i,j)} - L^{(i,j)})_{i=1,\dots,d}^{j=1,\dots,d} \to^{\mathcal{D}} \sqrt{T}(\eta^{(i,j)})_{i=1,\dots,d}^{j=1,\dots,d}. \tag{26}$$

Moreover, $\eta^{(i,j)}$ can be re-expressed for any $i = 1,\dots,d$ and any $j = 1,\dots,d$ as

$$\eta^{(i,j)} \quad = \quad \sum_{q=1}^{m} \Big(\sum_{r=1}^{m} dF^{(i,j,r)}(\theta^*)\big(\Gamma^{-1/2}\big)^{(r,q)} \Big) \xi^{(q)}. \tag{27}$$

We can deduce that

$$\mathrm{Cov}[\eta^{(i,j)}, \eta^{(k,l)}] = \sum_{q=1}^{m} \Big(\sum_{r=1}^{m} dF^{(i,j,r)}(\theta^*)\big(\Gamma^{-1/2}\big)^{(r,q)} \Big) \Big(\sum_{r=1}^{m} dF^{(k,l,r)}(\theta^*)\big(\Gamma^{-1/2}\big)^{(r,q)} \Big). \tag{28}$$

We obtain the consistency of the covariance estimator

$$\widehat{\mathrm{Cov}\left[\eta^{(i,j)}, \eta^{(k,l)}\right]} \to^{\mathbb{P}} \mathrm{Cov}[\eta^{(i,j)}, \eta^{(k,l)}]. \tag{29}$$

We show the feasible normalized CLT

$$\sqrt{nT}\left(\frac{\widehat{L}_T^{(i,j)}-L^{(i,j)}}{\sqrt{\widehat{\mathrm{Var}}\left[\eta^{(i,j)}\right]}}\right)_{i=1,\dots,d}^{j=1,\dots,d} \to^{\mathcal{D}} (\widetilde{\xi}^{(i,j)})_{i=1,\dots,d}^{j=1,\dots,d}, \tag{30}$$

where each component $\widetilde{\xi}^{(i,j)}$ *is a standard normal variable with correlation that satisfies*

$$\mathrm{Cor}\left[\widetilde{\xi}^{(i,j)},\widetilde{\xi}^{(k,l)}\right]=\frac{\mathrm{Cov}[\eta^{(i,j)},\eta^{(k,l)}]}{\sqrt{\mathrm{Var}\left[\eta^{(i,j)}\right]\mathrm{Var}\left[\eta^{(k,l)}\right]}}. \tag{31}$$

The consistency of the correlation estimator is obtained as

$$\widehat{\mathrm{Cor}\left[\widetilde{\xi}^{(i,j)},\widetilde{\xi}^{(k,l)}\right]} \to^{\mathbb{P}} \mathrm{Cor}\left[\widetilde{\xi}^{(i,j)},\widetilde{\xi}^{(k,l)}\right]. \tag{32}$$

For any $i=1,\dots,d$ and $j=1,\dots,d$ each component $\widetilde{\xi}^{(i,j)}$ is a standard normal variable, but the limit matrix $(\widetilde{\xi}^{(i,j)})_{i=1,\dots,d}^{j=1,\dots,d}$ is not a standard normal vector in Equation (30). In what follows, we give a feasible CLT with a standard normal vector in the limit since this is useful for providing the asymptotic theory of the multidimensional Wald test in Corollary 6. To obtain a standard normal vector in the limit, we rewrite the $d\times d$-dimensional matrix of latency estimators $(\widehat{L}_T^{(i,j)})_{i=1,\dots,d}^{j=1,\dots,d}$ as a d^2-dimensional vector of latencies

$$\widehat{\overline{L}}_T=(\widehat{L}_T^{(i,j)})_{i,j=1,\dots,d}=(\widehat{L}_T^{(1,1)},\widehat{L}_T^{(1,2)},\cdots,\widehat{L}_T^{(d,d)})^T$$

and we introduce the $d^2\times d^2$-dimensional asymptotic covariance matrix $\overline{\Gamma}^{-1}$ satisfying

$$(\overline{\Gamma}^{-1})_{i,j=1,\dots,d}^{k,l=1,\dots,d} \quad = \quad \mathrm{Cov}[\eta^{(i,j)},\eta^{(k,l)}]. \tag{33}$$

We estimate the asymptotic covariance matrix with

$$(\widehat{\overline{\Gamma}}_T^{-1})_{i,j=1,\dots,d}^{k,l=1,\dots,d} \quad = \quad \mathrm{Cov}\left[\widehat{\eta^{(i,j)},\eta^{(k,l)}}\right]. \tag{34}$$

We make the following assumption:

[C **]** We assume that the $d^2\times m$-dimensional matrix $\left(\sum_{r=1}^m dF^{(i,j,r)}(\theta^*)\left(\Gamma^{-1/2}\right)^{(r,q)}\right)_{i,j=1,\dots,d}^{q=1,\dots,m}$ has full rank.

Corollary 3. *We assume that Condition **[A]**, Condition **[B]** and Condition **[C]** hold. As $n \to +\infty$, we have the CLT*

$$\sqrt{n}(\widehat{\overline{L}}_T - \overline{L}) \to^{\mathcal{D}} \sqrt{T}\overline{\Gamma}^{-1/2}\overline{\xi}, \tag{35}$$

where $\overline{\xi}$ follows a d^2-dimensional standard normal vector. We obtain the consistency for the asymptotic covariance matrix

$$\widehat{\overline{\Gamma}}_T \to^{\mathbb{P}} \overline{\Gamma}. \tag{36}$$

Moreover, we provide the feasible normalized CLT

$$\widehat{\overline{\Gamma}}_T^{1/2}\sqrt{nT}(\widehat{L}_T^{(i,j)} - L^{(i,j)})_{i,j=1,\dots,d} \to^{\mathcal{D}} \overline{\xi}. \tag{37}$$

4.3 Tests related to latency

The following corollary shows that the first Wald test statistic converges in distribution to a chi-squared distribution with one degree of freedom under the null hypothesis and is consistent under the alternative hypothesis. The proof is based on Proposition 2. We denote Q_u as the quantile function of the chi-squared distribution with one degree of freedom.

Corollary 4. *We assume that Condition **[A]** and Condition **[B]** hold. As $n \to +\infty$ and for any latency value $\widetilde{L} \in [0, T]$, the first Wald test statistic $W(\widetilde{L})$ converges in distribution to a chi-squared distribution with one degree of freedom under the null hypothesis $H_0(\widetilde{L}) : \{L^{(i,j)} = \widetilde{L}\}$ and is consistent under the alternative hypothesis $H_1(\widetilde{L}) : \{L^{(i,j)} \neq \widetilde{L}\}$, i.e. for any $0 < \alpha < 1$ we have*

$$\text{size}(\alpha) = \mathbb{P}\{W(\widetilde{L}) > Q_{1-\alpha} \mid H_0(\widetilde{L})\} \to \alpha, \tag{38}$$

$$\text{power}(\alpha) = \mathbb{P}\{W(\widetilde{L}) > Q_{1-\alpha} \mid H_1(\widetilde{L})\} \to 1. \tag{39}$$

The second Wald test statistic converges in distribution to a chi-squared distribution with one degree of freedom under the null hypothesis and is consistent. The proof is based on Proposition 2.

Corollary 5. *We assume that Condition **[A]** and Condition **[B]** hold. We also assume that* $(\eta^{(i,j)}, \eta^{(k,l)})$ *is a two-dimensional random vector. As* $n \to +\infty$*, the second Wald test statistic* W' *converges in distribution to a chi-squared distribution with one degree of freedom under the null hypothesis* $H_0' : \{L^{(i,j)} = L^{(k,l)}\}$ *and is consistent under the alternative hypothesis* $H_1' : \{L^{(i,j)} \neq L^{(k,l)}\}$*, i.e. for any* $0 < \alpha < 1$ *we have*

$$\mathrm{size}'(\alpha) = \mathbb{P}\{W' > Q_{1-\alpha} \mid H_0'\} \to \alpha, \tag{40}$$

$$\mathrm{power}'(\alpha) = \mathbb{P}\{W' > Q_{1-\alpha} \mid H_1'\} \to 1. \tag{41}$$

Finally, the following corollary shows the third Wald test statistic converging to a chi-squared distribution with q degrees of freedom under the null hypothesis and being consistent. The proof is based on Corollary 3. We denote $Q_u^{(q)}$ as the quantile function of the chi-squared distribution with q degrees of freedom.

Corollary 6. *We assume that Condition **[A]**, Condition **[B]** and Condition **[C]** hold. As* $n \to +\infty$ *and for any* $r \in \mathbb{R}$*, the third Wald test statistic* $\overline{W}(r)$ *converges in distribution to a chi-squared distribution with* q *degrees of freedom under the null hypothesis* $\overline{H}_0(r) : \{R\overline{L} = r\}$ *and is consistent under the alternative hypothesis* $\overline{H}_1(r) : \{R\overline{L} \neq r\}$*, i.e. for any* $0 < \alpha < 1$ *we have*

$$\overline{\mathrm{size}}(\alpha) = \mathbb{P}\{\overline{W}(r) > Q_{1-\alpha}^{(q)} \mid \overline{H}_0(r)\} \to \alpha, \tag{42}$$

$$\overline{\mathrm{power}}(\alpha) = \mathbb{P}\{\overline{W}(r) > Q_{1-\alpha}^{(q)} \mid \overline{H}_1(r)\} \to 1. \tag{43}$$

5 Empirical application

In this section we analyze the performance of the proposed model using the transaction data for the New York Stock Exchange (NYSE) and the Toronto Stock Exchange (TSX).

5.1 Data

Our sample period runs from January 2, 2020 to December 31, 2021. Each day, we construct a sample of stocks included in the S&P 500 index and the TSX composite index and that are traded in the NYSE and the TSX. We obtain trade and mid-quote price, i.e. the average price between best bid and ask prices, and time stamps from the consolidated trade history in the transaction Datascope database. Following Hasbrouck (2018) all the stock trades and quotes between 9.45 a.m. and 3.45 p.m. (EST) are considered. Our selection of characteristics for each stock and each day includes millisecond time stamps. These stock characteristics are used to obtain the estimates of latency and to test the hypotheses formulated in the following section.

We apply additional filters in the following order. First, we exclude trades and quotes with zero volumes and prices. Second, we drop a stock-day observation if it takes extreme values falling in the top or bottom 1% of the monthly cross-sections. Finally, each daily sample comprises the 798 stocks traded in the NYSE and the TSX.

Table 1 presents summary sample statistics over the final sample. The sample statistics are computed for the US and Canadian stock exchanges separately. The US market is characterized by a shorter duration, while the TSX experienced a lower standard deviation of trade durations.

Table 1: Summary statistics are reported for all S&P 500 and TSX composite index stocks traded in the NYSE and TSX. The daily average statistics are obtained over the sample from 02/01/2020 to 31/12/2021. Durations are expressed in milliseconds.

	Obs.	Mean	Std dev	Min	Max
	NYSE				
Trade Duration	273,845,398	33.304	149.052	1	63247.101
Mid-Quote Duration	3,361,459,591	2.987	4.983	1	72.370
	TSX				
Trade Duration	120,801,281	77.245	127.918	1	4336.670
Mid-Quote Duration	1,409,644,068	6.798	10.286	1	222.901

5.2 Hypotheses

To understand the evolution of latency in the NYSE and TSX we formulate testable hypotheses of interest. All hypotheses are tested for the whole sample including all trading days and p-values can be obtained with Corollary 6. To verify that our results of testing hypotheses are not distorted due to a multiple statistical inference problem, we implement a Bonferroni adjustment of Holm (1979) for all p-values. The adjusted p-values computed at the 1% level provide the identical conclusions about all hypotheses confirming the statistical robustness of our results. Another robustness check of our testing results is conducted following Bajgrowicz et al. (2016) and the results are in agreement with the Bonferroni corrected tests.

First, we conjecture if latency exists in both exchanges and for mid-quote and trade events.

Latency exists in the NYSE and TSX stock exchanges, i.e. $\overline{H}_0(0) : L^{(i,i)} = 0$ *for all* $i = 1, \ldots, 4$.

A p-value of 0.001 confirms the existence of latency at the 5% level. One may argue that latency is solely characterized by technological development of a stock exchange. In this case a technology arms race would not be defined by timing of orders and latency estimates are expected to be similar in both exchanges. If this is not the case, transforming competition on speed into competition on price is possible when firms strategically consider the timing of order submissions, see e.g. Budish et al. (2015).

Latency varies across the stock exchanges, i.e. $\overline{H}_0(0) : L^{(1,1)} = L^{(3,3)}$ *and* $L^{(2,2)} = L^{(4,4)}$.

A p-value of 0.002 provides evidence of rejecting the null hypothesis at the 5% level. This suggests that there exist additional sources of market information that must be taken into consideration when a new measure of latency is designed. This conjecture extends an idea of Riordan and Storkenmaier (2012) about interrelation between latency and price discovery.

The presence of latency is a feature of modern financial markets, but it is unclear if changes in latencies across stock exchanges create market co-movements. This conjecture is supported by Baron et al. (2019) who find latency to be used as a channel for cross-exchange interactions. We call this phenomenon co-latency. Co-latency can be considered as a channel of emerging spillovers between the market events corroborating findings of Malceniece et al. (2019) about the substantial impact of trading activity on co-movements in stock returns. Following Aït-Sahalia et al. (2015) the presence of spillovers is associated with statistically significant cross-excitation effects.

Co-latency is observed for different events (trades, quotes) within the stock exchanges, i.e. (*) $\overline{H}_0(0) : L^{(1,2)} = L^{(2,1)} = 0$ *and* $L^{(3,4)} = L^{(4,3)} = 0$ *or (**)* $\overline{H}_0(0) : L^{(3,1)} = L^{(3,2)} = L^{(4,1)} = L^{(4,2)} = 0$ *and* $L^{(1,3)} = L^{(1,4)} = L^{(2,3)} = L^{(2,4)} = 0$.

Rejecting (*) with p-value=0.003 confirms the existence of co-latency in both exchanges. However, (**) is not rejected (p-value=0.203), justifying the co-location argument of Brogaard et al. (2015) and confirming that co-latency does not spread across exchanges.

5.3 Latency in the US and Canadian stock exchanges

We now discuss the estimates of latency obtained by the MLE procedure presented in Section 3. Trade and mid-quote time stamps of all the stocks are used to estimate Model (4) for each day and for both the NYSE and the TSX. In this case the kernel matrix $h^{(i,j)}$ is 4×4-dimensional and the market interaction between the US and Canadian stock exchanges is captured by the cross excitation terms $h^{(i,j)}$ when $i \neq j$. The shape of the kernel matrix $h^{(i,j)}$ follows the gamma specification discussed in Appendix B.2 which is a special case of mixture of generalized gamma kernels presented in Section 3.2 with parameters α, β and D. Following the results from the previous section we discuss only (co)latency estimates within the NYSE and the TSX captured by $h^{(1,1)}, h^{(2,2)}, h^{(3,3)}, h^{(4,4)}$ and $h^{(1,2)}, h^{(2,1)}, h^{(3,4)}, h^{(4,3)}$. This is verified by the testing results in the previous section and in agreement with the colocation argument of Shkilko and Sokolov (2020) implying the region specific nature of trading activity.

Delay D is discussed now as an important factor contributing to latency. Parameter estimates $\widehat{D_T^{(i,j)}}$ are presented in Figure 1 for each trading day.[1] The delay parameters $D^{(1,1)}$, $D^{(2,2)}$ for the NYSE and $D^{(3,3)}$, $D^{(4,4)}$ for the TSX trade and mid-quote events are statistically different from zero changing between 1 and 5 milliseconds. A substantial drop in delay from almost 4 to 2 milliseconds observed in Figure 1d happened in November 26th, 2021 which was the worst day of year for North American stock markets. On this day S&P 500 index dropped more than 2% due to a new Covid variant found in South Africa triggering a shift from risk assets and accelerating the speed of trading. The cross-exciting parameter $D^{(1,2)}$ capturing delay in the NYSE trades due to mid-quote events in the same exchange fluctuates between 1.4 and 4 milliseconds. In February and March 2020, active trading during the start of COVID pandemic creates longer delays $D^{(1,2)}$ and $D^{(3,4)}$ showing that market participants respond more quickly to trades when information is flowing. Delays of responding quotes to trades $D^{(2,1)}$ and

[1] The estimates of parameters $\alpha^{(i,j)}$ and $\beta^{(i,j)}$ are not presented here, but available in Appendix D.

$D^{(4,3)}$ are longer for both the NYSE and the TSX suggesting that in both markets trades absorb market information faster.

The estimates of latency and co-latency $\widehat{L_T^{(i,j)}}$ for the NYSE and the TSX obtained for each day over the sample are presented in Figure 2. The latency estimates $\widehat{L_T^{(i,i)}}$ for the US and Canada change between 2 and 5 milliseconds over the sample. Overall the TSX is faster with the latency just below 4 milliseconds for trades and mid-quotes. In February and March 2020, the start of COVID-19 pandemic, co-latency in the NYSE and TSX observed from Figures 2(e),(g) jumped a few times between 1 and almost 4 milliseconds. This pattern is attributed to suspending floor trading due to COVID-19 pandemic in February and March 2020. Another common pattern for both exchanges observed from Figures 2 (e),(f),(g),(h) is a faster reaction of trading co-latency in response to quotes. Confidence intervals for $\widehat{L_T^{(1,2)}}$ and $\widehat{L_T^{(3,4)}}$ are wider comparing to $\widehat{L_T^{(2,1)}}$ and $\widehat{L_T^{(4,3)}}$ indicating uncertainty about the impact of quote events on trades which reduces in 2021 for $\widehat{L_T^{(4,3)}}$. Our findings indicate the existence of co-latency channel working in both directions between the US and Canada. This corresponds to Hoffmann (2014) where an ability of fast traders to revise their quotes quickly after news arrivals helps reducing market risks in some markets, i.e. the TSX and the NYSE in our case.

6 Conclusion

A novel statistical approach to estimating latency, defined as the time it takes to learn about an event and generate response to this event, is described. Outside of finance this definition helps understanding and modelling delay in reactions to events for point processes. The problem is formulated to be solved by the comprehensive use of stochastic analysis techniques. More specifically, we have considered the class of parametric Hawkes models, which circumvents the use of more detailed datasets which may not even be available. We define latency as a known

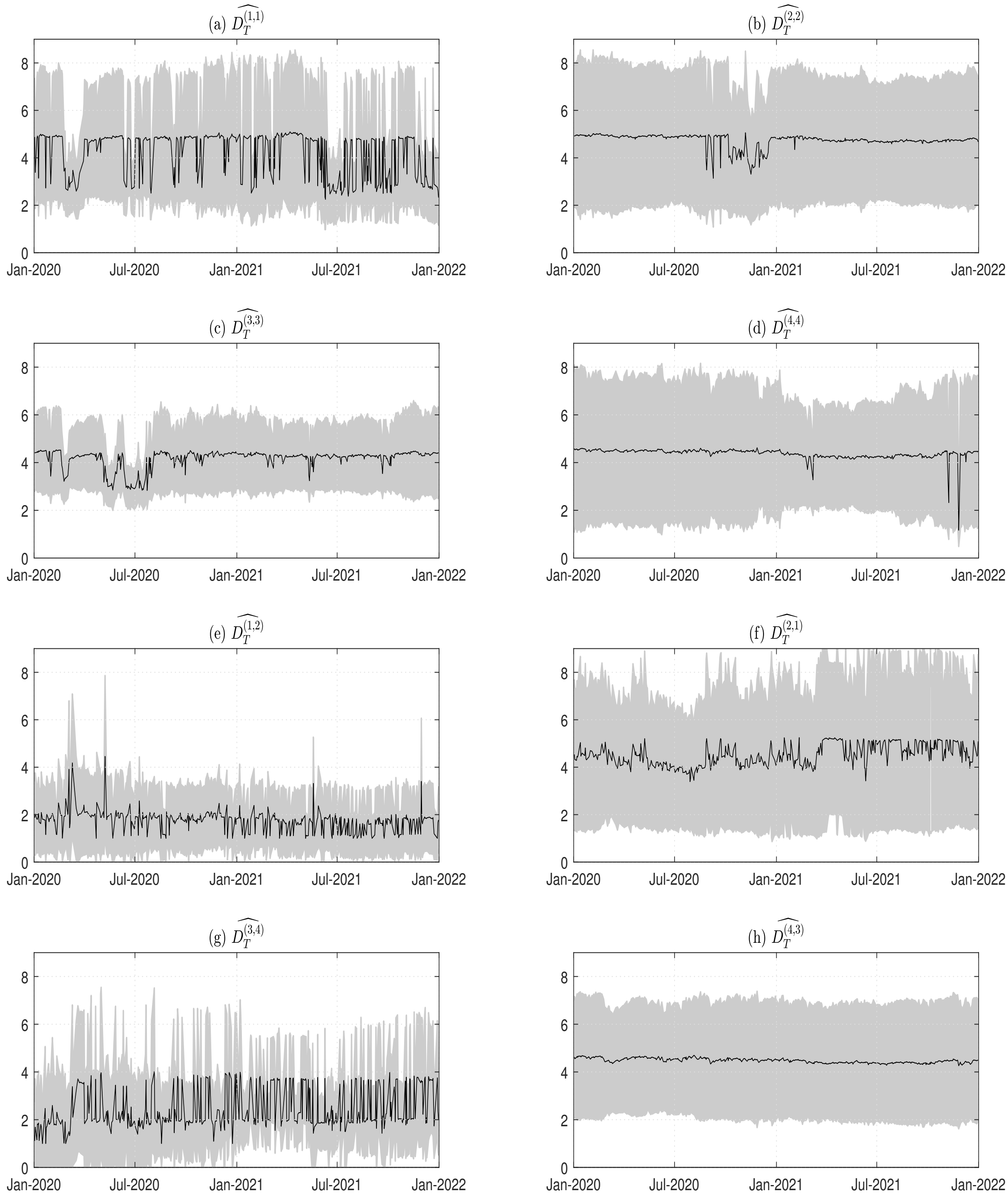

Figure 1: Parameter estimates $\widehat{D_T^{(i,j)}}$ obtained from Equation (4). Parameter estimates for each day in the NYSE and TSX are shown. 90% confidence intervals are also presented.

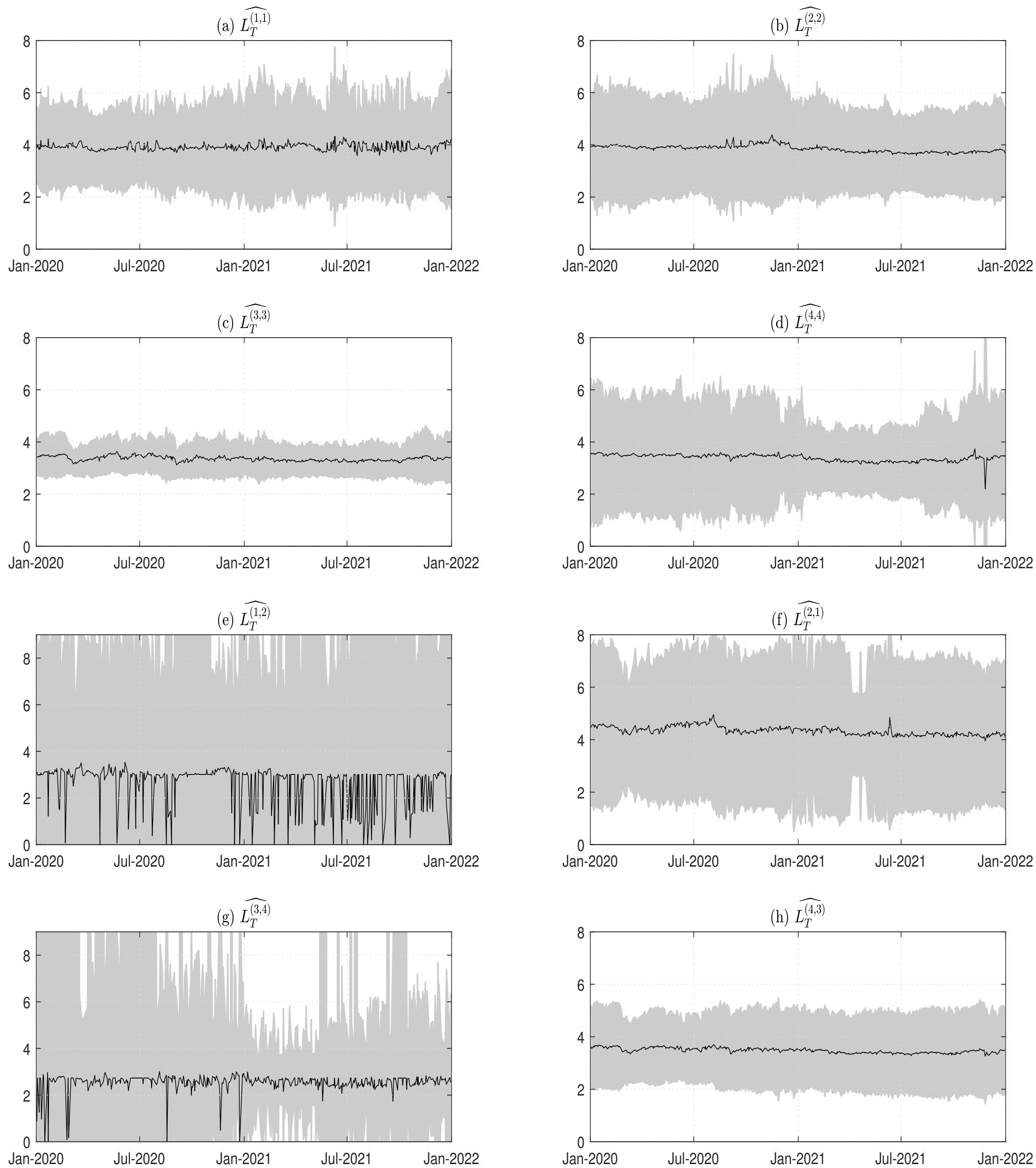

Figure 2: Latency and co-latency estimates $\widehat{L_T^{(i,j)}}$ obtained from Model (4). Each graph shows co-latency estimates for each day across all events in the NYSE and TSX. 90% confidence intervals are reported.

function of kernel parameters, typically the mode of kernel function. Since latency is not well-defined when the kernel is exponential, we consider maximum likelihood estimation in the mixture of generalized gamma kernels case and derive the feasible CLT with in-fill asymptotics. As a byproduct, CLT for a latency estimator, defined as the function of parameter estimates, and three tests were deduced. Latency estimates for the US and Canadian stock exchanges are found to vary between 1 and 6 milliseconds from 2020 to 2021. The existence of co-latency channel working in both directions between the US and Canada is also confirmed.

Supplementary material

Our numerical study is carried over in Appendix A. Examples are given in Appendix B. All proofs are shown in Appendix C. An additional empirical result belongs to Appendix D.

Acknowledgments

We thank Dylan Small (the co-Editor), the anonymous Associate Editor, four anonymous reviewers, Simon Clinet, Dobrislav Dobrev, participants of the Intrinsic Time Conference in Finance 2022 for discussions. Research assistantship from Fedor Fradkin is acknowledged.

Funding

Financial support from the Japanese Society for the Promotion of Science under grants 17K13718, 20K13470 and 23H00807 (Potiron) is gratefully acknowledged. Volkov gratefully acknowledges support from the Basic Research Program of the National Research University Higher School of Economics (HSE University).

References

Adams, RA, and Fournier, JJF. 2003. *Sobolev spaces.* Elsevier.

Aït-Sahalia, Y, and Jacod, J. 2014. *High-frequency financial econometrics.* Princeton University Press.

Aït-Sahalia, Y, Laeven, RJA, and Pelizzon, L. 2014. Mutual excitation in eurozone sovereign CDS. *Journal of Econometrics*, **183**, 151–167.

Aït-Sahalia, Y, Cacho-Diaz, J, and Laeven, RJA. 2015. Modeling financial contagion using mutually exciting jump processes. *Journal of Financial Economics*, **117**, 585–606.

Bacry, E, Delattre, S, Hoffmann, M, and Muzy, JF. 2013. Some limit theorems for Hawkes processes and application to financial statistics. *Stochastic Processes and their Applications*, **123**, 2475–2499.

Bajgrowicz, P, Scaillet, O, and Treccani, A. 2016. Jumps in high-frequency data: Spurious detections, dynamics, and news. *Management Science*, **62**(8), 2198–2217.

Baron, M, Brogaard, J, Hagströmer, B, and Kirilenko, A. 2019. Risk and return in high-frequency trading. *Journal of Financial and Quantitative Analysis*, **54**, 993–1024.

Barra, I, Borowska, A, and Koopman, SJ. 2018. Bayesian dynamic modeling of high-frequency integer price changes. *Journal of Financial Econometrics*, **16**(3), 384–424.

Bennedsen, M, Lunde, A, Shephard, N, and Veraart, AED. 2023. Inference and forecasting for continuous-time integer-valued trawl processes. *Journal of Econometrics*, **236**(2), 105476.

Bowsher, CG. 2007. Modelling security market events in continuous time: Intensity based, multivariate point process models. *Journal of Econometrics*, **141**, 876–912.

Brémaud, P, and Massoulié, L. 1996. Stability of nonlinear Hawkes processes. *The Annals of Probability*, 1563–1588.

Brogaard, J, Hagströmer, B, Nordén, L, and Riordan, R. 2015. *Trading fast and slow: Colocation and liquidity.*

Budish, E, Cramton, P, and Shim, J. 2015. The high-frequency trading arms race: Frequent batch auctions as a market design response. *The Quarterly Journal of Economics*, **130**, 1547–1621.

Cavaliere, G, Lu, Y, Rahbek, A, and Stærk-**O**stergaard, J. 2023. Bootstrap inference for Hawkes and general point processes. *Journal of Econometrics*, **235**, 133–165.

Chavez-Demoulin, V, Davison, AC, and McNeil, AJ. 2005. Estimating value-at-risk: a point process approach. *Quantitative Finance*, **5**, 227–234.

Chen, F, and Hall, P. 2013. Inference for a nonstationary self-exciting point process with an application in ultra-high frequency financial data modeling. *Journal of Applied Probability*, **50**, 1006–1024.

Clinet, S, and Potiron, Y. 2018. Statistical inference for the doubly stochastic self-exciting process. *Bernoulli*, **24**, 3469–3493.

Clinet, S, and Yoshida, N. 2017. Statistical inference for ergodic point processes and application to limit order book. *Stochastic Processes and their Applications*, **127**, 1800–1839.

Corradi, V, Distaso, W, and Fernandes, M. 2020. Testing for jump spillovers without testing for jumps. *Journal of the American Statistical Association*, **115**, 1214–1226.

Daley, DJ, and Vere-Jones, D. 2003. *An Introduction to the Theory of Point Processes*. 2 edn. Vol. 1. Springer-Verlag New York.

Daley, DJ, and Vere-Jones, D. 2008. *An introduction to the theory of point processes: General theory and structure*. 2 edn. Vol. 2. Springer Verlag.

Embrechts, P, Liniger, TJ, and Lin, L. 2011. Multivariate Hawkes processes: an application to financial data. *Journal of Applied Probability*, **48**(A), 367–378.

Engle, RF. 2000. The Econometrics of ultra-high-frequency data. *Econometrica*, **68**, 1–22.

Engle, RF, and Russell, JR. 1998. Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, **66**, 1127–1162.

Gagnon, L, and Karolyi, G. 2010. Multi-market trading and arbitrage. *Journal of Financial Economics*, **97**, 53–80.

Gámiz, ML, Mammen, E, Martínez-Miranda, MD, and Nielsen, JP. 2022. Missing link survival analysis with applications to available pandemic data. *Computational Statistics & Data Analysis*, **169**, 107405.

Gámiz, ML, Mammen, E, Martínez-Miranda, MD, and Nielsen, JP. 2023. Monitoring a developing pandemic with available data. *arXiv preprint arXiv:2308.09919*.

Harris, J. 1990. Reporting Delays and the Incidence of AIDS. *Journal of the American Statistical Association*, **85**, 915–924.

Hasbrouck, J. 2018. High-frequency quoting: Short-term volatility in bids and offers. *Journal of Financial and Quantitative Analysis*, **53**, 613–641.

Hasbrouck, J, and Saar, G. 2013. Low-latency trading. *Journal of Financial Markets*, **16**, 646–679.

Hawkes, AG. 1971a. Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, **33**, 438–443.

Hawkes, AG. 1971b. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, **58**, 83–90.

Hawkes, AG. 2018. Hawkes processes and their applications to finance: a review. *Quantitative Finance*, **18**(2), 193–198.

Heinen, A, and Rengifo, E. 2007. Multivariate autoregressive modeling of time series count data using copulas. *Journal of Empirical Finance*, **14**(4), 564–583.

Hoffmann, P. 2014. A dynamic limit order market with fast and slow traders. *Journal of Financial Economics*, **113**, 156–169.

Holm, S. 1979. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, **6**, 65–70.

Ikefuji, M, Laeven, RJA, Magnus, JR, and Yue, Y. 2022. Earthquake risk embedded in property prices: Evidence from five Japanese cities. *Journal of the American Statistical Association*, **117**, 82–93.

Jacod, J, and Protter, P. 2011. *Discretization of processes*. Vol. 67. Springer Science & Business Media.

Jacod, J, and Shiryaev, A. 2013. *Limit theorems for stochastic processes*. Vol. 288. Springer Science & Business Media.

Kallenberg, O. 1997. *Foundations of modern probability*. Vol. 2. Springer.

Karim, R, Laeven, RJA, and Mandjes, M. 2021. Exact and asymptotic analysis of general multivariate Hawkes processes and induced population processes. *arXiv preprint arXiv:2106.03560*.

Koopman, SJ, Lit, R, Lucas, A, and Opschoor, A. 2018. Dynamic discrete copula models for high-frequency stock price changes. *Journal of Applied Econometrics*, **33**(7), 966–985.

Kwan, TKJ. 2023. *Asymptotic analysis and ergodicity of the Hawkes process and its extensions*. Ph.D. thesis, UNSW Sydney.

Kwan, TKJ, Chen, F, and Dunsmuir, WTM. 2023. Alternative asymptotic inference theory for a nonstationary Hawkes process. *Journal of Statistical Planning and Inference*, **227**, 75–90.

Large, J. 2007. Measuring the resiliency of an electronic limit order book. *Journal of Financial Markets*, **10**, 1–25.

Liniger, TJ. 2009. *Multivariate Hawkes processes*. Ph.D. thesis, ETH Zurich.

Malceniece, L., Malcenieks, K., and Putniņš, T. 2019. High frequency trading and comovement in financial markets. *Journal of Financial Economics*, **134**, 381–399.

Ogata, Y. 1978. The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Annals of the Institute of Statistical Mathematics*, **30**, 243–261.

Ogata, Y. 1988. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, **83**, 9–27.

Ozaki, T. 1979. Maximum likelihood estimation of Hawkes' self-exciting point processes. *Annals of the Institute of Statistical Mathematics*, **31**, 145–155.

Riordan, R, and Storkenmaier, AS. 2012. Latency, liquidity and price discovery. *Journal of Financial Markets*, **15**, 416–437.

Rubin, I. 1972. Regular point processes and their detection. *IEEE Transactions on Information Theory*, **18**, 547–557.

Shkilko, A, and Sokolov, K. 2020. Every cloud has a silver lining: Fast trading, microwave connectivity, and trading costs. *The Journal of Finance*, **75**, 2899–2927.

van Lieshout, MNM. 2021. Infill asymptotics for adaptive kernel estimators of spatial intensity. *Australian & New Zealand Journal of Statistics*, **63**, 159–181.

Vere-Jones, D. 1978. Earthquake prediction-a statistician's view. *Journal of Physics of the Earth*, **26**, 129–146.

Vere-Jones, D, and Ozaki, T. 1982. Some examples of statistical estimation applied to earthquake data: I. Cyclic Poisson and self-exciting models. *Annals of the Institute of Statistical Mathematics*, **34**, 189–207.

Appendices

This part corresponds to the "supplementary material" of "Mutually exciting point processes with latency" by Yoann Potiron and Vladimir Volkov submitted to Journal of the American Statistical Association. Our numerical study is carried over in the Appendix A. Examples are given in the Appendix B. All proofs of the theory are shown in the Appendix C. Additional empirical results belong to the Appendix D.

A Numerical study

The performance of the model is now explored via a simple multidimensional simulation experiment. Consider the 5-dimensional specification of Equation (4) with intensity given by

$$\lambda^{(i)}(t,\theta^*) = n\nu^{*,(i)} + \sum_{j=1}^{5} \int_0^{t^-} nh^{(i,j)}(n(t-s), \theta_{ker}^{*,(i,j)}) dN_s^{(j)}, \tag{A1}$$

where $\nu^* = \begin{bmatrix} 0.01 & 0.01 & 0.01 & 0.01 & 0.01 \end{bmatrix}'$, $h(t,\theta)$ is a gamma kernel defined in Equation (B2) and the kernel parameters θ_{ker}^* are chosen such that

$$h(t,\theta_{ker}^*) =$$

$$\begin{bmatrix} 0.15\frac{t^8 exp(-t)}{\Gamma(9)} & 0.16\frac{t^{10} exp(-t/1.5)}{\Gamma(11)} & 0.14\frac{t^{10} exp(-t/1.2)}{\Gamma(11)} & 0.25\frac{t^5 exp(-t/2)}{\Gamma(6)} & 0.14\frac{t^9 exp(-t/1.1)}{\Gamma(10)} \\ 0.14\frac{t^5 exp(-t/2)}{\Gamma(6)} & 0.15\frac{t^{11} exp(-t/2)}{\Gamma(12)} & 0.24\frac{t^{10} exp(-t/1.5)}{\Gamma(11)} & 0.15\frac{t^9 exp(-t/2.1)}{\Gamma(10)} & 0.25\frac{t^8 exp(-t/1.7)}{\Gamma(9)} \\ 0.14\frac{t^7 exp(-t/1.8)}{\Gamma(8)} & 0.25\frac{t^8 exp(-t/1.2)}{\Gamma(9)} & 0.15\frac{t^9 exp(-t/1.6)}{\Gamma(10)} & 0.24\frac{t^{10} exp(-t/1.8)}{\Gamma(11)} & 0.15\frac{t^9 exp(-t/1.6)}{\Gamma(10)} \\ 0.25\frac{t^8 exp(-t/2)}{\Gamma(9)} & 0.14\frac{t^7 exp(-t/1.2)}{\Gamma(8)} & 0.14\frac{t^6 exp(-t/1.3)}{\Gamma(7)} & 0.24\frac{t^8 exp(-t/1.5)}{\Gamma(9)} & 0.16\frac{t^7 exp(-t/2)}{\Gamma(8)} \\ 0.24\frac{t^9 exp(-t/1.5)}{\Gamma(10)} & 0.15\frac{t^{5.5} exp(-t/2)}{\Gamma(6.5)} & 0.14\frac{t^6 exp(-t/1.5)}{\Gamma(7)} & 0.26\frac{t^{5.6} exp(-t/2.1)}{\Gamma(6.6)} & 0.15\frac{t^7 exp(-t/1.8)}{\Gamma(8)} \end{bmatrix}. \tag{A2}$$

From Equation (B2), we can deduce that the Hawkes process generated by the intensity (A1) has a baseline equal to $\widetilde{\nu}^{*,(i)} = n\nu^{*,(i)}$ and a gamma kernel $\widetilde{h}(t,\theta)$ with true value parameters

$(n\alpha^{(i,j)}, \frac{\beta^{(i,j)}}{n}, D^{(i,j)})$. Thus, there is a unique relationship between (ν^*, h) and $(\widetilde{\nu}^*, \widetilde{h})$. The choice of parameter values mimics the broad characteristics of the empirical data discussed in Section 5. From Equation (B3), latency and co-latency are obtained as a function of parameters equal to $\widetilde{L}^{(i,j)} = \frac{\beta^{(i,j)}(D^{(i,j)}-1)}{n}$. Thus, we obtain latency and co-latency values below 15 milliseconds. The simulation exercise involves 500 independent replications with the sample size order $n = 100,000$ and setting up $T = 1$ trading interval to generate the data. Since we recall that n corresponds to the order of the number of observations, we note that the sample size order $n = 100,000$ is more conservative than the average number of observations in our empirical study, see Table 1. The elements of the kernel $h(t, \theta^*_{ker})$ are estimated using the MLE approach presented in Section 3.1. The model defined in Equation (A1) meets the theoretical assumptions from Section 4.

Figure A1 illustrates histograms of the kernel estimates $\hat{h}$. The trapezoidal rule is used to numerically compute the integral in Equation (A1). The estimates of all kernel functions are close to their theoretical values and the confidence intervals behave as expected. One case where the confidence intervals are especially narrow is represented by $\hat{h}^{(3,2)}$. This is expected as the kernel specification has the highest mean rate. All kernels approach zero within 50 intervals.

Now we verify the CLT of the latency estimator. Histograms of latency and co-latency estimates $\widehat{L_T^{(i,j)}}$ are presented in Figure A2. The obtained variance lies within the range of 2-3 intervals. Table A1 verifies the finite sample properties. The bias ranges from -0.10 to -0.30, which only affects around 1% of latency values. The estimated latency has relatively similar in sample standard error, which is calculated from $(\widehat{L_T^{(i,j)}} - L^{(i,j)})$, and estimated standard error $\sqrt{\text{Cov}\left[\widehat{\eta^{(i,j)}, \eta^{(k,l)}}\right]}$. This indicates that our variance estimator performs reasonably well with increasing n. To confirm the behavior of variance estimators and asymptotic Gaussianity, we provide histograms of the standardized errors related to the latency estimator $\widehat{L_T^{(i,j)}}$ in Figure A3.

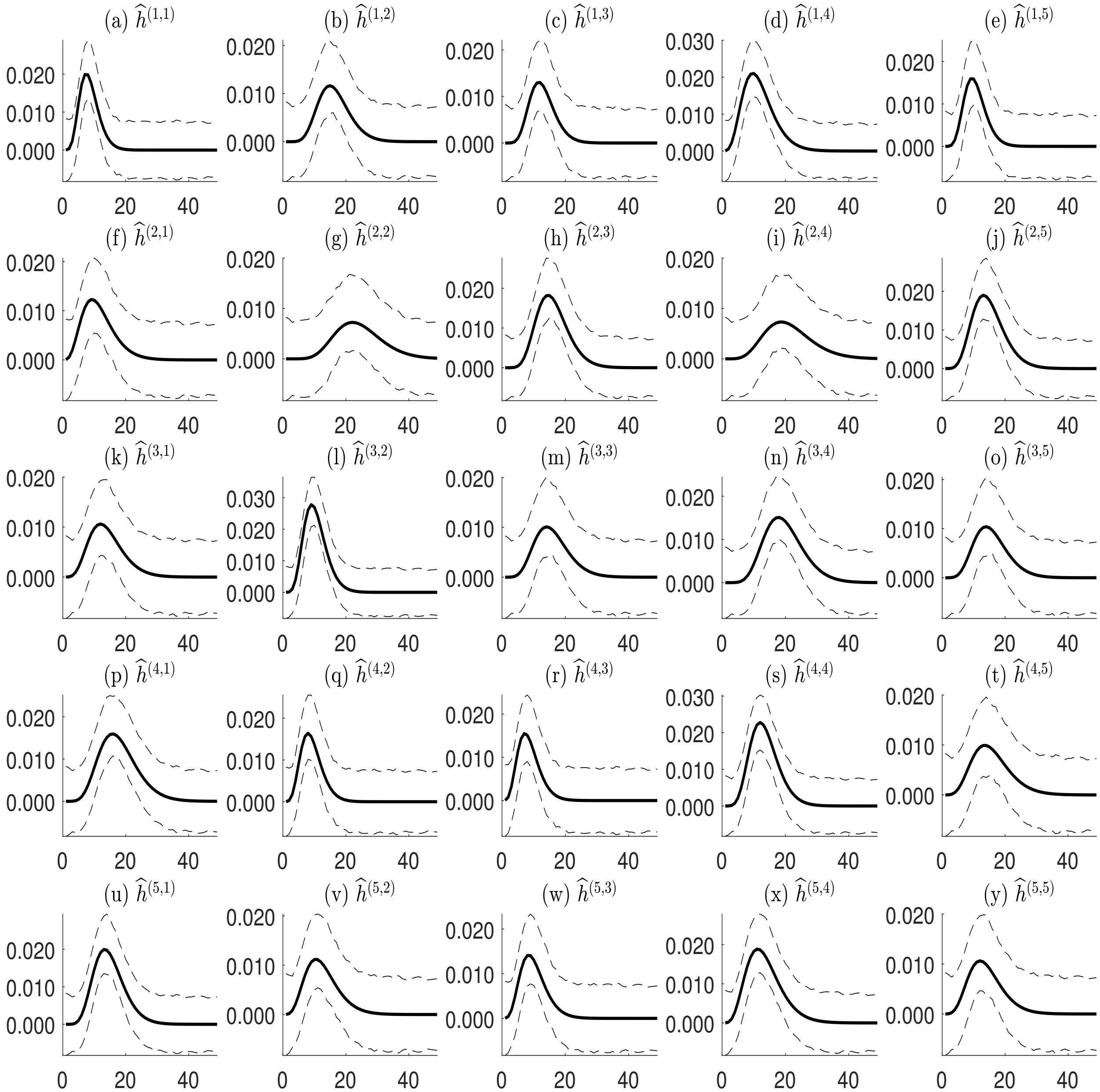

Figure A1: Histograms of the kernel estimates $\hat{h}$ over 50 time intervals. The solid line represents the estimated kernels. The confidence intervals (dashed line) are represented by 2.5% and 97.5% percentiles. The histograms are generated from 500 independent replications with the sample size order $n = 100,000$.

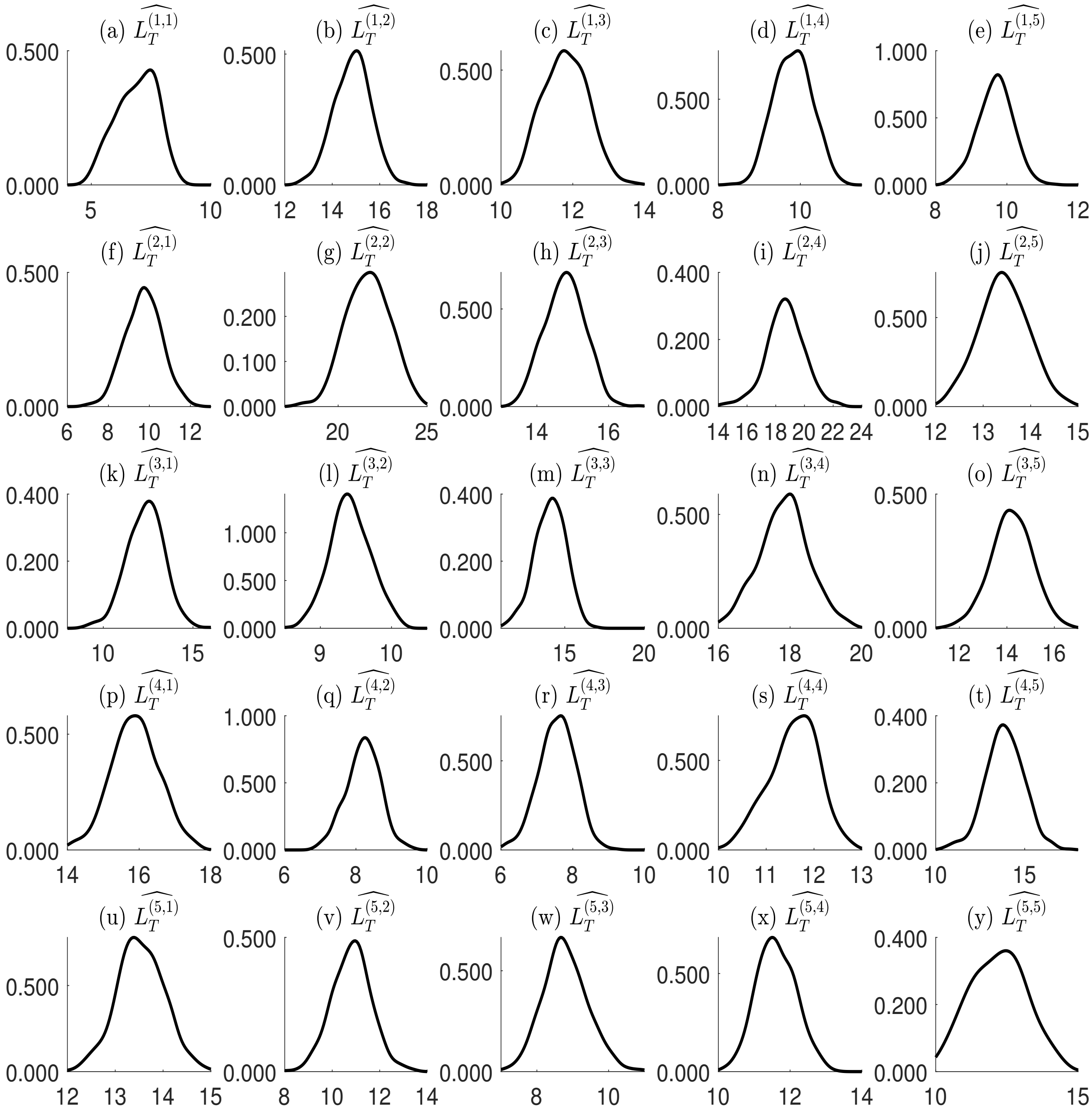

Figure A2: Histograms of latency and co-latency estimates $\widehat{L_T^{(i,j)}}$. The histograms are generated from 500 independent replications with the sample size order $n = 100,000$.

Table A1: Finite sample properties of the latency estimator $\widehat{L_T^{(i,j)}}$. The bias is computed as an average of $(\widehat{L_T^{(i,j)}} - L^{(i,j)})$, in sample standard error is calculated from $(\widehat{L_T^{(i,j)}} - L^{(i,j)})$, and the estimated standard error is $\sqrt{\mathrm{Cov}\left[\widehat{\eta^{(i,j)}, \eta^{(k,l)}}\right]}$. The estimates are generated from 500 independent replications with the sample size order $n = 100,000$.

	Bias	In st. error	Est. st. error		Bias	In st. error	Est. st. error
$\widehat{L_T^{(1,1)}}$	-0.11	2.50	2.62	$\widehat{L_T^{(1,2)}}$	-0.15	2.25	3.85
$\widehat{L_T^{(1,3)}}$	-0.20	1.84	3.43	$\widehat{L_T^{(1,4)}}$	-0.18	1.35	3.13
$\widehat{L_T^{(1,5)}}$	-0.21	1.45	3.11	$\widehat{L_T^{(2,1)}}$	-0.30	2.63	3.11
$\widehat{L_T^{(2,2)}}$	-0.13	2.35	3.18	$\widehat{L_T^{(2,3)}}$	-0.22	1.65	3.84
$\widehat{L_T^{(2,4)}}$	-0.25	3.95	4.31	$\widehat{L_T^{(2,5)}}$	-0.16	1.54	3.66
$\widehat{L_T^{(3,1)}}$	-0.26	3.08	3.50	$\widehat{L_T^{(3,2)}}$	-0.18	1.84	3.06
$\widehat{L_T^{(3,3)}}$	-0.19	2.28	3.31	$\widehat{L_T^{(3,4)}}$	-0.17	2.08	4.22
$\widehat{L_T^{(3,5)}}$	-0.20	2.62	3.76	$\widehat{L_T^{(4,1)}}$	-0.11	1.98	3.98
$\widehat{L_T^{(4,2)}}$	-0.20	1.42	2.86	$\widehat{L_T^{(4,3)}}$	-0.24	1.51	2.75
$\widehat{L_T^{(4,4)}}$	-0.41	1.53	3.39	$\widehat{L_T^{(4,5)}}$	-0.13	3.20	3.72
$\widehat{L_T^{(5,1)}}$	-0.12	1.47	3.68	$\widehat{L_T^{(5,2)}}$	-0.18	2.46	3.28
$\widehat{L_T^{(5,3)}}$	-0.20	1.86	2.96	$\widehat{L_T^{(5,4)}}$	-0.15	1.63	3.40
$\widehat{L_T^{(5,5)}}$	-0.19	2.04	3.03				

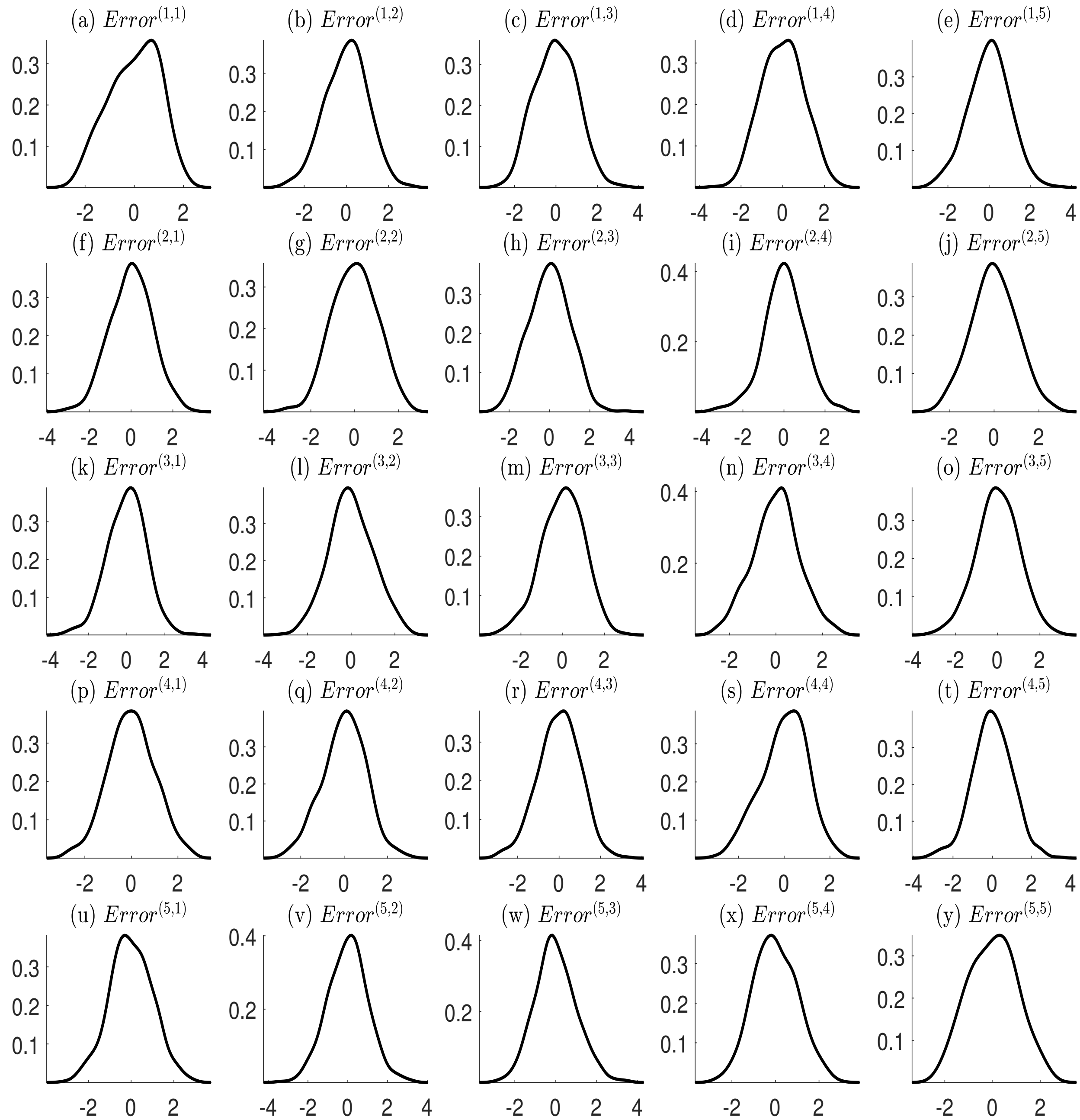

Figure A3: Histograms of the standardized errors for the latency estimator $\widehat{L_T^{(i,j)}}$. The histograms are generated from 500 independent replications with the sample size order $n = 100,000$.

Now we conduct hypothesis testing to confirm the size and power of the tests provided in Section 4.3 for different sample sizes. While the sample size order in our empirical application is larger than $n = 100,000$ providing a better size and power, we aim to demonstrate that the proposed latency test can be used in other areas of statistics. We compare the sample size orders of $n = 10,000$, $n = 50,000$ and $n = 100,000$. The sample size order of $n = 50,000$ is similar to a size of dataset from Ogata (1988) who considered earthquakes of magnitude 6 or more that occurred in Japan and its vicinity over almost 100 years. The significance level is set up at 5% level. Table A2 reports power and size at the 5% level of the tests: $H_A : L^{(1,1)} = 0$, $H_B : L^{(1,2)} = 0$, $H_C : L^{(1,3)} = 0$, $H_D : L^{(1,4)} = 0$, $H_E : L^{(1,5)} = 0$ against one-sided alternatives. For smaller samples the tests are slightly undersized in most cases but approach 5% for the sample size order of $n = 100,000$. The power is bigger than 0.93 in all cases and approaches 1. This highlights the potential of applying our method not only in finance but also in other areas of statistics such as seismology.

Table A2: Power and size at the 5% level of the tests: $H_A : L^{(1,1)} = 0$, $H_B : L^{(1,2)} = 0$, $H_C : L^{(1,3)} = 0$, $H_D : L^{(1,4)} = 0$, $H_E : L^{(1,5)} = 0$ against one-sided alternatives. 500 independent replications are used for simulation.

Null	H_A	H_B	H_C	H_D	H_E	H_A	H_B	H_C	H_D	H_E
Sample order			Size					Power		
10,000	0.040	0.016	0.010	0.044	0.011	0.932	0.944	0.964	0.956	0.948
50,000	0.032	0.036	0.060	0.036	0.044	0.938	0.951	0.952	0.958	0.952
100,000	0.039	0.044	0.046	0.048	0.046	0.974	0.996	1.000	0.959	1.000

B Examples

In this section, we provide five examples of kernels, i.e. exponential, gamma, Weibull, generalized gamma, and mixture of several kernels, which meet the assumptions of our framework. We insist on the fact that latency is not well-defined when the kernel is exponential as the mode is always equal to 0 in that case. For the remaining examples, the latency is defined as the mode of the kernel.

B.1 Exponential kernel

The conventional exponential kernel is defined as

$$h^{(i,j)}(t, \theta_{ker}^{(i,j)}) = \alpha^{(i,j)} \frac{exp(-t/\beta^{(i,j)})}{\beta^{(i,j)}}, \quad \alpha^{(i,j)} \in (0, h_+], \beta^{(i,j)} \in \mathbb{R}_+^*. \tag{B1}$$

This is a particular case of the generalized gamma kernel (5) when $p^{(i,j)} = D^{(i,j)} = 1$. On the one hand, the exponential kernel satisfies the assumptions of this paper, so it is a valid kernel form. On the other hand, we insist on the fact that latency is not well-defined when the kernel is exponential as the mode is always equal to 0 in that case. Thus, the exponential kernel is not suitable for estimating latency.

B.2 Gamma kernel

The gamma kernel is defined as

$$h^{(i,j)}(t, \theta_{ker}^{(i,j)}) = \alpha^{(i,j)} \frac{t^{(D^{(i,j)}-1)} exp(-t/\beta^{(i,j)})}{(\beta^{(i,j)})^{D^{(i,j)}} \Gamma(D^{(i,j)})}, \quad \alpha^{(i,j)} \in (0, h_+], \beta^{(i,j)} \in \mathbb{R}_+^*, D^{(i,j)} \in \mathbb{R}_+^*. \tag{B2}$$

This is a particular case of the generalized gamma kernel (5) when $p^{(i,j)} = 1$. We define latency as the mode, which can be expressed as

$$\begin{aligned} L^{(i,j)} &= \beta^{(i,j)}(D^{(i,j)} - 1) \quad \forall i, j, D^{(i,j)} > 1. && \text{(B3)} \\ L^{(i,j)} &= 0 \quad \forall i, j, D^{(i,j)} \leq 1. \end{aligned}$$

If $L^{(i,j)} > 0$, or equivalently $D^{(i,j)} > 1$, a latency between an event in process j and its impact on process i is introduced. If $L^{(i,j)} = 0$, or equivalently $D^{(i,j)} \leq 1$, there is no latency between an event in process j and its impact on process i. Figure B1 illustrates an example of gamma kernel defined in Equation (B2) and exponential kernel defined in Equation (B1) for 20 intervals.

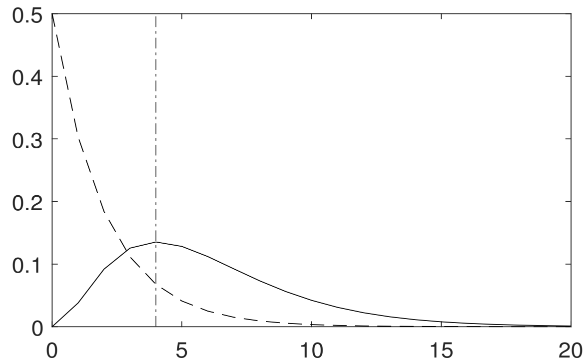

Figure B1: An example of gamma kernel defined in Equation (B2) and exponential kernel defined in Equation (B1) for 20 intervals. The solid line represents the gamma kernel with parameters $\alpha = 1$, $\beta = 2$, and $D = 3$ and the vertical line shows the latency $L = 4$. The dashed line represents the exponential kernel with $\alpha = 1$, $\beta = 2$, and $D = 1$.

B.3 Weibull kernel

The Weibull kernel is defined as

$$h^{(i,j)}(t, \theta_{ker}^{(i,j)}) = \alpha^{(i,j)} \frac{D^{(i,j)} t^{(D^{(i,j)}-1)} \exp(-(t/\beta^{(i,j)})^{D^{(i,j)}})}{(\beta^{(i,j)})^{D^{(i,j)}}}, \qquad \alpha^{(i,j)} \in (0, h_+], \beta^{(i,j)} \in \mathbb{R}_+^*, D^{(i,j)} \in \mathbb{R}_+^*. \tag{B4}$$

This is a particular case of the generalized gamma kernel (5) when $p^{(i,j)} = D^{(i,j)}$. We define latency as the mode, which can be expressed as

$$\begin{aligned} L^{(i,j)} &= \beta^{(i,j)} \Big(\frac{D^{(i,j)} - 1}{D^{(i,j)}} \Big)^{1/D^{(i,j)}} \quad \forall i, j, D^{(i,j)} > 1. \\ L^{(i,j)} &= 0 \quad \forall i, j, D^{(i,j)} \leq 1. \end{aligned} \tag{B5}$$

B.4 Generalized gamma kernel

The generalized gamma kernel is defined as

$$h^{(i,j)}(t,\theta_{ker}^{(i,j)}) = \alpha^{(i,j)} \frac{p^{(i,j)} t^{(D^{(i,j)}-1)} \exp(-(t/\beta^{(i,j)})^{p^{(i,j)}})}{(\beta^{(i,j)})^{D^{(i,j)}} \Gamma(D^{(i,j)}/p^{(i,j)})}, \alpha^{(i,j)} \in (0, h_+], \beta^{(i,j)}, D^{(i,j)}, p^{(i,j)} \in \mathbb{R}_+^* \quad \text{(B6)}$$

We define latency as the mode, which can be expressed as

$$\begin{aligned} L^{(i,j)} &= \beta^{(i,j)} \Big(\frac{D^{(i,j)}-1}{p^{(i,j)}}\Big)^{1/p^{(i,j)}} \quad \forall i,j, D^{(i,j)} > 1. \\ L^{(i,j)} &= 0 \quad \forall i,j, D^{(i,j)} \leq 1. \end{aligned} \tag{B7}$$

B.5 Mixture of several kernels

The mixture of exponential, gamma, Weibull and generalized gamma kernel is defined as

$$h^{(i,j)}(t,\theta_{ker}^{(i,j)}) = h_{exp}^{(i,j)}(t,\theta_{ker}^{(i,j)}) + h_{gam}^{(i,j)}(t,\theta_{ker}^{(i,j)}) + h_{Wei}^{(i,j)}(t,\theta_{ker}^{(i,j)}) + h_{gengam}^{(i,j)}(t,\theta_{ker}^{(i,j)}), \tag{B8}$$

where

$$\begin{aligned} h_{exp}^{(i,j)}(t,\theta_{ker}^{(i,j)}) &= \sum_{k=1}^{K_1^{(i,j)}} \alpha_{k,1}^{(i,j)} \frac{\exp(-t/\beta_{k,1}^{(i,j)})}{\beta_{k,1}^{(i,j)}}, \\ h_{gam}^{(i,j)}(t,\theta_{ker}^{(i,j)}) &= \sum_{k=1}^{K_2^{(i,j)}} \alpha_{k,2}^{(i,j)} \frac{t^{(D_{k,2}^{(i,j)}-1)} \exp(-t/\beta_{k,2}^{(i,j)})}{(\beta_{k,2}^{(i,j)})^{D_{k,2}^{(i,j)}} \Gamma(D_{k,2}^{(i,j)})}, \\ h_{Wei}^{(i,j)}(t,\theta_{ker}^{(i,j)}) &= \sum_{k=1}^{K_3^{(i,j)}} \alpha_{k,3}^{(i,j)} \frac{D_{k,3}^{(i,j)} t^{(D_{k,3}^{(i,j)}-1)} \exp(-(t/\beta_{k,3}^{(i,j)})^{D_{k,3}^{(i,j)}})}{(\beta_{k,3}^{(i,j)})^{D_{k,3}^{(i,j)}}}, \\ h_{gengam}^{(i,j)}(t,\theta_{ker}^{(i,j)}) &= \sum_{k=1}^{K_4^{(i,j)}} \alpha_{k,4}^{(i,j)} \frac{p_{k,4}^{(i,j)} t^{(D_{k,4}^{(i,j)}-1)} \exp(-(t/\beta_{k,4}^{(i,j)})^{p_{k,4}^{(i,j)}})}{(\beta_{k,4}^{(i,j)})^{D_{k,4}^{(i,j)}} \Gamma(D_{k,4}^{(i,j)}/p_{k,4}^{(i,j)})}, \end{aligned}$$

and we have that

$$\sum_{l=1}^{4} \sum_{k=1}^{K^{(i,j)}} \alpha_{k,l}^{(i,j)} \leq h_+.$$

A general formula for the mode in the mixture of several kernels case is beyond the scope of this paper.

C Proofs

C.1 Notations

Before we start the proofs, we need some more formal definitions. If z is a real number, a vector or a matrix, we define its norm as $\mid z \mid = \sum_k \mid z_k \mid$. When Z is a random variable, we define its L^p-norm as $\mid\mid Z \mid\mid = \mathbb{E}[\mid Z \mid^p]^{1/p}$. When Y_n and Z_n are two sequences of random variables, we define the notation small tau as $Y_n = o_{\mathbb{P}}(Z_n)$, i.e. that $\frac{Y_n}{Z_n}\mathbf{1}_{\{Z_n \neq 0\}} \to^{\mathbb{P}} 0$, and the notation big tau $Y_n = O_{\mathbb{P}}(Z_n)$, i.e. that $\frac{Y_n}{Z_n}\mathbf{1}_{\{Z_n \neq 0\}}$ is stochastically bounded. Moreover, given a Borel space $(E, \mathbf{B}(E))$, $C_b(E, \mathbb{R})$ is defined as the set of continuous and bounded functions from the space E to $\mathbb{R}$. For a measure μ, let $\mathbb{L}^1(\mu)$ be the space of functions that are integrable with respect to μ. Finally, we define for any $i = 1, \dots, d$ the event times of the ith process as

$$(T_1^{(i)}, \dots, T_{N^{(i)}}^{(i)}).$$

Since N_t is a point process, its $\mathbf{F}$-intensity (4) can be re-expressed partly as the sum at jump times, i.e.

$$\lambda^{(i)}(t, \theta^*) = n\nu^{*,(i)} + \sum_{j=1}^{d} \sum_{k \in \mathbb{N}_* \text{ s.t. } T_k^{(j)} < t} nh^{(i,j)}(n(t - T_k^{(j)}), \theta_{ker}^{*,(i,j)}). \tag{C1}$$

C.2 Time transformation and some lemmas

Our proof strategy follows the general machinery of Clinet and Yoshida (2017), which consider large-T asymptotics. To rewrite our problem with in-fill asymptotics as a problem with large-T asymptotics, we consider a time transformation as in Kwan et al. (2023). More specifically, we define the time-transformed filtration as

$$\begin{aligned} \overline{\mathbf{F}}_n &= (\overline{\mathcal{F}}_{t,n})_{t \in [0, nT]}, \\ \overline{\mathcal{F}}_{t,n} &= \mathcal{F}_{\frac{t}{n}, n}. \end{aligned}$$

From now on, we implicitly assume that all the defined quantities are $\overline{\mathbf{F}}_n$-adapted. For any $i = 1, \ldots, d$ we define the ith process of the time-transformed point process as

$$\begin{aligned} \overline{N}_n^{(i)} : [0, nT] &\rightarrow \mathbb{N} \\ t &\mapsto \overline{N}_{t,n}^{(i)} = N_{\frac{t}{n},n}^{(i)}, \end{aligned} \tag{C2}$$

with corresponding jump times

$$(\overline{T}_{1,n}^{(i)}, \ldots, \overline{T}_{N_n^{(i)},n}^{(i)})$$

defined such that $\overline{T}_{k,n}^{(i)} = nT_{k,n}^{(i)}$ and the rescaled time-transformed stochastic $\overline{\mathbf{F}}_n$-intensity process as

$$\begin{aligned} \overline{\lambda}_n^{(i)} : [0, nT] \times \Theta &\rightarrow \mathbb{R}^+ \\ (t, \theta) &\mapsto \overline{\lambda}_n^{(i)}(t, \theta) = \frac{\lambda_n^{(i)}(\frac{t}{n}, \theta)}{n}. \end{aligned} \tag{C3}$$

In this first lemma, we rewrite the rescaled time-transformed stochastic $\overline{\mathbf{F}}_n$-intensity in terms of the time-transformed point process.

Lemma C1. *For any $(t, \theta) \in [0, nT] \times \Theta$ and any $i = 1, \ldots, d$ we have that*

$$\overline{\lambda}_n^{(i)}(t, \theta) = \nu^{(i)} + \sum_{j=1}^{d} \int_0^{t^-} h^{(i,j)}((t-s), \theta_{ker}^{(i,j)}) d\overline{N}_{s,n}^{(j)}. \tag{C4}$$

Proof. If we substitute Equation (4) into Definition (C3), we obtain

$$\overline{\lambda}_n^{(i)}(t, \theta) = \nu^{(i)} + \sum_{j=1}^{d} \int_0^{t^-} h^{(i,j)}((t-s), \theta_{ker}^{(i,j)}) dN_{\frac{s}{n},n}^{(j)}. \tag{C5}$$

Finally, we can conclude by substituting Definition (C2) into Equation (C5). □

The next lemma shows that $\overline{N}_n$ is a multidimensional Hawkes process with a mixture of generalized gamma kernels.

Lemma C2. *We have that*

$$\overline{M}_{t,n} = \overline{N}_{t,n} - \int_0^t \overline{\lambda}_n(s, \theta^*) ds \tag{C6}$$

is a $\overline{\mathbf{F}}_n$-local martingale. In particular, this implies that $\overline{N}_n$ is a multidimensional Hawkes process with a mixture of generalized gamma kernels and related $\overline{\mathbf{F}}_n$-intensity $\overline{\lambda}_n(.,\theta^*)$.

Proof. By definition of a compensator, we have that

$$M_{t,n} = N_{t,n} - \int_0^t \lambda_n(s,\theta^*)ds \tag{C7}$$

is a $\mathbf{F}_n$-local martingale. First, we will show that Equation (C6) is a $\overline{\mathbf{F}}_n$-local martingale. In fact we have

$$\begin{aligned} \overline{M}_{t,n} &= \overline{N}_{t,n} - \int_0^t \overline{\lambda}_n(s,\theta^*)ds \\ &= N_{\frac{t}{n},n} - \int_0^t \frac{\lambda_n(\frac{s}{n},\theta^*)}{n}ds \\ &= N_{\frac{t}{n},n} - \int_0^{\frac{t}{n}} \lambda_n(y,\theta^*)dy \\ &= M_{\frac{t}{n},n}, \end{aligned}$$

where we used Equation (C6) in the first equality, Equation (C2) and Equation (C3) in the second equality, integral change of variable in the third equality and Equation (C7) in the fourth equality. Now, as $M_{t,n}$ is a $\mathbf{F}_n$-local martingale, it is clear that the time-transformed local martingale $M_{\frac{t}{n},n}$ is a $\overline{\mathbf{F}}_n$-local martingale. Then, it means that $\overline{M}_{\frac{t}{n},n}$ is a $\overline{\mathbf{F}}_n$-local martingale, thus we have shown the lemma. Second, we can deduce that $\overline{N}_n$ is a multidimensional Hawkes process with a mixture of generalized gamma kernels and related $\overline{\mathbf{F}}_n$-intensity $\overline{\lambda}_n(.,\theta^*)$ by Theorem 3.17 (p. 32) in Jacod and Shiryaev (2013). □

We also define the log likelihood process of the time-transformed process as

$$\bar{l}_n(\theta) = \sum_{i=1}^d \int_0^{Tn} \log(\overline{\lambda}_n^{(i)}(t,\theta))d\overline{N}_{t,n}^{(i)} - \sum_{i=1}^d \int_0^{Tn} \overline{\lambda}_n^{(i)}(t,\theta)dt, \tag{C8}$$

and $\widehat{\overline{\theta}}_n$ any maximizer of it. The following lemma states that a.s. the MLE on $[0,T]$ of the standard point process is equal to the MLE on $[0,nT]$ of the time-transformed point process.

Lemma C3. *We have that a.s.*

$$\widehat{\overline{\theta}}_n = \widehat{\theta}_{T,n}$$

Proof. By the time-transformed process form and Lemma C2, the result follows. □

C.3 Adaptation of some lemmas in the mixture of generalized gamma kernels case

The following lemma shows that the time-transformed $\overline{\mathbf{F}}_n$-intensity, together with its first three derivatives, are in L^p for any $p \in \mathbb{N}$ with $p \geq 2$. This corresponds to Condition **[A2] (i)** (p. 1804) in Clinet and Yoshida (2017). This extends Lemma A.5 (p. 1833) in Clinet and Yoshida (2017) which is restricted to the exponential kernel case to the mixture of generalized gamma kernels case.

Lemma C4. *We assume that Condition **[A]** holds. For any $i = 1, \dots, p$, the $\overline{\mathbf{F}}_n$-intensity process and their first derivatives satisfy for any $p \in \mathbb{N}$, $p \geq 2$,*

$$\sup_{n \in \mathbb{N}, t \in [0, nT]} \sum_{l=0}^{3} || \sup_{\theta \in \Theta} | \partial_\theta^l \overline{\lambda}_n^{(i)}(t, \theta) | ||_p < +\infty$$

Proof. Without loss of generality, we will show the statement only for integers of the form 2^p with $p \in \mathbb{N}^*$. By Lemma C1, we have that

$$\overline{\lambda}_n^{(i)}(t, \theta) = \nu^{(i)} + \sum_{j=1}^{d} \int_0^{t^-} h^{(i,j)}((t-s), \theta_{ker}^{(i,j)}) d\overline{N}_{s,n}^{(j)}.$$

Given that $\nu^{(i)}$ is bounded above as Θ is itself assumed to be bounded and bounded below by Condition **[A] (i)**, there is no loss of generality assuming that $\nu^{(i)} = 0$ in the rest of this proof. Thus, it remains to show for $i, j = 1, \dots, d$ and $l = 0, \dots, 3$ that

$$\sup_{n \in \mathbb{N}, t \in [0, nT]} \mathbb{E}\Big[\Big| \int_0^t \sup_{\theta \in \Theta} | \partial_\theta^l h^{(i,j)}((t-s), \theta_{ker}^{(i,j)}) d\overline{N}_{s,n}^{(j)} | \Big|^{2^p} \Big] < +\infty.$$

Applying the triangular inequality, it is then sufficient to show that

$$\sup_{n\in\mathbb{N},t\in[0,nT]} \mathbb{E}\Big[\Big|\int_0^t \sup_{\theta\in\Theta} \mid \partial_\theta^l h^{(i,j)}((t-s),\theta_{ker}^{(i,j)}) \mid d\overline{N}_{s,n}^{(j)}\Big|^{2^p}\Big] < +\infty.$$

Because the term inside the integral is positive, it is sufficient to show that uniformly in $n\in\mathbb{N}$ we have

$$\mathbb{E}\Big[\Big|\int_0^{nT} \sup_{\theta\in\Theta} \mid \partial_\theta^l h^{(i,j)}((nT-s),\theta_{ker}^{(i,j)}) \mid d\overline{N}_{s,n}^{(j)}\Big|^{2^p}\Big] < +\infty.$$

In view of Equation (5), this can be rewritten as

$$\mathbb{E}\Big[\Big|\int_0^{nT} \sup_{\theta\in\Theta}\Big|\partial_\theta^l \sum_{k=1}^{K^{(i,j)}} P_k^{(i,j)}((nT-s),\theta_{ker}^{(i,j)}) \exp(-((nT-s)/\beta_k^{(i,j)})^{p_k^{(i,j)}})\Big|d\overline{N}_{s,n}^{(i)}\Big|^{2^p}\Big] < +\infty,$$

where $P_k^{(i,j)}(t,\theta_{ker}^{(i,j)})$ is defined as

$$P_k^{(i,j)}(t,\theta_{ker}^{(i,j)}) = \alpha_k^{(i,j)} \frac{p_k^{(i,j)} t^{(D_k^{(i,j)}-1)}}{(\beta_k^{(i,j)})^{D_k^{(i,j)}}\Gamma(D_k^{(i,j)}/p_k^{(i,j)})}. \tag{C9}$$

Since Θ is assumed to be bounded, there exists $\beta_+\in\mathbb{R}_+^*$ such that we have uniformly $\beta_k^{(i,j)}\le\beta_+$. Then, an application of that inequality along with Condition **[A](iii)** yields

$$\begin{aligned}
&\mathbb{E}\Big[\Big|\int_0^{nT} \sup_{\theta\in\Theta}\Big|\partial_\theta^l \sum_{k=1}^{K^{(i,j)}} P_k^{(i,j)}((nT-s),\theta_{ker}^{(i,j)}) \exp(-((nT-s)/\beta_k^{(i,j)})^{p_k^{(i,j)}})\Big|d\overline{N}_{s,n}^{(i)}\Big|^{2^p}\Big]\\
&\quad\le \mathbb{E}\Big[\Big|\int_0^{nT} e^{-((nT-s)/\beta_+)^{p_-}} \sup_{\theta\in\Theta}\Big|\partial_\theta^l \sum_{k=1}^{K^{(i,j)}} P_k^{(i,j)}((nT-s),\theta_{ker}^{(i,j)})\Big|d\overline{N}_{s,n}^{(i)}\Big|^{2^p}\Big].
\end{aligned}$$

In what follows, we will show by induction that

$$\begin{aligned}
&\mathbb{E}\Big[\Big|\int_0^{nT} e^{-((nT-s)/\beta_+)^{p_-}} \sup_{\theta\in\Theta}\Big|\partial_\theta^l \sum_{k=1}^{K^{(i,j)}} P_k^{(i,j)}((nT-s),\theta_{ker}^{(i,j)})\Big|d\overline{M}_{t,n}^{(i)}\Big|^{2^p}\Big] \\
\le\ & K_p\mathbb{E}\Big[\int_0^{Tn} e^{-p((nT-s)/\beta_+)^{p_-}} \sup_{\theta\in\Theta}\Big|\partial_\theta^l \sum_{k=1}^{K^{(i,j)}} P_k^{(i,j)}((nT-s),\theta_{ker}^{(i,j)})\Big|^{2p}\overline{\lambda}_n^{(i)}(t,\theta^*)dt\Big]\\
&+K_p\mathbb{E}\Big[\Big|\int_0^{Tn} e^{-((nT-s)/\beta_+)^{p_-}} \sup_{\theta\in\Theta}\Big|\partial_\theta^l \sum_{k=1}^{K^{(i,j)}} P_k^{(i,j)}((nT-s),\theta_{ker}^{(i,j)})\Big|^2dt\Big|^{2^{p-1}}\Big].
\end{aligned}\tag{C10}$$

We define

$$f(t) = e^{-((nT-s)/\beta_+)^{p_-}} \sup_{\theta\in\Theta}\Big|\partial_\theta^l \sum_{k=1}^{K^{(i,j)}} P_k^{(i,j)}((nT-s),\theta_{ker}^{(i,j)})\Big|.$$

We consider the case $p = 1$. We can calculate

$$\begin{aligned}
\mathbb{E}\Big[\Big|\int_0^{Tn} f(t)d\overline{M}_{t,n}^{(i)}\Big|^2\Big] &= \mathbb{E}\Big[\int_0^{Tn} f(t)^2 d\langle \overline{M}_n^{(i)}, \overline{M}_n^{(i)}\rangle_t\Big] \\
&= \mathbb{E}\Big[\int_0^{Tn} f(t)^2 \overline{\lambda}_n^{(i)}(t,\theta^*)dt\Big],
\end{aligned}$$

where the first equality was obtained with Itô isometry for point process martingales and the second equality is due to Equation (C6). This implies that Inequality (C10) holds in the case $p = 1$. We investigate now the case $p \geq 2$. By the Burkholder-Davis-Gundy inequality (see, e.g., Equation (2.1.32) in Jacod and Protter (2011)), we obtain

$$\begin{aligned}
\mathbb{E}\Big[\Big|\int_0^{Tn} f(t)d\overline{M}_{t,n}^{(i)}\Big|^{2^p}\Big] &\leq D_p\mathbb{E}\Big[\Big|\int_0^{Tn} f(t)^2 d\overline{N}_{t,n}^{(i)}\Big|^{2^{p-1}}\Big] \\
&\leq 2^{p-1}D_p\mathbb{E}\Big[\Big|\int_0^{Tn} f(t)^2 d\overline{M}_{t,n}^{(i)}\Big|^{2^{p-1}}\Big] \\
&\quad +2^{p-1}D_p\mathbb{E}\Big[\Big|\int_0^{Tn} f(t)^2 \overline{\lambda}_n^{(i)}(t,\theta^*)dt\Big|^{2^{p-1}}\Big].
\end{aligned}$$

Now, an induction argument yields that for some constant $Q_p > 0$:

$$\mathbb{E}\Big[\Big|\int_0^{Tn} f(t)d\overline{M}_{t,n}^{(i)}\Big|^{2^p}\Big] \leq Q_p\sum_{q=1}^{p}\mathbb{E}\Big[\Big|\int_0^{Tn} f(t)^{2q}\overline{\lambda}_n^{(i)}(t,\theta^*)dt\Big|^{2^{p-q}}\Big].$$

If we can show that for any $q = 1, \ldots, p$ we have

$$\begin{aligned}
\Big|\int_0^{Tn} f(t)^{2q}\overline{\lambda}_n^{(i)}(t,\theta^*)dt\Big|^{2^{p-q}} &\leq \int_0^{Tn} f(t)^{2p}\overline{\lambda}_n^{(i)}(t,\theta^*)dt \\
&\quad +\Big|f(t)^2dt\Big|^{2^{p-1}},
\end{aligned} \tag{C11}$$

then Inequality (C10) is shown with $K_p = pQ_p$. We prove now that Inequality (C11) holds. We write

$$g_n^{(i)}(t) = \frac{f(t)}{\Big|\int_0^{Tn} f(t)^2\overline{\lambda}_n^{(i)}(t,\theta^*)dt\Big|^{\frac{1}{2}}}. \tag{C12}$$

We then obtain that

$$\begin{aligned}\Big|\int_0^{Tn} g_n^{(i)}(t)^{2q}\overline{\lambda}_n^{(i)}(t,\theta^*)dt\Big|^{2^{p-q}} &= \Big|\int_0^{Tn} g_n^{(i)}(t)^{2q-2}\mu_n^{(i)}(dt)\Big|^{2^{p-q}} \\ &\leq \Big|\int_0^{Tn} g_n^{(i)}(t)^{2p}\overline{\lambda}_n^{(i)}(t,\theta^*)dt\Big|^{\frac{2^{p-1}-2^{-q}}{2^{p-1}-1}} \\ &\leq \Big|\int_0^{Tn} g_n^{(i)}(t)^{2p}\overline{\lambda}_n^{(i)}(t,\theta^*)dt\Big|^{\frac{2^{p-1}-2^{-q}}{2^{p-1}-1}} + 1, \qquad \text{(C13)}\end{aligned}$$

where the equality is due to the fact that $\mu_n^{(i)}(dt) := g_n^{(i)}(t)^{2q}\overline{\lambda}_n^{(i)}(t,\theta^*)dt$ is a probability measure on $[0,Tn]$ and the first inequality comes from Jensen's inequality. If we reexpress Inequality (C13) with Definition (C12), we can show Inequality (C10).

We obviously have that

$$\begin{aligned}&\mathbb{E}\Big[\Big|\int_0^{nT} e^{-((nT-s)/\beta_+)^{p-}}\sup_{\theta\in\Theta}\Big|\partial_\theta^l\sum_{k=1}^{K^{(i,j)}} P_k^{(i,j)}((nT-s),\theta_{ker}^{(i,j)})\Big|d\overline{N}_{s,n}^{(i)}\Big|^{2^p}\Big] \\ &= \mathbb{E}\Big[\Big|\int_0^{nT} e^{-((nT-s)/\beta_+)^{p-}}\sup_{\theta\in\Theta}\Big|\partial_\theta^l\sum_{k=1}^{K^{(i,j)}} P_k^{(i,j)}((nT-s),\theta_{ker}^{(i,j)})\Big|d\overline{M}_{s,n}^{(i)}\Big|^{2^p}\Big] \\ &+\Big(\mathbb{E}\Big[\Big|\int_0^{nT} e^{-((nT-s)/\beta_+)^{p-}}\sup_{\theta\in\Theta}\Big|\partial_\theta^l\sum_{k=1}^{K^{(i,j)}} P_k^{(i,j)}((nT-s),\theta_{ker}^{(i,j)})\Big|d\overline{N}_{s,n}^{(i)}\Big|^{2^p}\Big] \\ &-\mathbb{E}\Big[\Big|\int_0^{nT} e^{-((nT-s)/\beta_+)^{p-}}\sup_{\theta\in\Theta}\Big|\partial_\theta^l\sum_{k=1}^{K^{(i,j)}} P_k^{(i,j)}((nT-s),\theta_{ker}^{(i,j)})\Big|d\overline{M}_{s,n}^{(i)}\Big|^{2^p}\Big]\Big).\end{aligned}$$

We use now Inequality (C10), and we obtain that

$$\begin{aligned}&\mathbb{E}\Big[\Big|\int_0^{nT} e^{-((nT-s)/\beta_+)^{p-}}\sup_{\theta\in\Theta}\Big|\partial_\theta^l\sum_{k=1}^{K^{(i,j)}} P_k^{(i,j)}((nT-s),\theta_{ker}^{(i,j)})\Big|d\overline{N}_{t,n}^{(i)}\Big|^{2^p}\Big] \\ \leq\;& K_p\mathbb{E}\Big[\int_0^{Tn} e^{-p((nT-s)/\beta_+)^{p-}}\sup_{\theta\in\Theta}\Big|\partial_\theta^l\sum_{k=1}^{K^{(i,j)}} P_k^{(i,j)}((nT-s),\theta_{ker}^{(i,j)})\Big|^{2p}\overline{\lambda}_n^{(i)}(t,\theta^*)dt\Big] \\ &+K_p\mathbb{E}\Big[\Big|\int_0^{Tn} e^{-((nT-s)/\beta_+)^{p-}}\sup_{\theta\in\Theta}\Big|\partial_\theta^l\sum_{k=1}^{K^{(i,j)}} P_k^{(i,j)}((nT-s),\theta_{ker}^{(i,j)})\Big|^2 dt\Big|^{2^{p-1}}\Big] \\ &+\Big(\mathbb{E}\Big[\Big|\int_0^{nT} e^{-((nT-s)/\beta_+)^{p-}}\sup_{\theta\in\Theta}\Big|\partial_\theta^l\sum_{k=1}^{K^{(i,j)}} P_k^{(i,j)}((nT-s),\theta_{ker}^{(i,j)})\Big|d\overline{N}_{s,n}^{(i)}\Big|^{2^p}\Big] \\ &-\mathbb{E}\Big[\Big|\int_0^{nT} e^{-((nT-s)/\beta_+)^{p-}}\sup_{\theta\in\Theta}\Big|\partial_\theta^l\sum_{k=1}^{K^{(i,j)}} P_k^{(i,j)}((nT-s),\theta_{ker}^{(i,j)})\Big|d\overline{M}_{s,n}^{(i)}\Big|^{2^p}\Big]\Big) \\ :=\;& I_n + II_n + III_n.\end{aligned}$$

We now show that $I_n < \infty$ uniformly in $n \in \mathbb{N}$. We can calculate

$$\begin{aligned}\frac{I_n}{K_p} &= \mathbb{E}\Big[\int_0^{Tn} e^{-p((nT-s)/\beta_+)^{p-}} \sup_{\theta \in \Theta} \Big|\partial_\theta^l \sum_{k=1}^{K^{(i,j)}} P_k^{(i,j)}((nT-s), \theta_{ker}^{(i,j)})\Big|^{2p} \overline{\lambda}_n^{(i)}(t, \theta^*) dt\Big] \\ &= \int_0^{Tn} e^{-p((nT-s)/\beta_+)^{p-}} \sup_{\theta \in \Theta} \Big|\partial_\theta^l \sum_{k=1}^{K^{(i,j)}} P_k^{(i,j)}((nT-s), \theta_{ker}^{(i,j)})\Big|^{2p} \mathbb{E}\Big[\overline{\lambda}_n^{(i)}(t, \theta^*)\Big] dt,\end{aligned}$$

where we use the definition of I_n in the first equality, and the second equality is due to Tonelli's theorem along with the fact that $\mathbb{E}[aX] = a\mathbb{E}[X]$ for any random variable X and any nonrandom $a \in \mathbb{R}$. It remains to prove that uniformly in $n \in \mathbb{N}$ we have

$$\int_0^{Tn} e^{-p((nT-s)/\beta_+)^{p-}} \sup_{\theta \in \Theta} \Big|\partial_\theta^l \sum_{k=1}^{K^{(i,j)}} P_k^{(i,j)}((nT-s), \theta_{ker}^{(i,j)})\Big|^{2p} \mathbb{E}\Big[\overline{\lambda}_n^{(i)}(t, \theta^*)\Big] dt < \infty.$$

By the kernel definition (see Equation (5) and Equation (C9)), since Θ is bounded itself and by Condition **[A](iii)-(iv)** we have that $P_k^{(i,j)}(t,.)$ is in $C^3(\overline{\Theta})$. We can deduce that uniformly in $n \in \mathbb{N}$ we have

$$\sup_{\theta \in \Theta} \Big|\partial_\theta^l \sum_{k=1}^{K^{(i,j)}} P_k^{(i,j)}((nT-s), \theta_{ker}^{(i,j)})\Big|^{2p} \leq C.$$

The proof of $I_n < \infty$ amounts to showing that uniformly in $n \in \mathbb{N}$ we have

$$\int_0^{Tn} e^{-p((nT-s)/\beta_+)^{p-}} \mathbb{E}\Big[\overline{\lambda}_n^{(i)}(t, \theta^*)\Big] dt < \infty.$$

By Condition **[A] (v)** and Definition (C3), we obtain uniformly in $n \in \mathbb{N}$ and in $s \in [0, nT]$ that

$$\mathbb{E}\Big[\overline{\lambda}_n^{(i)}(t, \theta^*)\Big] \leq C.$$

We thus obtain that

$$I_n \leq C \int_0^{Tn} e^{-p((nT-s)/\beta_+)^{p-}} dt.$$

By a change of variable in the integral, we can deduce that

$$\int_0^{nT} e^{-p((nT-s)/\beta_+)^{p-}} dt = \int_0^{nT} e^{-p(u/\beta_+)^{p-}} du.$$

The obtained term can be dominated uniformly in $n \in \mathbb{N}$ by

$$\begin{aligned}\int_0^{nT} e^{-p(u/\beta_+)^{p_-}} du &\leq \int_0^{\infty} e^{-p(u/\beta_+)^{p_-}} du \\ &= C_1.\end{aligned} \tag{C14}$$

We have thus proven that $I_n < \infty$ uniformly in $n \in \mathbb{N}$. The proof for II_n and III_n follows with the same arguments. □

In what follows, we provide the definition of ergodicity in our time-transformed framework. This extends Definition 3.1 in Clinet and Yoshida (2017) which does not consider any time transformation. See also Kwan (2023) for a similar time-transformed framework.

Definition C1. *(ergodicity) We assume that $(E, \mathbf{B}(E))$ is a Borel space, and $X_n : \Omega \times [0, nT] \to E$ a sequence of stochastic processes adapted to the time-transformed filtration. We say that X_n is ergodic if there exists a mapping $\pi : C_b(E, \mathbb{R}) \to \mathbb{R}$ such that for any $\psi \in C_b(E, \mathbb{R})$ we have*

$$\frac{1}{nT} \int_0^{nT} \psi(X_{s,n}) ds \to^{\mathbb{P}} \pi(\psi).$$

The following definition introduces the notion of mixing to our time-transformed framework. This extends the definition from Section 3.4 in Clinet and Yoshida (2017) which does not consider any time transformation. See also Kwan (2023) for a similar time-transformed framework.

Definition C2. *(mixing) We assume that $(E, \mathbf{B}(E))$ is a Borel space, and $X_n : \Omega \times [0, nT] \to E$ a sequence of stochastic processes adapted to the time-transformed filtration. We say that X_n is C-mixing, for some set of functions C from E to $\mathbb{R}$, if for any ϕ, $\psi \in C$, the following convergence holds*

$$\rho_u := \sup_{n \in \mathbb{N}^*} \sup_{s \in [0, nT-u]} | \operatorname{Cov}[\phi(X_{s,n}), \psi(X_{s+u,n})] | \to 0 \text{ as } u \to +\infty.$$

The following proposition states that $X_n^{(i)}$ is mixing in the sense of Definition C2, stable, and ergodic in the sense of Definition C1. This corresponds to Condition **[A3]** (p. 1805) and Condition **[M1]** (p. 1815) in Clinet and Yoshida (2017). This extends Lemma 3.16 (p. 1815) and Lemma A.6 (p. 1834) in Clinet and Yoshida (2017) which are restricted to the exponential kernel case to the mixture of generalized gamma kernels case. This also extends their general machinery by proving first that $X_n^{(i)}$ is mixing and stable, and then this implies its ergodicity. Finally, this extends Kwan (2023) who considers the non-exponential kernel case but can only show the ergodicity of $(\overline{\lambda}_n^{(i)}(., \theta^*), \overline{\lambda}_n^{(i)}(., \theta))$ but not the ergodicity of $X_n^{(i)}$. The stability is a direct consequence of Theorem 1 and Lemma 4 in Brémaud and Massoulié (1996), along with Condition **[A] (v)**.

Proposition C1. *We assume that Condition **[A] (iii) and (v)** hold. For any $i = 1, \ldots, d$ and any $\theta \in \Theta$, $X_n^{(i)}$ is:*

(i) $C_b(E, \mathbb{R})$-mixing in the sense of Definition C2.

(ii) stable, i.e. there exists a E-valued random variable $\overline{\lambda}_{lim}^{(i)}$ such that

$$X_{nT,n}^{(i)} \to^{\mathcal{D}} (\overline{\lambda}_{lim}^{(i)}(\theta^*), \overline{\lambda}_{lim}^{(i)}(\theta), \partial_\theta \overline{\lambda}_{lim}^{(i)}(\theta)).$$

(iii) ergodic in the sense of Definition C1, i.e. there exists a mapping $\pi_{\theta^*}^{(i)} : C_b(E, \mathbb{R}) \times \Theta \to \mathbb{R}$ such that for any $(\psi, \theta) \in C_b(E, \mathbb{R}) \times \Theta$ we have $\frac{1}{nT} \int_0^{nT} \psi(X_{s,n}^{(i)}) ds \to^{\mathbb{P}} \pi_{\theta^*}^{(i)}(\psi, \theta)$, where $\pi_{\theta^*}^{(i)}(\psi, \theta) = \mathbb{E}[\psi(\overline{\lambda}_{lim}^{(i)}(\theta^*), \overline{\lambda}_{lim}^{(i)}(\theta), \partial_\theta \overline{\lambda}_{lim}^{(i)}(\theta))]$.

Proof. We first show (i). We first define the truncated version of $X_{s,n}^{(i)}$ at time $t \leq s$ as

$$\widetilde{X}_{t,s,n}^{(i)} := (\overline{\lambda}_n^{(i)}(s, \theta^*), \sum_{j=1}^d \int_t^{s^-} h^{(i,j)}(s-u, \theta) d\overline{N}_{u,n}^{(i)}, \sum_{j=1}^d \int_t^{s^-} \partial_\theta h^{(i,j)}(s-u, \theta) d\overline{N}_{u,n}^{(i)}).$$

By considering ϕ, $\psi \in C_b(E, \mathbb{R})$, we can reexpress $\rho_u^{(i)}$ as

$$\begin{aligned}
\rho_u^{(i)} &= \sup_{n \in \mathbb{N}^*} \sup_{s \in [0, nT-u]} | \operatorname{Cov}[\phi(X_{s,n}^{(i)}), \psi(X_{s+u,n}^{(i)})] | \\
&= \sup_{n \in \mathbb{N}^*} \sup_{s \in [0, nT-u]} | \operatorname{Cov}[\phi(X_{s,n}^{(i)}), \psi(X_{s+u,n}^{(i)}) - \psi(\widetilde{X}_{s+v,s+u,n}^{(i)}) + \psi(\widetilde{X}_{s+v,s+u,n}^{(i)})] |,
\end{aligned}$$

where we use Definition C2 in the first equality, and we have $v \leq s - u$ in the second equality. Using the triangular inequality, we can dominate $\rho_u^{(i)}$ as

$$\begin{aligned}\rho_u^{(i)} &\leq \sup_{n \in \mathbb{N}^*} \sup_{s \in [0, nT-u]} \mid \mathrm{Cov}[\phi(X_{s,n}^{(i)}), \psi(X_{s+u,n}^{(i)}) - \psi(\widetilde{X}_{s+v,s+u,n}^{(i)})] \mid \\ &\quad + \sup_{n \in \mathbb{N}^*} \sup_{s \in [0, nT-u]} \mid \mathrm{Cov}[\phi(X_{s,n}^{(i)}), \psi(\widetilde{X}_{s+v,s+u,n}^{(i)})] \mid \\ &:= I_u + II_u.\end{aligned}$$

Since Θ is assumed to be bounded, there exists $\beta_+ \in \mathbb{R}_+^*$ such that we have uniformly $\beta_k^{(i,j)} \leq \beta_+$. Then, an application of that inequality along with Condition **[A](iii)** yields that the intensity process is decreasing exponentially. Thus, we can deduce by similar arguments from the proof of Lemma A.6 (pp. 1834-1836) in Clinet and Yoshida (2017) that $I_u \to 0$ and $II_u \to 0$. This in turn implies that

$$\rho_u^{(i)} \to 0 \text{ as } u \to +\infty.$$

The stability (ii) is a direct consequence of Theorem 1 and Lemma 4 in Brémaud and Massoulié (1996), along with Condition **[A] (v)**.

We now show the ergodicity (iii). For $\psi \in C_b(E, \mathbb{R})$ we define $V_n^{(i)}(\psi, \theta)$ as

$$V_n^{(i)}(\psi, \theta) = \frac{1}{nT} \int_0^{nT} \psi(X_{s,n}^{(i)}) ds. \tag{C15}$$

We consider $\pi^{(i)}(\psi, \theta) = \mathbb{E}[\psi(\overline{\lambda}_{lim}^{(i)}(\theta^*), \overline{\lambda}_{lim}^{(i)}(\theta), \partial_\theta \overline{\lambda}_{lim}^{(i)}(\theta))]$. Establishing ergodicity amounts to showing the convergence in probability, i.e. $V_n^{(i)}(\psi, \theta) \to^{\mathbb{P}} \pi^{(i)}(\psi, \theta)$. In what follows, we show a stronger statement, i.e. the L^2-convergence. We calculate

$$\begin{aligned}\mathbb{E}[(V_n^{(i)}(\psi, \theta) - \pi^{(i)}(\psi, \theta))^2] &= \mathrm{Var}[V_n^{(i)}(\psi, \theta)] + (\mathbb{E}[V_n^{(i)}(\psi, \theta)] - \pi^{(i)}(\psi, \theta))^2 \\ &:= I_n + II_n\end{aligned}$$

where the equality is due to the fact that for any random variable X and any nonrandom $a \in \mathbb{R}$

we have $\mathbb{E}[(X-a)^2] = \mathrm{Var}[X] + (\mathbb{E}[X]-a)^2$. For the first term, we have that

$$\begin{aligned}
I_n &= \mathrm{Var}[V_n^{(i)}(\psi,\theta)] \\
&= \mathrm{Var}\left[\frac{1}{nT}\int_0^{nT}\psi(X_{s,n}^{(i)})ds\right] \\
&= \frac{1}{n^2T^2}\mathrm{Var}\left[\int_0^{nT}\psi(X_{s,n}^{(i)})\right] \\
&= \frac{1}{n^2T^2}\mathrm{Var}\left[\lim_{K\to\infty}\frac{nT}{K}\sum_{k=0}^{K-1}\psi(X_{knT/K,n}^{(i)})\right] \\
&= \frac{1}{n^2T^2}\lim_{K\to\infty}\mathrm{Var}\left[\frac{nT}{K}\sum_{k=0}^{K-1}\psi(X_{knT/K,n}^{(i)})\right] \\
&= \frac{1}{n^2T^2}\lim_{K\to\infty}\frac{n^2T^2}{K^2}\mathrm{Var}\left[\sum_{k=0}^{K-1}\psi(X_{knT/K,n}^{(i)})\right] \\
&= \frac{1}{n^2T^2}\lim_{K\to\infty}\frac{n^2T^2}{K^2}\sum_{k=0}^{K-1}\sum_{l=0}^{K-1}\mathrm{Cov}\left[\psi(X_{knT/K,n}^{(i)}),\psi(X_{lnT/K,n}^{(i)})\right] \\
&= \frac{1}{n^2T^2}\int_0^{nT}\int_0^{nT}\mathrm{Cov}\left[\psi(X_{s,n}^{(i)}),\psi(X_{u,n}^{(i)})\right]dsdu,
\end{aligned}$$

where the second equality is obtained via Definition (C15), the third equality and the sixth equality are due to the fact that for any nonrandom $a \in \mathbb{R}$ and any random variable X we have $\mathrm{Var}[aX] = a^2\,\mathrm{Var}[X]$, we used the approximation of Riemann sum in the fourth equality and eight equality, the fifth equality is an application of dominated convergence theorem, and the seventh equality corresponds to Bienayme's identity. By Definition C1 and results obtained in (i), we can bound I_n as

$$I_n \le \frac{1}{n^2T^2}\int_0^{nT}\int_0^{nT}\rho_{|s-u|}^{(i)}dsdu.$$

Then, splitting the integral into two terms leads to

$$I_n \le \frac{1}{n^2T^2}\int_0^{nT}\int_0^{nT}\rho_{|s-u|}^{(i)}\mathbf{1}_{\{|s-u|\le\sqrt{nT}\}}dsdu + \frac{1}{n^2T^2}\int_0^{nT}\int_0^{nT}\rho_{|s-u|}^{(i)}\mathbf{1}_{\{|s-u|>\sqrt{nT}\}}dsdu$$

Since there exists $\rho^{(i)}_{\max} > 0$ such that for any $t \geq 0$ we have $\rho^{(i)}_t \leq \rho^{(i)}_{\max}$, we can deduce that

$$\begin{aligned}
\frac{1}{n^2T^2}\int_0^{nT}\int_0^{nT}\rho^{(i)}_{|s-u|}\mathbf{1}_{\{|s-u|\leq\sqrt{nT}\}}dsdu &\leq \frac{\rho^{(i)}_{\max}}{n^2T^2}\int_0^{nT}\int_0^{nT}\mathbf{1}_{\{|s-u|\leq\sqrt{nT}\}}dsdu \\
&= O(\frac{1}{\sqrt{nT}}) \\
&\to 0.
\end{aligned}$$

We also have that

$$\begin{aligned}
\frac{1}{n^2T^2}\int_0^{nT}\int_0^{nT}\rho^{(i)}_{|s-u|}\mathbf{1}_{\{|s-u|>\sqrt{nT}\}}dsdu &\leq \frac{\sup\limits_{y>\sqrt{nT}}\rho^{(i)}_y}{n^2T^2}\int_0^{nT}\int_0^{nT}\mathbf{1}_{\{|s-u|>\sqrt{nT}\}}dsdu \\
&\leq \sup_{y>\sqrt{nT}}\rho^{(i)}_y.
\end{aligned}$$

Since $\rho_u \to^{u\to\infty} 0$ by (i), we also deduce that

$$\sup_{y>\sqrt{nT}}\rho^{(i)}_y \to 0.$$

This implies that

$$\frac{1}{n^2T^2}\int_0^{nT}\int_0^{nT}\rho^{(i)}_{|s-u|}\mathbf{1}_{\{|s-u|>\sqrt{nT}\}}dsdu \to 0,$$

and thus $I_n \to 0$. For the second term, we know by (ii) that $X^{(i)}_{s,n} \to^{\mathcal{D}} (\overline{\lambda}^{(i)}_{lim}(\theta^*), \overline{\lambda}^{(i)}_{lim}(\theta), \partial_\theta\overline{\lambda}^{(i)}_{lim}(\theta))$. In particular, convergence in distribution implies convergence in expectation of any bounded function, thus we obtain that $\mathbb{E}[\psi(X^{(i)}_{s,n})] \to \mathbb{E}[\psi(\overline{\lambda}^{(i)}_{lim}(\theta^*), \overline{\lambda}^{(i)}_{lim}(\theta), \partial_\theta\overline{\lambda}^{(i)}_{lim}(\theta))]$. This can be reexpressed as $\mathbb{E}[\psi(X^{(i)}_{s,n})] \to \pi^{(i)}_{\theta^*}(\psi,\theta)$. In particular, this implies that $(\mathbb{E}[\psi(X^{(i)}_{s,n})] - \pi^{(i)}_{\theta^*}(\psi,\theta))^2 \to 0$, i.e. $II_n \to 0$. □

Since the functions that we will be using in our proofs will not necessarily be bounded, we need to extend from $C_b(E,\mathbb{R})$ to $C_\uparrow(E,\mathbb{R})$ the space of functions in which the ergodicity holds. We also give a more explicit form to the mapping $\pi^{(i)}_{\theta^*}(\psi,\theta)$. The following proposition extends Proposition 3.8 (pp. 1806-1807) in Clinet and Yoshida (2017). The proof follows the arguments from the proof of Proposition 3.8 (pp. 1822-1824) in Clinet and Yoshida (2017).

Proposition C2. *We assume that Condition **[A]** holds. Then, for any $\theta \in \Theta$ and for any $i = 1, \dots, d$, the following properties hold*

(i) The ergodicity, i.e. Proposition C1(iii), still holds for any $\psi \in C_{\uparrow}(E, \mathbb{R})$. In particular, the mapping $\pi^{(i)}_{\theta^*}(., \theta)$ can be extended to $C_{\uparrow}(E, \mathbb{R})$. Moreover, for any $\psi \in C_{\uparrow}(E, \mathbb{R})$ the convergence is uniform in θ.

(ii) There exists a probability measure $\Pi^{(i)}_{\theta^*}$ on $(E, \mathbf{B}(E))$ such that for any $\psi \in C_{\uparrow}(E, \mathbb{R})$, we have $\pi^{(i)}_{\theta^*}(\psi, \theta) = \int_E \psi(x) \Pi^{(i)}_{\theta^*}(dx)$. In particular, $C_{\uparrow}(E, \mathbb{R}) \subset \mathbb{L}^1(\Pi^{(i)}_{\theta^*})$.

Proof. We can use the arguments from the proof of Proposition 3.8 (pp. 1822-1824) in Clinet and Yoshida (2017). □

C.4 Proofs of consistency

We define

$$\overline{\mathbb{Y}}_n(\theta) = \frac{1}{nT}(\bar{l}_n(\theta) - \bar{l}_n(\theta^*)) \tag{C16}$$

and also the asymptotic rescaled of the time-transformed log likelihood as

$$\overline{\mathbb{Y}}(\theta) = \sum_{i=1}^{d} \int_E (\log(\frac{v}{u})u - (v - u))\Pi^{(i)}_{\theta^*}(du, dv, dw). \tag{C17}$$

In the following lemma, we will prove that $\overline{\mathbb{Y}}_n(\theta)$ goes to $\overline{\mathbb{Y}}(\theta)$ uniformly in $\theta \in \Theta$ and in probability. This extends Lemma 3.10 (p. 1807) in Clinet and Yoshida (2017) which is restricted to the large-T asymptotics to the in-fill asymptotics.

Lemma C5. *We assume that Condition **[A]** holds. We have that*

$$\sup_{\theta \in \Theta} | \overline{\mathbb{Y}}_n(\theta) - \overline{\mathbb{Y}}(\theta) | \to^{\mathbb{P}} 0.$$

Proof. We can rewrite $\overline{\mathbb{Y}}_n(\theta)$ as

$$\begin{aligned}
\overline{\mathbb{Y}}_n(\theta) &= \frac{1}{nT}(\bar{l}_n(\theta) - \bar{l}_n(\theta^*)) \\
&= \frac{1}{nT}\Big(\sum_{i=1}^{d}\int_0^{Tn}\log(\overline{\lambda}_n^{(i)}(t,\theta))d\overline{N}_{t,n}^{(i)} - \sum_{i=1}^{d}\int_0^{Tn}\overline{\lambda}_n^{(i)}(t,\theta)dt \\
&\quad -\sum_{i=1}^{d}\int_0^{Tn}\log(\overline{\lambda}_n^{(i)}(t,\theta^*))d\overline{N}_{t,n}^{(i)} + \sum_{i=1}^{d}\int_0^{Tn}\overline{\lambda}_n^{(i)}(t,\theta^*)dt\Big) \\
&= \frac{1}{nT}\sum_{i=1}^{d}\int_0^{Tn}\log\Big(\frac{\overline{\lambda}_n^{(i)}(t,\theta)}{\overline{\lambda}_n^{(i)}(t,\theta^*)}\Big)d\overline{N}_{t,n}^{(i)} - \frac{1}{nT}\sum_{i=1}^{d}\int_0^{Tn}(\overline{\lambda}_n^{(i)}(t,\theta) - \overline{\lambda}_n^{(i)}(t,\theta^*))dt \\
&= \frac{1}{nT}\sum_{i=1}^{d}\int_0^{Tn}\log\Big(\frac{\overline{\lambda}_n^{(i)}(t,\theta)}{\overline{\lambda}_n^{(i)}(t,\theta^*)}\Big)d\overline{M}_{t,n}^{(i)} \\
&\quad -\frac{1}{nT}\sum_{i=1}^{d}\int_0^{Tn}\Big(\overline{\lambda}_n^{(i)}(t,\theta) - \overline{\lambda}_n^{(i)}(t,\theta^*) - \log\Big(\frac{\overline{\lambda}_n^{(i)}(t,\theta)}{\overline{\lambda}_n^{(i)}(t,\theta^*)}\Big)\overline{\lambda}_n^{(i)}(t,\theta^*)\Big)dt \\
&:= \sum_{i=1}^{d} I_n^{(i)}(nT,\theta) + \sum_{i=1}^{d} II_n^{(i)}(nT,\theta),
\end{aligned}$$

where we use Equation (C16) in the first equality, Equation (C8) in the second equality, algebraic manipulation in the third equality, Equation (C6) and algebraic manipulation in the fourth equality.

We first show that the martingale term disappears uniformly asymptotically in probability, i.e. that

$$\sup_{\theta\in\Theta}\Big|\sum_{i=1}^{d} I_n^{(i)}(nT,\theta)\Big| \to^{\mathbb{P}} 0.$$

As an application of Lemma C4 along with Condition **[A]**, for any $i = 1,\dots,d$ and any $\theta \in \Theta$ we can deduce that $I_n^{(i)}(t,\theta)$ is an L^p-integrable martingale for any $p \in \mathbb{N}$, with $p \geq 2$. We can thus apply Sobolev's inequality (see, e.g, Theorem 4.2, part 1, case A in Kallenberg (1997)), and for some big enough $p \in \mathbb{N}$ we obtain

$$\mathbb{E}\Big[\Big|\sup_{\theta\in\Theta} I_n^{(i)}(nT,\theta)\Big|^p\Big] \leq C\Big(\int_\Theta d\theta\mathbb{E}\Big[\Big|I_n^{(i)}(nT,\theta)\Big|^p\Big] + \int_\Theta d\theta\mathbb{E}\Big[\Big|\partial\theta I_n^{(i)}(nT,\theta)\Big|^p\Big]\Big). \tag{C18}$$

The first term in the right hand-side of Equation (C18) can be bounded by

$$
\begin{aligned}
\int_\Theta d\theta \mathbb{E}\Big[\Big|I_n^{(i)}(nT,\theta)\Big|^p\Big] &\leq C \sup_{\theta\in\Theta} \mathbb{E}\Big[\Big|I_n^{(i)}(nT,\theta)\Big|^p\Big] \\
&= C \sup_{\theta\in\Theta} \mathbb{E}\Big[\Big|\frac{1}{nT}\int_0^{Tn} \log\Big(\frac{\overline{\lambda}_n^{(i)}(t,\theta)}{\overline{\lambda}_n^{(i)}(t,\theta^*)}\Big) d\overline{M}_{t,n}^{(i)}\Big|^p\Big] \\
&= C \sup_{\theta\in\Theta} \mathbb{E}\Big[\frac{1}{(nT)^p}\Big|\int_0^{Tn} \log\Big(\frac{\overline{\lambda}_n^{(i)}(t,\theta)}{\overline{\lambda}_n^{(i)}(t,\theta^*)}\Big) d\overline{M}_{t,n}^{(i)}\Big|^p\Big] \\
&\leq C \sup_{\theta\in\Theta} \mathbb{E}\Big[\frac{1}{(nT)^p}\Big|\int_0^{Tn} \log\Big(\frac{\overline{\lambda}_n^{(i)}(t,\theta)}{\overline{\lambda}_n^{(i)}(t,\theta^*)}\Big)^2 \overline{\lambda}_n^{(i)}(t,\theta^*) dt\Big|^{\frac{p}{2}}\Big] \\
&\leq C \sup_{\theta\in\Theta} \mathbb{E}\Big[\frac{1}{(nT)^{\frac{p}{2}-1}}\int_0^{Tn} \log\Big(\frac{\overline{\lambda}_n^{(i)}(t,\theta)}{\overline{\lambda}_n^{(i)}(t,\theta^*)}\Big)^p \overline{\lambda}_n^{(i)}(t,\theta^*)^{\frac{p}{2}} dt\Big] \\
&\leq C \sup_{\theta\in\Theta} \mathbb{E}\Big[\frac{1}{(nT)^{\frac{p}{2}-1}}\int_0^{Tn} \log\Big(\frac{\overline{\lambda}_n^{(i)}(t,\theta)}{\nu_-}\Big)^p \overline{\lambda}_n^{(i)}(t,\theta^*)^{\frac{p}{2}} dt\Big],
\end{aligned}
$$

where the first equality is obtained with $I_n(t,\theta)$ definition, the second inequality is a consequence to Burkholder-Davis-Gundy inequality, the third inequality comes from Jensen's inequality, and the fourth inequality is due to Condition **[A](i)-(ii)**. We can continue to bound the first term in the right hand-side of Equation (C18) by

$$
\begin{aligned}
\int_\Theta d\theta \mathbb{E}\Big[\Big|I_n^{(i)}(nT,\theta)\Big|^p\Big] &\leq \sup_{\theta\in\Theta} \mathbb{E}\Big[\frac{1}{(nT)^{\frac{p}{2}-1}}\int_0^{Tn} \log\Big(\frac{\overline{\lambda}_n^{(i)}(t,\theta)}{\nu_-}\Big)^p \overline{\lambda}_n^{(i)}(t,\theta^*)^{\frac{p}{2}} dt\Big] \\
&= \frac{1}{(nT)^{\frac{p}{2}-1}} \sup_{\theta\in\Theta} \mathbb{E}\Big[\int_0^{Tn} \log\Big(\frac{\overline{\lambda}_n^{(i)}(t,\theta)}{\nu_-}\Big)^p \overline{\lambda}_n^{(i)}(t,\theta^*)^{\frac{p}{2}} dt\Big] \\
&= \frac{1}{(nT)^{\frac{p}{2}-1}} \sup_{\theta\in\Theta} \int_0^{Tn} \mathbb{E}\Big[\log\Big(\frac{\overline{\lambda}_n^{(i)}(t,\theta)}{\nu_-}\Big)^p \overline{\lambda}_n^{(i)}(t,\theta^*)^{\frac{p}{2}}\Big] dt \\
&= \frac{1}{(nT)^{\frac{p}{2}-2}} \sup_{\theta\in\Theta} \sup_{n\in\mathbb{N}, t\in[0,nT]} \mathbb{E}\Big[\log\Big(\frac{\overline{\lambda}_n^{(i)}(t,\theta)}{\nu_-}\Big)^p \overline{\lambda}_n^{(i)}(t,\theta^*)^{\frac{p}{2}}\Big] \\
&\leq \frac{1}{(nT)^{\frac{p}{2}-2}} \sup_{\theta\in\Theta} \sup_{n\in\mathbb{N}, t\in[0,nT]} \sqrt{\mathbb{E}\Big[\log\Big(\frac{\overline{\lambda}_n^{(i)}(t,\theta)}{\nu_-}\Big)^{2p}\Big]}\sqrt{\mathbb{E}\Big[\overline{\lambda}_n^{(i)}(t,\theta^*)^p\Big]} \\
&\leq \frac{1}{(nT)^{\frac{p}{2}-2}} \sup_{\theta\in\Theta} \sup_{n\in\mathbb{N}, t\in[0,nT]} \sqrt{\mathbb{E}\Big[\Big(1+\frac{\overline{\lambda}_n^{(i)}(t,\theta)}{\nu_-}\Big)^{2p}\Big]}\sqrt{\mathbb{E}\Big[\overline{\lambda}_n^{(i)}(t,\theta^*)^p\Big]} \\
&\leq \frac{1}{(nT)^{\frac{p}{2}-2}} \sup_{n\in\mathbb{N}, t\in[0,nT]} \sqrt{\mathbb{E}\Big[\sup_{\theta\in\Theta}\Big(1+\frac{\overline{\lambda}_n^{(i)}(t,\theta)}{\nu_-}\Big)^{2p}\Big]}\sqrt{\mathbb{E}\Big[\sup_{\theta\in\Theta}\overline{\lambda}_n^{(i)}(t,\theta^*)^p\Big]},
\end{aligned}
$$

where the first equality corresponds to the fact that $\mathbb{E}[aX] = a\mathbb{E}[X]$ for any random variable X and any nonrandom $a \in \mathbb{R}$, the second equality is explained by Tonelli's theorem, the second inequality is a consequence of Cauchy-Schwarz inequality, we used the fact that $\sup \mathbb{E}[.] \leq \mathbb{E}[\sup .]$ in the fourth inequality. Using the arguments from the proof of Lemma C4 along with Condition **[A]**, we can show that

$$\frac{1}{(nT)^{\frac{p}{2}-2}} \sup_{n \in \mathbb{N}, t \in [0,nT]} \sqrt{\mathbb{E}\Big[\sup_{\theta \in \Theta} \Big(1 + \frac{\overline{\lambda}_n^{(i)}(t,\theta)}{\nu_-}\Big)^{2p}\Big]} \sqrt{\mathbb{E}\Big[\sup_{\theta \in \Theta} \overline{\lambda}_n^{(i)}(t,\theta^*)^p\Big]} \to 0,$$

which implies that for $i = 1, \dots, d$ we have

$$\int_\Theta d\theta \mathbb{E}\Big[\Big|I_n^{(i)}(nT,\theta)\Big|^p\Big] \to 0.$$

With the same arguments, we can also show that for $i = 1, \dots, d$ we have

$$\int_\Theta d\theta \mathbb{E}\Big[\Big|\partial\theta I_n^{(i)}(nT,\theta)\Big|^p\Big] \to 0.$$

Thus, we can deduce by Equation (C18) for $i = 1, \dots, d$ that

$$\mathbb{E}\Big[\Big|\sup_{\theta \in \Theta} I_n^{(i)}(nT,\theta)\Big|^p\Big] \to 0. \tag{C19}$$

We can deduce that

$$\begin{aligned}
\mathbb{E}\Big[\Big|\sup_{\theta \in \Theta}\Big|\sum_{i=1}^d I_n^{(i)}(nT,\theta)\Big|^p\Big] &\leq C\mathbb{E}\Big[\sup_{\theta \in \Theta}\sum_{i=1}^d \Big|I_n^{(i)}(nT,\theta)\Big|^p\Big] \\
&\leq C\mathbb{E}\Big[\sum_{i=1}^d \sup_{\theta \in \Theta}\Big|I_n^{(i)}(nT,\theta)\Big|^p\Big] \\
&= C\sum_{i=1}^d \mathbb{E}\Big[\sup_{\theta \in \Theta}\Big|I_n^{(i)}(nT,\theta)\Big|^p\Big] \\
&= C\sum_{i=1}^d \mathbb{E}\Big[\Big|\sup_{\theta \in \Theta} I_n^{(i)}(nT,\theta)\Big|^p\Big] + o_{\mathbb{P}}(1) \\
&\to 0,
\end{aligned}$$

where the first inequality is a consequence of the fact that $|\sum_{i=1}^d a_i|^p \leq C\sum_{i=1}^d |a_i|^p$, the second inequality follows as $\sup \sum \leq \sum \sup$, the first equality corresponds to the fact that

$\mathbb{E}[aX] = a\mathbb{E}[X]$ for any random variable X and any nonrandom $a \in \mathbb{R}$, the second equality comes from Equation (C6) and the martingaleness of $\overline{M}_{t,n}$, and the convergence is due to Equation (C19). To prove that $\mid \sum_{i=1}^{d} II_n^{(i)}(nT, \theta) - \overline{\mathbb{Y}}(\theta) \mid \to^{\mathbb{P}} 0$, we can use Proposition C2 along with Condition **[A]**. □

We provide now the following lemma, which is the classical nondegeneracy condition on $\overline{\mathbb{Y}}$. This corresponds to Condition **[A](iv)** (p. 1807) in Clinet and Yoshida (2017). This extends Lemma A.7 (p. 1836) in Clinet and Yoshida (2017) which is restricted to the large-T asymptotics to the in-fill asymptotics.

Lemma C6. *We assume that Condition **[A] (iii) and (v)** hold. For any $\theta \in \Theta - \{\theta^*\}$, we have that $\overline{\mathbb{Y}}(\theta) \neq 0$.*

Proof. We assume that $\theta \in \Theta$ and that $\overline{\mathbb{Y}}(\theta) = 0$. In view of Equation (C17), we can deduce that

$$0 = \sum_{i=1}^{d} \int_E (\log(\frac{v}{u})u - (v-u))\Pi_{\theta^*}^{(i)}(du, dv, dw).$$

By Proposition C1 along with Condition **[A] (iii) and (v)**, this equation can be reexpressed as

$$0 = \sum_{i=1}^{d} \mathbb{E}\Big[(\log(\frac{\overline{\lambda}_{lim}^{(i)}(\theta)}{\overline{\lambda}_{lim}^{(i)}(\theta^*)})\overline{\lambda}_{lim}^{(i)}(\theta^*) - (\overline{\lambda}_{lim}^{(i)}(\theta) - \overline{\lambda}_{lim}^{(i)}(\theta^*)))\Big].$$

For any $i = 1, \ldots, d$ we also have that

$$0 \geq (\log(\frac{\overline{\lambda}_{lim}^{(i)}(\theta)}{\overline{\lambda}_{lim}^{(i)}(\theta^*)})\overline{\lambda}_{lim}^{(i)}(\theta^*) - (\overline{\lambda}_{lim}^{(i)}(\theta) - \overline{\lambda}_{lim}^{(i)}(\theta^*))).$$

This yields that for any $i = 1, \ldots, d$ a.s.

$$0 = (\log(\frac{\overline{\lambda}_{lim}^{(i)}(\theta)}{\overline{\lambda}_{lim}^{(i)}(\theta^*)})\overline{\lambda}_{lim}^{(i)}(\theta^*) - (\overline{\lambda}_{lim}^{(i)}(\theta) - \overline{\lambda}_{lim}^{(i)}(\theta^*))).$$

We can then deduce that for any $i = 1, \dots, d$ a.s.

$$\overline{\lambda}^{(i)}_{lim}(\theta^*) = \overline{\lambda}^{(i)}_{lim}(\theta).$$

By injectivity of the function $\theta \mapsto (\overline{\lambda}^{(1)}_{lim}(\theta), \dots, \overline{\lambda}^{(d)}_{lim}(\theta))$, we can infer that $\theta^* = \theta$. $\square$

We provide the proof of consistency in what follows. This extends Theorem 3.9 (p. 1807) in Clinet and Yoshida (2017) which is restricted to the large-T asymptotics to the in-fill asymptotics.

Proof of Equation (18) in Theorem 1. We have by Lemma C3 a.s. $\widehat{\overline{\theta}}_n = \widehat{\theta}_{T,n}$. Since the consistency is a convergence in probability, we can replace $\widehat{\theta}_{T,n}$ by $\widehat{\overline{\theta}}_n$ in the rest of this proof. In view of the expression $\overline{\mathbb{Y}}(\theta)$, we can see that $\overline{\mathbb{Y}}(\theta) \leq 0$ for any $\theta \in \Theta$ and $\overline{\mathbb{Y}}(\theta^*) = 0$. As an application of Lemma C6 along with Condition **[A] (iii) and (v)**, we can deduce that θ^* is a global maximum of $\overline{\mathbb{Y}}$. By Lemma C5 along with Condition **[A]**, the consistency is directly proven. $\square$

C.5 Proofs of the CLT

We start with the following lemma. This extends Lemma A.1 (p. 1824) in Clinet and Yoshida (2017) which is restricted to the large-T asymptotics to the in-fill asymptotics.

Lemma C7. *We assume that Condition **[A]** holds. For any $\theta \in \Theta$, we have that $\bar{l}_n(\theta)$ is a.s. finite and admits a derivative in θ that satisfies*

$$\partial_\theta \bar{l}_n(\theta) = \sum_{i=1}^{d} \int_0^{Tn} \frac{\partial_\theta \overline{\lambda}^{(i)}_n(t,\theta)}{\overline{\lambda}^{(i)}_n(t,\theta)} d\overline{N}^{(i)}_{t,n} - \sum_{i=1}^{d} \int_0^{Tn} \partial_\theta \overline{\lambda}^{(i)}_n(t,\theta) dt.$$

Moreover, we have that $\bar{l}_n(\theta)$ is twice differentiable and that its Hessian matrix satisfies

$$\begin{aligned}
\partial^2_\theta \bar{l}_n(\theta) \quad = \quad & \sum_{i=1}^{d} \int_0^{Tn} \partial_\theta \frac{\partial_\theta \overline{\lambda}^{(i)}_n(t,\theta)}{\overline{\lambda}^{(i)}_n(t,\theta)} d\overline{M}^{(i)}_{t,n} - \sum_{i=1}^{d} \int_0^{Tn} (\partial_\theta \overline{\lambda}^{(i)}_n)^{\otimes 2}(t,\theta)) \overline{\lambda}^{(i)}_n(t,\theta)^{-2} \overline{\lambda}^{(i)}_n(t,\theta^*) dt \\
& + \sum_{i=1}^{d} \int_0^{Tn} (\partial^2_\theta \overline{\lambda}^{(i)}_n)(t,\theta)) \overline{\lambda}^{(i)}_n(t,\theta)^{-1} (\overline{\lambda}^{(i)}_n(t,\theta) - \overline{\lambda}^{(i)}_n(t,\theta^*)) dt.
\end{aligned}$$

Proof. From Equation (C8), we define $\bar{l}_n(\theta) := \sum_{i=1}^d \bar{l}_n^{(i),I}(\theta) - \bar{l}_n^{(i),II}(\theta)$. First, we show that for any $\theta \in \Theta$ and any $i = 1, \ldots, d$ we have that $\bar{l}_n^{(i),I}(\theta) - \bar{l}_n^{(i),II}(\theta)$ is a.s. finite and admits a derivative in θ satisfying

$$\begin{aligned} \partial_\theta \int_0^{nT} \log(\overline{\lambda}_n^{(i)}(t,\theta)) d\overline{N}_{t,n}^{(i)} &= \int_0^{nT} \partial_\theta \log(\overline{\lambda}_n^{(i)}(t,\theta)) d\overline{N}_{t,n}^{(i)}, & \text{(C20)} \\ \partial_\theta \int_0^{nT} \overline{\lambda}_n^{(i)}(t,\theta) dt &= \int_0^{Tn} \partial_\theta \overline{\lambda}_n^{(i)}(t,\theta) dt. & \text{(C21)} \end{aligned}$$

By Equation (C1), $\bar{l}_n^{(i),I}(\theta)$ can be reexpressed as

$$\bar{l}_n^{(i),I}(\theta) = \sum_{k \in \mathbb{N}_* \text{ s.t. } \overline{T}_k^{(i)} < nT} \log(\overline{\lambda}_n^{(i)}(\overline{T}_k^{(i)}, \theta)). \tag{C22}$$

As $\bar{l}_n^{(i),I}(\theta)$ is a finite sum, and since $\overline{\lambda}_n^{(i)}(\overline{T}_k^{(i)}, \theta) > 0$ by Condition **[A] (i)-(ii)** , it is a.s. finite. In addition, since log and $\overline{\lambda}_n^{(i)}(\overline{T}_k^{(i)}, .)$ admit derivative in $\theta \in \Theta$ by Lemma C4 along with Condition **[A]**, then $\bar{l}_n^{(i),I}$ also admits a derivative by the chain rule. As the sum is finite and by linearity of the derivative operator, we deduce that

$$\partial_\theta \sum_{k \in \mathbb{N}_* \text{ s.t. } \overline{T}_k^{(i)} < nT} \log(\overline{\lambda}_n^{(i)}(\overline{T}_k^{(i)}, \theta)) = \sum_{k \in \mathbb{N}_* \text{ s.t. } \overline{T}_k^{(i)} < nT} \partial_\theta \log(\overline{\lambda}_n^{(i)}(\overline{T}_k^{(i)}, \theta)).$$

By Equation (C22), this equality can be reexpressed as

$$\partial_\theta \int_0^{nT} \log(\overline{\lambda}_n^{(i)}(t,\theta)) d\overline{N}_{t,n}^{(i)} = \int_0^{nT} \partial_\theta \log(\overline{\lambda}_n^{(i)}(t,\theta)) d\overline{N}_{t,n}^{(i)}.$$

The term $\bar{l}_n^{(i),II}(\theta)$ will be a.s. finite if we can show that its L^1-norm is finite. We have that its L^1-norm can be bounded as

$$\begin{aligned} \mathbb{E}\Big[\Big|\bar{l}_n^{(i),II}(\theta)\Big|\Big] &= \mathbb{E}\Big[\Big|\int_0^{nT} \log(\overline{\lambda}_n^{(i)}(t,\theta)) dt\Big|\Big] \\ &\leq \mathbb{E}\Big[\int_0^{nT} \Big|\log(\overline{\lambda}_n^{(i)}(t,\theta))\Big| dt\Big] \\ &\leq \mathbb{E}\Big[\int_0^{nT} |\overline{\lambda}_n^{(i)}(t,\theta)|^{-1} dt\Big] + \mathbb{E}\Big[\int_0^{nT} \Big|\overline{\lambda}_n^{(i)}(t,\theta)) - 1\Big| dt\Big] \\ &\leq \frac{nT}{\nu_-} + \mathbb{E}\Big[\int_0^{nT} \Big|\overline{\lambda}_n^{(i)}(t,\theta)) - 1\Big| dt\Big] \\ &\leq \frac{nT}{\nu_-} + C, \end{aligned}$$

where we use the definition of $\bar{l}_n^{(i),II}(\theta)$ in the equality, the triangular inequality in the first inequality, the fact that $|\log(z)| \leq z^{-1} + |z-1|$ for any $z \in \mathbb{R}_*^+$ together with the linearity of the expectation operator in the second inequality, the third inequality is explained by Condition **[A] (i)-(ii)** and the fourth inequality by Lemma C4 along with Condition **[A]**. We have thus shown that the L^1-norm of $\bar{l}_n^{(i),II}(\theta)$ is finite, so that $\bar{l}_n^{(i),II}(\theta)$ is a.s. finite. By extending the arguments, we can prove that

$$\begin{aligned} \int_0^{Tn} |\partial_\theta \overline{\lambda}_n^{(i)}(t,\theta)|dt &\leq \int_0^{Tn} |\sup_{\theta \in \Theta} \partial_\theta \overline{\lambda}_n^{(i)}(t,\theta)|dt \\ &\leq C. \end{aligned}$$

Now, an application of the dominated convergence theorem yields Equation (C21). We can prove the case $\partial_\theta^2 \bar{l}_n(\theta)$ with the same arguments.

□

We note that by Equation (C6), $\partial_\theta \bar{l}_n(\theta^*)$ can be reexpressed as

$$\partial_\theta \bar{l}_n(\theta^*) = \sum_{i=1}^{d} \int_0^{Tn} \frac{\partial_\theta \overline{\lambda}_n^{(i)}(t,\theta^*)}{\overline{\lambda}_n^{(i)}(t,\theta^*)} d\overline{M}_{t,n}^{(i)}. \tag{C23}$$

We provide the proof of the CLT in what follows. This extends Theorem 3.11 (p. 1809) in Clinet and Yoshida (2017) which is restricted to the large-T asymptotics to the in-fill asymptotics.

Proof of Equations (19), (20) and (21) in Theorem 1. First, we have a.s. $\widehat{\overline{\theta}}_n = \widehat{\theta}_{T,n}$ by Lemma C3 and thus we can replace $\widehat{\theta}_{T,n}$ by $\widehat{\overline{\theta}}_n$ in the rest of this proof since it will not affect convergence in distribution. We obtain by a Taylor expansion that

$$\partial_\theta \bar{l}_n(\widehat{\overline{\theta}}_n) = \partial_\theta \bar{l}_n(\theta^*) + \partial_\theta^2 \bar{l}_n(\zeta_n)(\widehat{\overline{\theta}}_n - \theta^*),$$

where ζ_n is between $\widehat{\overline{\theta}}_n$ and θ^*. Since $\widehat{\overline{\theta}}_n$ is defined as the maximizer of $\bar{l}_n(.)$, we deduce that $\partial_\theta \bar{l}_n(\widehat{\overline{\theta}}_n) = 0$. This yields that

$$0 = \partial_\theta \bar{l}_n(\theta^*) + \partial_\theta^2 \bar{l}_n(\zeta_n)(\widehat{\overline{\theta}}_n - \theta^*).$$

If we multiply by $\frac{-\Gamma^{-1}}{\sqrt{nT}}$, we obtain that

$$0 = \frac{-\Gamma^{-1}}{\sqrt{nT}} \partial_\theta \bar{l}_n(\theta^*) + \frac{-\Gamma^{-1}}{\sqrt{nT}} \partial_\theta^2 \bar{l}_n(\zeta_n)(\widehat{\overline{\theta}}_n - \theta^*).$$

This equation can be reexpressed as

$$0 = \frac{-\Gamma^{-1}}{\sqrt{nT}} \partial_\theta \bar{l}_n(\theta^*) + \frac{-\Gamma^{-1}}{nT} \partial_\theta^2 \bar{l}_n(\zeta_n) \sqrt{nT}(\widehat{\overline{\theta}}_n - \theta^*).$$

To prove Equation (19), it remains to show that

$$\begin{aligned} \frac{-\Gamma^{-1}}{\sqrt{nT}} \partial_\theta \bar{l}_n(\widehat{\overline{\theta}}_n) &\to^{\mathcal{D}} \quad \Gamma^{-1/2}\xi., & \text{(C24)} \\ \frac{-\Gamma^{-1}}{nT} \partial_\theta^2 \bar{l}_n(\zeta_n) &\to^{\mathbb{P}} \quad 1. & \text{(C25)} \end{aligned}$$

Then, Equation (19) easily follows using Slutsky's theorem. We prove now Equation (C24). By Equation (C23), we have that

$$\frac{-\Gamma^{-1}}{\sqrt{nT}} \partial_\theta \bar{l}_n(\widehat{\overline{\theta}}_n) = \frac{-\Gamma^{-1}}{\sqrt{nT}} \sum_{i=1}^{d} \int_0^{Tn} \frac{\partial_\theta \overline{\lambda}_n^{(i)}(t, \theta^*)}{\overline{\lambda}_n^{(i)}(t, \theta^*)} d\overline{M}_{t,n}^{(i)}.$$

For $u \in [0, 1]$, we define $S_{u,n}$ as

$$S_{u,n} = \frac{-\Gamma^{-1}}{\sqrt{nT}} \sum_{i=1}^{d} \int_0^{uTn} \frac{\partial_\theta \overline{\lambda}_n^{(i)}(t, \theta^*)}{\overline{\lambda}_n^{(i)}(t, \theta^*)} d\overline{M}_{t,n}^{(i)}. \qquad \text{(C26)}$$

We use Theorem $VIII.3.24$ in Jacod and Shiryaev (2013). We can calculate that

$$\begin{aligned} \langle S_n, S_n \rangle_u \quad &= \quad \frac{\Gamma^{-2}}{nT} \sum_{i=1}^{d} \int_0^{uTn} \frac{\partial_\theta \overline{\lambda}_n^{(i)}(t, \theta^*)^2}{\overline{\lambda}_n^{(i)}(t, \theta^*)} dt. \\ &\to^{\mathbb{P}} \quad u\Gamma^{-1}. \end{aligned}$$

We prove now that Lindeberg's condition is satisfied. For any $\epsilon > 0$ we have

$$\begin{aligned}
\mathbb{E}\Big[\sum_{s\le u}|\Delta S_{s,n}|^2\mathbf{1}_{|\Delta S_{s,n}|>\epsilon}\Big] &\le \mathbb{E}\Big[\frac{1}{a}\sum_{s\le u}|\Delta S_{s,n}|^3\Big] \\
&= \mathbb{E}\Big[\frac{1}{a}\sum_{s\le u}\Big|\Delta\frac{-\Gamma^{-1}}{\sqrt{nT}}\sum_{i=1}^{d}\int_0^{uTn}\frac{\partial_\theta\overline{\lambda}_n^{(i)}(t,\theta^*)}{\overline{\lambda}_n^{(i)}(t,\theta^*)}d\overline{M}_{t,n}^{(i)}\Big|^3\Big] \\
&= \mathbb{E}\Big[\frac{1}{a}\sum_{s\le u}\Big|\Delta\frac{-\Gamma^{-1}}{\sqrt{nT}}\sum_{i=1}^{d}\int_0^{uTn}\frac{\partial_\theta\overline{\lambda}_n^{(i)}(t,\theta^*)}{\overline{\lambda}_n^{(i)}(t,\theta^*)}d\overline{N}_{t,n}^{(i)}\Big|^3\Big] \\
&= \mathbb{E}\Big[\frac{1}{a}\sum_{i=1}^{d}\int_0^{uTn}\Big|\frac{-\Gamma^{-1}}{\sqrt{nT}}\frac{\partial_\theta\overline{\lambda}_n^{(i)}(t,\theta^*)}{\overline{\lambda}_n^{(i)}(t,\theta^*)}\Big|^3 d\overline{N}_{t,n}^{(i)}\Big] \\
&= \mathbb{E}\Big[\frac{1}{a}\sum_{i=1}^{d}\int_0^{uTn}\Big|\frac{-\Gamma^{-1}}{\sqrt{nT}}\frac{\partial_\theta\overline{\lambda}_n^{(i)}(t,\theta^*)}{\overline{\lambda}_n^{(i)}(t,\theta^*)}\Big|^3 \overline{\lambda}_n^{(i)}(t,\theta^*)dt\Big] \\
&\le \mathbb{E}\Big[\frac{1}{a}\Big|\frac{-\Gamma^{-1}}{\sqrt{nT}}\Big|^3\sum_{i=1}^{d}\int_0^{uTn}\Big|\frac{\partial_\theta\overline{\lambda}_n^{(i)}(t,\theta^*)}{\overline{\lambda}_n^{(i)}(t,\theta^*)}\Big|^3 \overline{\lambda}_n^{(i)}(t,\theta^*)dt\Big] \\
&\le \mathbb{E}\Big[\frac{1}{a}\Big|\frac{-\Gamma^{-1}}{\sqrt{nT}}\Big|^3\sum_{i=1}^{d}\int_0^{uTn}\sup_{\theta\in\Theta}\Big|\frac{\partial_\theta\overline{\lambda}_n^{(i)}(t,\theta^*)}{\overline{\lambda}_n^{(i)}(t,\theta^*)}\Big|^3 \overline{\lambda}_n^{(i)}(t,\theta^*)dt\Big],
\end{aligned}$$

where we used the fact that $\mathbf{1}_{|\Delta S_{s,n}|>\epsilon} \le \frac{1}{a}|\Delta S_{s,n}|$ in the first inequality, the first equality is due to Definition (C26), the second equality is explained by the fact that the drift part does not jump, the third and fourth equality are a consequence of the form of $d\overline{N}_{t,n}^{(i)}$. We can continue to bound the Linderberg's term by

$$\begin{aligned}
\mathbb{E}\Big[\sum_{s\le u}|\Delta S_{s,n}|^2\mathbf{1}_{|\Delta S_{s,n}|>\epsilon}\Big] &\le \mathbb{E}\Big[\frac{1}{a}\Big|\frac{-\Gamma^{-1}}{\sqrt{nT}}\Big|^3\sum_{i=1}^{d}\int_0^{uTn}\sup_{\theta\in\Theta}\Big|\frac{\partial_\theta\overline{\lambda}_n^{(i)}(t,\theta^*)}{\overline{\lambda}_n^{(i)}(t,\theta^*)}\Big|^3\overline{\lambda}_n^{(i)}(t,\theta^*)dt\Big] \\
&\le \mathbb{E}\Big[\frac{1}{a}\Big|\frac{-\Gamma^{-1}}{\sqrt{nT}}\Big|^3\sum_{i=1}^{d}\int_0^{uTn}\sup_{\theta\in\Theta}\frac{|\partial_\theta\overline{\lambda}_n^{(i)}(t,\theta^*)|^3}{\overline{\lambda}_n^{(i)}(t,\theta^*)^2}dt\Big] \\
&\le \mathbb{E}\Big[\frac{1}{a\nu_-^2}\Big|\frac{-\Gamma^{-1}}{\sqrt{nT}}\Big|^3\sum_{i=1}^{d}\int_0^{uTn}\sup_{\theta\in\Theta}|\partial_\theta\overline{\lambda}_n^{(i)}(t,\theta^*)|^3dt\Big] \\
&= \frac{1}{a\nu_-^2}\Big|\frac{-\Gamma^{-1}}{\sqrt{nT}}\Big|^3\sum_{i=1}^{d}\int_0^{uTn}\mathbb{E}\Big[\sup_{\theta\in\Theta}|\partial_\theta\overline{\lambda}_n^{(i)}(t,\theta^*)|^3\Big]dt \\
&\le \frac{CunT}{a\nu_-^2}\Big|\frac{-\Gamma^{-1}}{\sqrt{nT}}\Big|^3 \\
&\to 0,
\end{aligned}$$

where the third inequality is due to Condition **[A] (i)-(ii)**, and Lemma C4 along with Condition **[A]** is used for the fourth inequality. We have thus shown that Lindeberg's condition holds, so that Equation (C24) is satisfied. We prove now Equation (C25), i.e. that $\frac{-\Gamma^{-1}}{nT}\partial_\theta^2\overline{l}_n(\zeta_n)\to^{\mathbb{P}}1$. It is sufficient to prove that

$$|\Gamma+(nT)^{-1}\partial_\theta^2\overline{l}_n(\zeta_n)|\to^{\mathbb{P}}0.$$

If we define V_n as a shrinking ball centered on θ^* it is sufficient to show that

$$\sup_{\theta\in V_n}|\Gamma+(nT)^{-1}\partial_\theta^2\overline{l}_n(\zeta_n)|\to^{\mathbb{P}}0. \tag{C27}$$

We can reexpress Equation (C27) as the sum of a martingale term and a drift term. For the martingale term, we can notice that $\partial_\theta\frac{\partial_\theta\overline{\lambda}_n^{(i)}(t,\theta)}{\overline{\lambda}_n^{(i)}(t,\theta)}$ and $\partial_\theta^2\frac{\partial_\theta\overline{\lambda}_n^{(i)}(t,\theta)}{\overline{\lambda}_n^{(i)}(t,\theta)}$ are dominated by polynoms in $\partial_\theta^k\overline{\lambda}_n^{(i)}(t,\theta)$ and $\frac{1}{\overline{\lambda}_n^{(i)}(t,\theta)}$ for $k=0,1,2,3$. By an application of Sobolev's inequality, Lemma C4 along with Condition **[A]**, we obtain for p big enough that

$$\mathbb{E}\Big|\sup_{\theta\in\Theta}\frac{1}{nT}\int_0^{nT}\partial_\theta\frac{\partial_\theta\overline{\lambda}_n^{(i)}(t,\theta)}{\overline{\lambda}_n^{(i)}(t,\theta)}d\overline{M}_{t,n}^{(i)}\Big|^p=O((nT)^{-p/2}).$$

Given that we have L^p convergence implies convergence in probability, we can deduce that

$$\sum_{i=1}^{d} \sup_{\theta\in\Theta} \frac{1}{nT} \int_0^{nT} \partial_\theta \frac{\partial_\theta \overline{\lambda}_n^{(i)}(t,\theta)}{\overline{\lambda}_n^{(i)}(t,\theta)} d\overline{M}_{t,n}^{(i)} \to^{\mathbb{P}} 0.$$

We have that

$$\begin{aligned}
\mathbb{E} \sup_{\theta\in V_n} \Big| \frac{1}{nT} \int_0^{nT} \frac{\partial_\theta^2 \overline{\lambda}_n^{(i)}(t,\theta)}{\overline{\lambda}_n^{(i)}(t,\theta)} (\overline{\lambda}_n^{(i)}(t,\theta) \\
-\overline{\lambda}_n^{(i)}(t,\theta^*))dt\Big| \quad &\leq \quad \frac{1}{nT} \int_0^{nT} \mathbb{E} \sup_{\theta\in V_n} \Big| \frac{\partial_\theta^2 \overline{\lambda}_n^{(i)}(t,\theta)}{\overline{\lambda}_n^{(i)}(t,\theta)} (\overline{\lambda}_n^{(i)}(t,\theta) - \\
&\quad\quad \overline{\lambda}_n^{(i)}(t,\theta^*))\Big| dt \\
&\leq \quad \frac{1}{nT\nu_-} \int_0^{nT} \mathbb{E} \sup_{\theta\in V_n} \Big| \partial_\theta^2 \overline{\lambda}_n^{(i)}(t,\theta)(\overline{\lambda}_n^{(i)}(t,\theta) \\
&\quad\quad -\overline{\lambda}_n^{(i)}(t,\theta^*))\Big| dt \\
&\leq \quad \frac{1}{nT\nu_-} \int_0^{nT} \sqrt{\mathbb{E} \sup_{\theta\in V_n} \Big| \partial_\theta^2 \overline{\lambda}_n^{(i)}(t,\theta) \Big|^2} \\
&\quad\quad \times \sqrt{\mathbb{E} \sup_{\theta\in V_n} \Big| \overline{\lambda}_n^{(i)}(t,\theta) - \overline{\lambda}_n^{(i)}(t,\theta^*) \Big|^2} dt \\
&\leq \quad \frac{1}{nT\nu_-} \int_0^{nT} \sqrt{\sup_{n\in\mathbb{N}, t\in[0,nT]} \mathbb{E}\Big[\sup_{\theta\in\Theta} \Big| \partial_\theta^2 \overline{\lambda}_n^{(i)}(t,\theta) \Big|^2 \Big]} \\
&\quad\quad \times \sqrt{\mathbb{E} \sup_{\theta\in V_n} \Big| \overline{\lambda}_n^{(i)}(t,\theta) - \overline{\lambda}_n^{(i)}(t,\theta^*) \Big|^2} dt \\
&\leq \quad \frac{C}{nT\nu_-} \int_0^{nT} \\
&\quad\quad \sqrt{\mathbb{E} \sup_{\theta\in V_n} \Big| \overline{\lambda}_n^{(i)}(t,\theta) - \overline{\lambda}_n^{(i)}(t,\theta^*) \Big|^2} dt,
\end{aligned} \tag{C28}$$

where we use the triangular inequality and linearity of expectation operator in the first inequality, the second inequality is due to Condition **[A] (i)-(ii)**, the third inequality corresponds to Cauchy-Schwarz inequality, and the fifth inequality comes from Lemma C4 along with Condition **[A]**. Now, we obtain by a Taylor expansion that

$$\overline{\lambda}_n^{(i)}(t,\theta) - \overline{\lambda}_n^{(i)}(t,\theta^*) = \partial_\theta \overline{\lambda}_n^{(i)}(t,\widetilde{\theta})(\theta-\theta^*),$$

where $\widetilde{\theta}$ is between θ and θ^*. Applying square operator on both sides of the equation, we obtain that

$$\left|\overline{\lambda}_n^{(i)}(t,\theta) - \overline{\lambda}_n^{(i)}(t,\theta^*)\right|^2 = \left|\partial_\theta \overline{\lambda}_n^{(i)}(t,\widetilde{\theta})\right|^2 \left|\theta - \theta^*\right|^2.$$

We can easily deduce that

$$\sup_{\theta\in V_n}\left|\overline{\lambda}_n^{(i)}(t,\theta) - \overline{\lambda}_n^{(i)}(t,\theta^*)\right|^2 \leq \sup_{\theta\in\Theta}\left|\partial_\theta \overline{\lambda}_n^{(i)}(t,\theta)\right|^2 \left|\theta - \theta^*\right|^2. \tag{C29}$$

We have that

$$\begin{aligned}
\mathbb{E}\sup_{\theta\in V_n}\Bigg|\frac{1}{nT}\int_0^{nT}\frac{\partial_\theta^2\overline{\lambda}_n^{(i)}(t,\theta)}{\overline{\lambda}_n^{(i)}(t,\theta)}(\overline{\lambda}_n^{(i)}(t,\theta) & \\
-\overline{\lambda}_n^{(i)}(t,\theta^*))dt\Bigg| &\leq \frac{C}{nT\nu_-}\int_0^{nT}\sqrt{\mathbb{E}\sup_{\theta\in V_n}\left|\overline{\lambda}_n^{(i)}(t,\theta)-\overline{\lambda}_n^{(i)}(t,\theta^*)\right|^2}dt \\
&\leq \frac{C}{nT\nu_-}\int_0^{nT}\sqrt{\mathbb{E}\sup_{\theta\in\Theta}\left|\partial_\theta\overline{\lambda}_n^{(i)}(t,\theta)\right|^2\left|\theta-\theta^*\right|^2}dt \\
&\leq \frac{C}{\nu_-}\left|\theta-\theta^*\right|\sqrt{\sup_{n\in\mathbb{N},t\in[0,nT]}\mathbb{E}\sup_{\theta\in\Theta}\left|\partial_\theta\overline{\lambda}_n^{(i)}(t,\theta)\right|^2} \\
&\leq \frac{CK}{\nu_-}\left|\theta-\theta^*\right| \\
&\to 0.
\end{aligned}$$

where the first inequality comes from Equation (C28), the second inequality is due to Equation (C29), the fourth inequality is deduced by Lemma C4 along with Condition **[A]**, and the convergence is due to the fact that $\theta \in V_n$ with V_n shrinking to θ. Since L^1 convergence implies convergence in probability we obtain that

$$\sup_{\theta\in V_n}\left|\frac{1}{nT}\int_0^{nT}\frac{\partial_\theta^2\overline{\lambda}_n^{(i)}(t,\theta)}{\overline{\lambda}_n^{(i)}(t,\theta)}(\overline{\lambda}_n^{(i)}(t,\theta) - \overline{\lambda}_n^{(i)}(t,\theta^*))dt\right| \to^{\mathbb{P}} 0.$$

For the drift term, we define the process as

$$U_n^{(i)}(\theta) = \frac{1}{nT}\int_0^{nT}(\partial_\theta)^{\otimes 2}\overline{\lambda}_n^{(i)}(t,\theta)\overline{\lambda}_n^{(i)}(t,\theta)^{-2}\overline{\lambda}_n^{(i)}(t,\theta^*)dt.$$

When evaluated at θ^*, this process is equal to

$$U_n^{(i)}(\theta^*) = \frac{1}{nT}\int_0^{nT} (\partial_\theta)^{\otimes 2}\overline{\lambda}_n^{(i)}(t,\theta^*)\overline{\lambda}_n^{(i)}(t,\theta^*)^{-1}dt.$$

Using the arguments from the proof of the martingale case, we can show for any $i = 1,\dots,d$ that

$$|U_n^{(i)}(\theta^*) - U_n^{(i)}(\theta)| \to^{\mathbb{P}} 0.$$

Then the conclusion follows by writing Γ as the limit of $\sum_{i=1}^d U_n^{(i)}(\theta^*)$ and an application of Proposition C2 along with Condition **[A]**. Finally, the consistency of the asymptotic variance estimator, i.e. Equation (20), follows given its definition (17), the definition of the Fisher Hessian matrix in Equation (16), along with the consistency of $\widehat{\theta}_{T,n}$ (see Equation (18) in Theorem 1). The feasible CLT, i.e. Equation (21), is obtained via the standard CLT (see Equation (19)) together with Slutsky's theorem. □

C.6 Proofs of CLT for latency

We first give the proof of Proposition 2.

Proof of Proposition 2. By Equation (8) and Equation (7), we can reexpress $\widehat{L}_{T,n} - L$ as

$$\widehat{L}_{T,n} - L \quad = \quad F(\widehat{\theta}_{T,n}) - F(\theta^*_{ker}). \tag{C30}$$

By Condition **[B]**, we obtain for any $i = 1,\dots,d$ and any $j = 1,\dots,d$ by componentwise Taylor expansion that

$$\begin{aligned} F^{(i,j)}(\widehat{\theta}_{T,n}) - F^{(i,j)}(\theta^*_{ker}) &= dF^{(i,j)}(\theta^*_{ker})(\widehat{\theta}_{T,n} - \theta^*_{ker}) \\ &+(\widehat{\theta}_{T,n} - \theta^*_{ker})^T d^2F^{(i,j)}(\tilde{\theta})(\widehat{\theta}_{T,n} - \theta^*_{ker}), \end{aligned} \tag{C31}$$

where $d^2F^{(i,j)}(\theta)$ corresponds the $(m \times m)$-dimensional Hessian matrix of the (i,j)-index of F at point θ, and $\tilde{\theta}$ is between $\widehat{\theta}_{T,n}$ and θ^*_{ker}. To show the consistency, i.e. Equation (25), we can calculate for any $i = 1, \ldots, d$ and any $j = 1, \ldots, d$ that

$$\begin{aligned}
\widehat{L}_{T,n}^{(i,j)} - L^{(i,j)} &= F^{(i,j)}(\widehat{\theta}_{T,n}) - F^{(i,j)}(\theta^*_{ker}) \\
&= dF^{(i,j)}(\theta^*)(\widehat{\theta}_{T,n} - \theta^*_{ker}) + (\widehat{\theta}_{T,n} - \theta^*_{ker})^T d^2 F^{(i,j)}(\tilde{\theta})(\widehat{\theta}_{T,n} - \theta^*_{ker}) \\
&= O_{\mathbb{P}}(||\widehat{\theta}_{T,n} - \theta^*_{ker}||) + O_{\mathbb{P}}(||\widehat{\theta}_{T,n} - \theta^*_{ker}||^2) \\
&\to^{\mathbb{P}} 0,
\end{aligned}$$

where we use Equation (C30) in the first equality, the second equality is due to Equation (C31), the third equality is a consequence to the fact that $\overline{\Omega}$ is a compact set and F is twice continuously differentiable by Condition **[B]** so that $dF^{(i,j)}(\theta^*_{ker})$ and $d^2F^{(i,j)}(\tilde{\theta})$ are bounded, and the convergence is obtained via the consistency of $\widehat{\theta}_{T,n}$ (see Equation (18) in Theorem 1 along with Condition **[A]**). To prove the CLT, i.e. Equation (26), we can calculate

$$\begin{aligned}
\sqrt{nT}(\widehat{L}_{T,n}^{(i,j)} - L^{(i,j)})_{i=1,\ldots,d}^{j=1,\ldots,d} &= \sqrt{nT}(F^{(i,j)}(\widehat{\theta}_{T,n}) - F^{(i,j)}(\theta^*_{ker}))_{i=1,\ldots,d}^{j=1,\ldots,d} \\
&= \sqrt{nT}\big(dF^{(i,j)}(\theta^*_{ker})(\widehat{\theta}_{T,n}) - \theta^*_{ker})\big)_{i=1,\ldots,d}^{j=1,\ldots,d} \\
&\quad + \sqrt{nT}\big((\widehat{\theta}_{T,n} - \theta^*_{ker})^T d^2F^{(i,j)}(\tilde{\theta})(\widehat{\theta}_{T,n} - \theta^*_{ker})\big)_{i=1,\ldots,d}^{j=1,\ldots,d} \\
&= \sqrt{nT}\big(dF^{(i,j)}(\theta^*_{ker})(\widehat{\theta}_{T,n}) - \theta^*_{ker})\big)_{i=1,\ldots,d}^{j=1,\ldots,d} + \sqrt{nT}O_{\mathbb{P}}(||\widehat{\theta}_{T,n} - \theta^*_{ker}||^2) \\
&= \sqrt{nT}\big(dF^{(i,j)}(\theta^*_{ker})(\widehat{\theta}_{T,n}) - \theta^*_{ker})\big)_{i=1,\ldots,d}^{j=1,\ldots,d} + O_{\mathbb{P}}(||\widehat{\theta}_{T,n} - \theta^*_{ker}||) \\
&\to^{\mathcal{D}} (dF^{(i,j)}(\theta^*_{ker})\Gamma^{-1/2}\xi)_{i=1,\ldots,d}^{j=1,\ldots,d}
\end{aligned}$$

where we use Equation (C30) in the first equality, the second equality is a consequence to Equation (C31), the third equality is due to the fact that $\overline{\Omega}$ is a compact set and F is twice continuously differentiable by Condition **[B]** so that $d^2F^{(i,j)}(\tilde{\theta})$ is bounded, the fourth equality is a consequence to the CLT of $\widehat{\theta}_{T,n}$ (see Equation (19) in Theorem 1 along with Condition **[A]**), and the convergence is obtained via the consistency and CLT of $\widehat{\theta}_{T,n}$ (see Equation (18) and Equation (19) in Theorem 1 along with Condition **[A]**). We show now Equation (27), i.e.

we reexpress $\eta^{(i,j)}$ as

$$\begin{aligned}
\eta^{(i,j)} &= dF^{(i,j)}(\theta^*_{ker})\Gamma^{-1/2}\xi \\
&= dF^{(i,j)}(\theta^*_{ker})\Big(\sum_{q=1}^{m}\big(\Gamma^{-1/2}\big)^{(1,q)}\xi^{(q)},\ldots,\sum_{q=1}^{m}\big(\Gamma^{-1/2}\big)^{(m,q)}\xi^{(q)}\Big)^T \\
&= \sum_{r=1}^{m}\Big(dF^{(i,j,r)}(\theta^*_{ker})\sum_{q=1}^{m}\big(\Gamma^{-1/2}\big)^{(r,q)}\xi^{(q)}\Big) \\
&= \sum_{q=1}^{m}\Big(\sum_{r=1}^{m}dF^{(i,j,r)}(\theta^*_{ker})\big(\Gamma^{-1/2}\big)^{(r,q)}\Big)\xi^{(q)},
\end{aligned}$$

where the first equality is due to Equation (22), the second and third equalities are matrix calculation, the fourth equality is from algebraic manipulations by inverting sums. Equation (28) can be deduced directly by using the fact that ξ follows a m-dimensional standard normal vector. The consistency of the covariance estimator, i.e. Equation (29), is due to the consistency of $\widehat{\theta}_{T,n}$ (see Equation (18) in Theorem 1 along with Condition **[A]**), the consistency of $\widehat{\Gamma}_T$ (see Equation (20) in Theorem 1), along with Condition **[B]**. The normalized feasible CLT, i.e. Equation (30) and Equation (31), is deduced via the standard CLT (see Equation (26)), the consistency of the covariance estimator (see Equation (29)), together with Slutsky's theorem. Finally, we can show Equation (32) since

$$\begin{aligned}
\widehat{\text{Cor}\left[\widetilde{\xi}^{(i,j)},\widetilde{\xi}^{(k,l)}\right]} &= \frac{\widehat{\text{Cov}[\eta^{(i,j)},\eta^{(k,l)}]}}{\sqrt{\widehat{\text{Var}\left[\eta^{(i,j)}\right]}\widehat{\text{Var}\left[\eta^{(k,l)}\right]}}} \\
&= \frac{\sum_{q=1}^{m}\Big(\sum_{r=1}^{m}dF^{(i,j,r)}(\widehat{\theta}_T)\big(\widehat{\Gamma}_T^{-1/2}\big)^{(r,q)}\Big)\Big(\sum_{r=1}^{m}dF^{(k,l,r)}(\widehat{\theta}_T)\big(\widehat{\Gamma}_T^{-1/2}\big)^{(r,q)}\Big)}{\sqrt{\widehat{\text{Var}\left[\eta^{(i,j)}\right]}\widehat{\text{Var}\left[\eta^{(k,l)}\right]}}} \\
&\to \frac{\sum_{q=1}^{m}\Big(\sum_{r=1}^{m}dF^{(i,j,r)}(\theta^*)\big(\Gamma^{-1/2}\big)^{(r,q)}\Big)\Big(\sum_{r=1}^{m}dF^{(k,l,r)}(\theta^*)\big(\Gamma^{-1/2}\big)^{(r,q)}\Big)}{\sqrt{\text{Var}\left[\eta^{(i,j)}\right]\text{Var}\left[\eta^{(k,l)}\right]}} \\
&= \frac{\text{Cov}[\eta^{(i,j)},\eta^{(k,l)}]}{\sqrt{\text{Var}\left[\eta^{(i,j)}\right]\text{Var}\left[\eta^{(k,l)}\right]}} \\
&= \text{Cor}\left[\widetilde{\xi}^{(i,j)},\widetilde{\xi}^{(k,l)}\right],
\end{aligned}$$

where the first equality is due to Equation (24), the second equality comes from Equation (23), the convergence is due to the consistency of $\widehat{\theta}_{T,n}$ (see Equation (18) in Theorem 1 along with Condition **[A]**) together with the consistency of $\widehat{\Gamma}_T$ (see Equation (20) in Theorem 1 along with Condition **[A]**) in the numerator and to Equation (29) in the denominator, the third equality corresponds to Equation (28), and the fourth equality is obtained by Equation (31). □

We provide now the proof of Corollary 3.

Proof of Corollary 3. We obtain by the CLT of the latency estimator (see Equation (26) in Proposition 2 along with Condition **[A]** and Condition **[B]**) the asymptotic matrix $(\eta^{(i,j)})_{i=1,\dots,d}^{j=1,\dots,d}$, which can be reexpressed as in Equation (27). We can then deduce Equation (35) by Condition **[C]**. The consistency of the asymptotic covariance matrix inverse, i.e. Equation (36), follows directly from the consistency of the covariance estimator in Equation (29). The feasible normalized CLT, i.e. Equation (37), is deduced via the standard CLT (see Equation (35)), the consistency of the asymptotic covariance matrix inverse (see Equation (36)) together with Slutsky's theorem. □

C.7 Proofs of the tests related to latency

We finally give the proofs of the corollaries related to latency tests. We start with the proof of Corollary 4.

Proof of Corollary 4. The size of the first Wald test statistic $W(\widetilde{L})$, i.e. Equation (38), can be shown converging in distribution to a chi-squared distribution with one degree of freedom using its definition (see Equation (9)), the CLT of the latency matrix estimator and the consistency of the covariance estimator (see Equations (26) and (29) in Proposition 2 along with Condition **[A]** and Condition **[B]**), and the form of the chi-squared distribution with one degree of freedom. The power of the first Wald test statistic $W(\widetilde{L})$, i.e. Equation (39), goes to 1 as an application of the CLT of the latency matrix estimator and the consistency of the covariance estimator (see

Equations (26) and (29) in Proposition 2 along with Condition **[A]** and Condition **[B]**) along with its definition (see Equation (9)). $\square$

We provide the proof of Corollary 5 in what follows.

Proof of Corollary 5. Under the null hypothesis $H_0' : \{L^{(i,j)} = L^{(k,l)}\}$, we can calculate that

$$\begin{aligned}
\sqrt{nT}(\widehat{L}_T^{(i,j)} - \widehat{L}_T^{(k,l)}) &= \sqrt{nT}(\widehat{L}_T^{(i,j)} - L^{(i,j)}) + \sqrt{nT}(L^{(k,l)} - \widehat{L}_T^{(k,l)}) + \sqrt{nT}(L^{(i,j)} - L^{(k,l)}) \\
&= \sqrt{nT}(\widehat{L}_T^{(i,j)} - L^{(i,j)}) + \sqrt{nT}(L^{(k,l)} - \widehat{L}_T^{(k,l)}) \\
&\to^{\mathcal{D}} \eta^{(i,j)} - \eta^{(k,l)},
\end{aligned}$$

where the first equality corresponds to algebraic manipulation, the second equality is due to the fact that under H_0' we have that $L^{(i,j)} = L^{(k,l)}$, and the convergence comes from the CLT of the latency matrix estimator (see Equation (26) in Proposition 2 along with Condition **[A]** and Condition **[B]**). If we write

$$\widetilde{\eta} = \eta^{(i,j)} - \eta^{(k,l)}, \tag{C32}$$

we know that $\widetilde{\eta}$ is normally distributed since we assume that $(\eta^{(i,j)}, \eta^{(k,l)})$ is a two-dimensional random vector. We have that the mean of $\widetilde{\eta}$ is null by its definition in Equation (22). It remains to calculate its variance. We obtain that

$$\begin{aligned}
\text{Var}\left[\widetilde{\eta}\right] &= \text{Var}\left[\eta^{(i,j)} - \eta^{(k,l)}\right] \\
&= \text{Var}\left[\eta^{(i,j)}\right] + \text{Var}\left[\eta^{(k,l)}\right] - 2\,\text{Cov}\left[\eta^{(i,j)}, \eta^{(k,l)}\right]
\end{aligned}$$

where the first equality comes from Equation (C32), and the second equality corresponds to a well-known variance-covariance property. By the consistency of the covariance estimator (see Equation (29) in Proposition 2 along with Condition **[A]** and Condition **[B]**), $\text{Var}\left[\widetilde{\eta}\right]$ can be consistently estimated as

$$\widehat{\text{Var}\left[\widetilde{\eta}\right]} = \widehat{\text{Var}\left[\eta^{(i,j)}\right]} + \widehat{\text{Var}\left[\eta^{(k,l)}\right]} - 2\widehat{\text{Cov}\left[\eta^{(i,j)}, \eta^{(k,l)}\right]}.$$

As a consequence, we obtain that W' converges in distribution to a chi-square with one degree of freedom. Under the alternative hypothesis $H_1' : \{L^{(i,j)} \neq L^{(k,l)}\}$, we can calculate that

$$\begin{aligned} \sqrt{nT}(\widehat{L}_T^{(i,j)} - \widehat{L}_T^{(k,l)}) &= \sqrt{nT}(\widehat{L}_T^{(i,j)} - L^{(i,j)}) + \sqrt{nT}(L^{(k,l)} - \widehat{L}_T^{(k,l)}) + \sqrt{nT}(L^{(i,j)} - L^{(k,l)}) \\ &= O_{\mathbb{P}}(1) + \sqrt{nT}(L^{(i,j)} - L^{(k,l)}), \end{aligned}$$

where the first equality corresponds to algebraic manipulation, and the second equality comes from the CLT of the latency matrix estimator (see Equation (26) in Proposition 2 along with Condition **[A]** and Condition **[B]**). Finally, we have by the consistency of the covariance estimator (see Equation (29) in Proposition 2 along with Condition **[A]** and Condition **[B]**) and by the fact that Ω is bounded that $\text{Var}\left[\widetilde{\eta}\right]$ is uniformly bounded. Thus, we can deduce that the second Wald statistic W' diverges under H_1'. □

Finally, we give the proof of Corollary 6.

Proof of Corollary 6. The size of the third Wald test statistic $\overline{W}(r)$, i.e. Equation (42), can be shown converging in distribution to a chi-squared distribution with q degrees of freedom using its definition (see Equation (11)), the CLT of the latency vector estimator and the consistency of the asymptotic covariance matrix inverse (see Equations (35) and (36) in Corollary 3 along with Condition **[A]**, Condition **[B]** and Condition **[C]**), and the form of the chi-squared distribution with q degrees of freedom. The power of the third Wald test statistic $\overline{W}(r)$, i.e. Equation (43), goes to 1 as an application of the CLT of the latency vector estimator and the consistency of the asymptotic covariance matrix inverse (see Equations (35) and (36) in Corollary 3 along with Condition **[A]**, Condition **[B]** and Condition **[C]**) along with its definition (see Equation (11)).

□

D Additional empirical results

The parameter estimates $\widehat{\alpha_T^{(i,j)}}$ are presented in Figure D1. Recall that α captures a size of jump associated with trading or quoting intensity. The size of jump for self-exciting effects in plots (a) and (c) is larger in periods of intensive trading - March 2020 when traders utilized the market shift due to the start of COVID-19 pandemic. Interestingly the cross-exciting effects from quotes to trades in the US, plots (f), and in Canada, plot (h), dropped during this period showing a change towards a trade driven market.

The parameter estimates $\widehat{\beta^{(i,j)}}$ are presented in Figure D2. At the beginning of COVID-19 pandemic in March 2020, decay captured by β changed from 1 to almost 2 reflecting acceleration in trading reactions for self-excitation parts in the US and Canada (plots (a) and (c)). Cross-exciting parameters $\widehat{\beta_T^{(4,3)}}$ and $\widehat{\beta_T^{(2,1)}}$ for quotes due to trades are stable during this period.

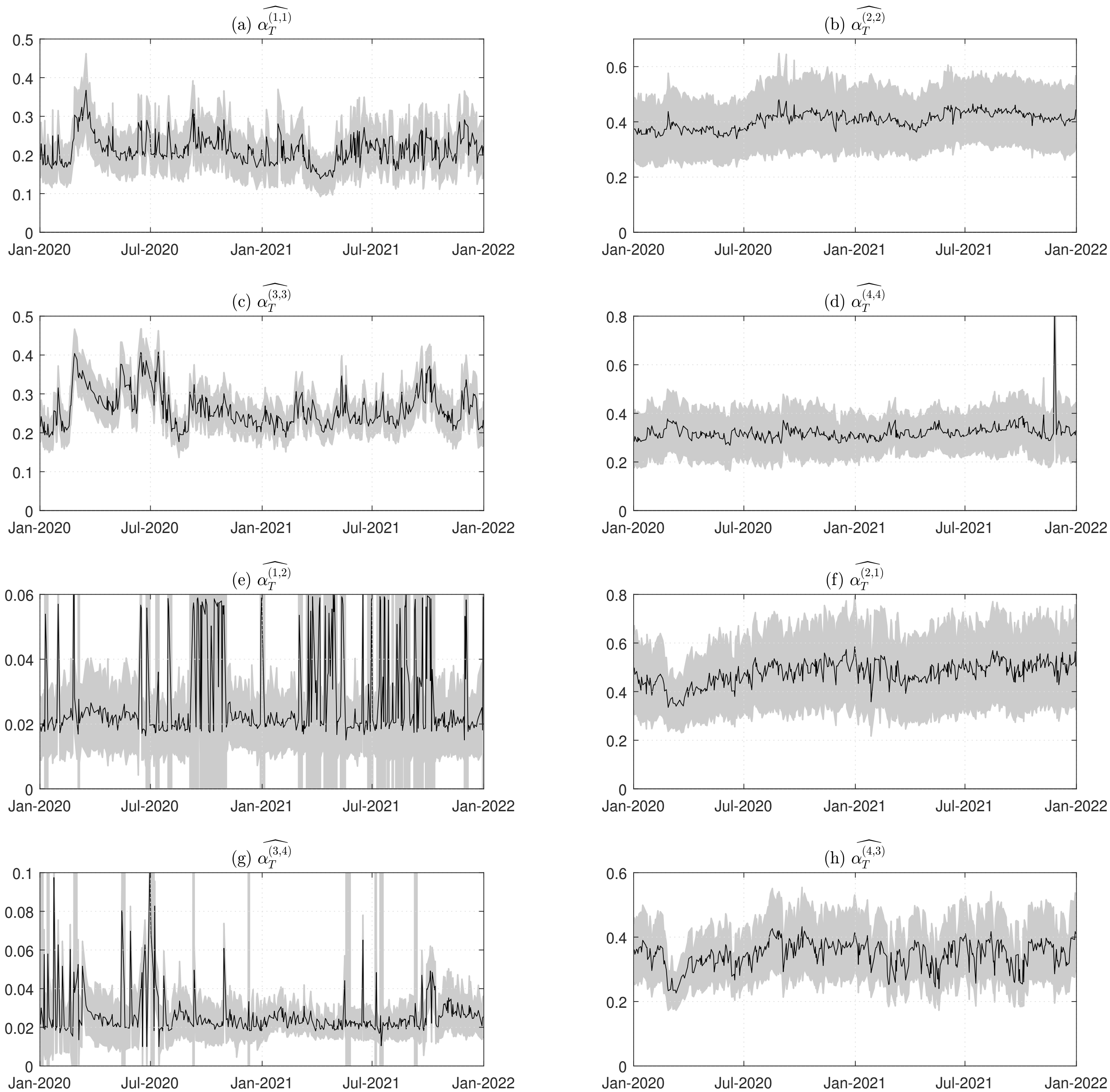

Figure D1: Parameter estimates $\widehat{\alpha_T^{(i,j)}}$ for each day for events in the NYSE and TSX. 90% confidence intervals are reported.

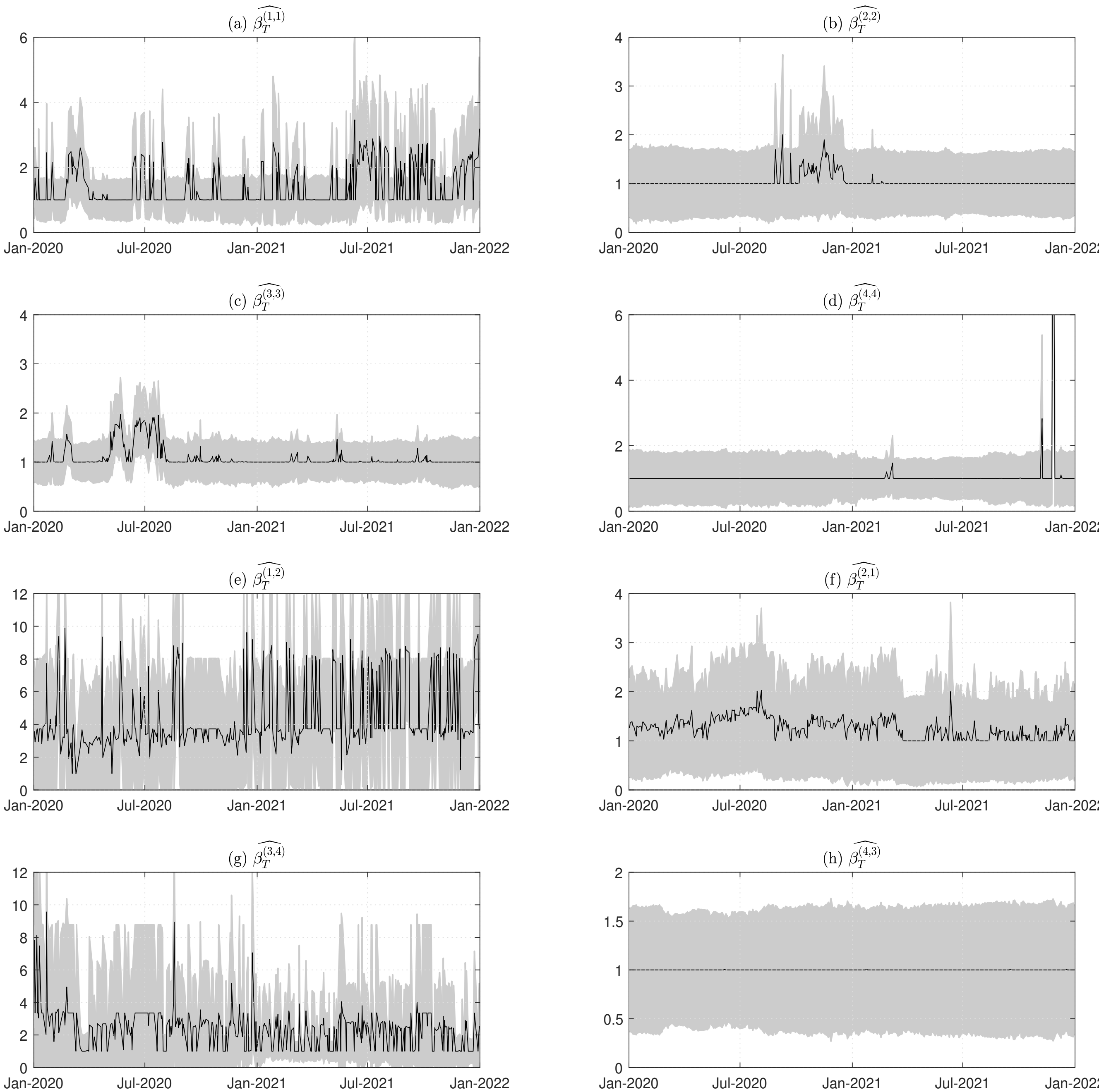

Figure D2: Parameter estimates $\widehat{\beta_T^{(i,j)}}$ for each day for events in the NYSE and TSX. 90% confidence intervals are reported.