Estimation of branching ratio for Hawkes processes with Itô semimartingale baseline

Yoann Potiron* Olivier Scaillet† Vladimir Volkov‡ and Seunghyeon Yu§

Abstract

We study Hawkes self-exciting processes where the baseline is driven by an Itô semimartingale. We consider estimation of the branching ratio, i.e., the integral of the kernel. The estimation procedure is based on empirical average and variance of local point process estimates. We characterize feasible statistics induced by central limit theory for the estimation procedure. We develop a test for any branching ratio value. We also propose a test for a branching ratio value that does not depend on the number of observations against the branching ratio is dependent and tends to unity as the number of observations increases. The results are obtained with in-fill asymptotics. Simulation studies corroborate asymptotic theory and show that we improve branching ratio estimation for this more realistic baseline. An empirical application on high-frequency data of the E-mini S&P500 future contracts shows that the branching ratio is around 0.7 and 0.8, while alternative methods are positively biased. We interpret the branching ratio as a measure of resiliency of the limit order book.

Keywords: branching ratio; kernel integral; self-exciting Hawkes processes; Itô semimartingale baseline; in-fill asymptotics; high-frequency data; resiliency

*Faculty of Business and Commerce, Keio University. 2-15-45 Mita, Minato-ku, Tokyo, 108-8345, Japan. E-mail: potiron@fbc.keio.ac.jp website: http://www.fbc.keio.ac.jp/~potiron

†(CONTACT) Swiss Finance Institute, University of Geneva. E-mail: olivier.scaillet@unige.ch

‡School of Business and Economics, University of Tasmania. E-mail: vladimir.volkov@utas.edu.au

§Kellogg School of Management, Northwestern University. E-mail: seunghyeon.yu@kellogg.northwestern.edu

1 Introduction

Point processes are widely used in statistics to characterize event times. The main stylized fact in this strand of literature, the presence of event clustering in time, motivates the so-called Hawkes self-exciting processes (see [Hawkes, 1971a] and [Hawkes, 1971b]). If we define N_t as the aggregated number of events up to time t and λ_t its corresponding intensity, a standard definition of a Hawkes self-exciting process is given by

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

Here, $\mu > 0$ is the Poisson baseline and ϕ is the non-negative exciting kernel, i.e., $\phi(t) \geq 0$ for any $t \in \mathbb{R}^+$. The particular case $\phi = 0$ corresponds to a classical Poisson process, thus we can view Hawkes processes as a natural extension of Poisson processes.

An early application of Hawkes processes evolves in seismology (see [Rubin, 1972], [Vere-Jones, 1978], [Ozaki, 1979], [Vere-Jones and Ozaki, 1982], [Ogata, 1978] and [Ogata, 1988]). [Ikefuji et al., 2022] analyze the impact of earthquake risk based on marked Hawkes processes. There are also some applications in financial econometrics (see [Yu, 2004], [Bowsher, 2007], [Embrechts et al., 2011], but also [Aït-Sahalia et al., 2014], and [Corradi et al., 2020]), finance (see [Large, 2007], [Aït-Sahalia et al., 2015] and [Fulop et al., 2015]) and in quantitative finance (see [Chavez-Demoulin et al., 2005], but also the papers [Bacry et al., 2013], [Jaisson and Rosenbaum, 2015] and [Morariu-Patrichi and Pakkanen, 2022]). See also [Liniger, 2009] and [Hawkes, 2018] with the references therein. [Cavaliere et al., 2023] develop a bootstrap approach, while [Christensen and Kolokolov, 2024] propose an unbounded intensity model for point processes. [Potiron and Volkov, 2025] consider estimation of latency.

This paper concerns the estimation of the branching ratio, i.e., the integral of the kernel function $\mathrm{BR} = \int_0^\infty \phi(t)\, dt$. To ensure stability of the Hawkes processes, the branching ratio has to be strictly smaller than unity. The interpretation of this branching ratio is closely related to the Poisson cluster representation for Hawkes processes (see [Hawkes and Oakes, 1974]). More specifically, Hawkes processes can be expressed as a population dynamic where each new individual duration time follows a

Poisson distribution with parameter μ. Then, each new individual gives birth to children according to a non-homogeneous Poisson process with kernel $\phi(t)$, and so on. In terms of population interpretation, this branching ratio is the average number of children of an individual.

The main application of the branching ratio lies in finance. From a financial perspective, the branching ratio can be interpreted as the average proportion of endogenous events. It is used as an empirical measure of the degree of endogeneity in the market. During a crisis, we then expect the branching ratio to be closer to unity. For example, [Filimonov and Sornette, 2012] consider the prediction of flash crashes based on the branching ratio. [Hardiman et al., 2013] report branching ratio values above 0.9. In [Filimonov and Sornette, 2012], branching ratio values between 0.7 and 0.8 are obtained. Since it could lead to different financial interpretations, a debate on the validity of these results is currently ongoing between people leaning towards values close to unity, or on the contrary way below. See also [Achab et al., 2018b] for the multidimensional case. There are also some applications in seismology (see [Bacry and Muzy, 2016]).

Empirical evidence suggests that the baseline is time-dependent, random and with possible jumps (see [Chen and Hall, 2013], [Clinet and Potiron, 2018] and [Rambaldi et al., 2015]). These features affect the estimation of the branching ratio. The empirical results very close to unity obtained in [Hardiman et al., 2013] are criticized in [Filimonov and Sornette, 2015] (Section 4.3–4.4); see also [Luo et al., 2024]. The authors point out that an estimation method only designed for constant baseline induces a bias when the baseline is time-dependent, random and with jumps. The bias is also visible in our numerical studies (see Figure 3).

In this paper, we consider Hawkes processes with a baseline driven by an Itô semimartingale with possible jumps, namely

$$\mu_t = \mu_0 + \int_0^t b_s \, ds + \int_0^t \sigma_s \, dW_s + (\delta \mathbf{1}_{\{|\delta| \leq 1\}}) \star (\underline{\mu} - \underline{\nu})_t + (\delta \mathbf{1}_{\{|\delta| > 1\}}) \star \underline{\mu}_t. \tag{1}$$

The Itô semimartingale baseline suits the three aforementioned empirical facts for the baseline intensity: time-dependence, randomness, and jumps. This framework is introduced in [Potiron et al., 2025]. It differs from the available literature on nonstationary Hawkes processes, that do not accommodate for the Itô semimartingale baseline. For example, [Chen and Hall, 2013], [Omi et al., 2017], [Kwan et al., 2023] and [Kwan, 2023] allow for a time-dependent baseline, with time-invariant kernel parameters. [Clinet and Potiron, 2018] and [Erdemlioglu et al., 2025] consider random time-dependent baseline and random time-dependent kernel parameters, where the parameters are smoother than the Itô semimartingale. [Roueff et al., 2016], [Roueff and Von Sachs, 2019] and [Mammen and Müller, 2023] propose nonparametric estimation. Spectral parametric estimation for misobserved Hawkes processes with a setting also covering a time-dependent baseline is given in [Cheysson and Lang, 2022].

Our inferential theory builds on in-fill asymptotics, i.e., when T is fixed and the number of observations on $[0, T]$ increases as $n \to \infty$. These asymptotics are popular with financial applications based on high-frequency data (see [Aït-Sahalia and Jacod, 2014]). The main reason why we use these asymptotics is that we observe in our empirical application a time-dependent branching ratio between different days. There already exists work to accommodate for in-fill asymptotics with Hawkes processes. In-fill asymptotic results from [Chen and Hall, 2013] are based on random observation times of order n. A single boosting of the baseline, i.e., $\lambda_t = \alpha\nu_t^* + \int_0^t h(t-s, \theta^*)dN_s$, is considered, where $\alpha \to \infty$ is a scaling sequence. [Clinet and Potiron, 2018] introduce a joint boosting of the baseline and the kernel, i.e., $\lambda(t) = n\nu_t^* + \int_0^t na_s^* \exp(-nb_s^*(t-s))dN_s$. [Kwan et al., 2023] revisit [Chen and Hall, 2013] with the same in-fill asymptotics as in [Clinet and Potiron, 2018], i.e. $\lambda_t = n\nu_t^* + \int_0^t na^* \exp(-nb^*(t-s))\, dN_s$. [Kwan, 2023], [Potiron and Volkov, 2025], [Potiron et al., 2025] and [Erdemlioglu et al., 2025] also use these in-fill asymptotics. There are compatible with the in-fill asymptotics of [Christensen and Kolokolov, 2024].

Most methods for estimation of the branching ratio are based on the estimation of the kernel. Numerical approximation of the kernel integral limit their use in practice. In the parametric case, [Ogata, 1978]) propose a method based on maximum likelihood estimation (MLE). See also [Clinet and Yoshida, 2017], [Potiron and Volkov, 2025] and [Potiron, 2025]. Based on [Veen and Schoenberg, 2008], [Marsan and Lengline, 2008] develop an expectation maximization (EM) estimator. [Lewis and Mohler, 2011] consider maximimum penalized likelihood estimation. [Omi et al., 2017] develop EM with time-dependent parametric baseline. In the nonparametric case, [Bacry and Muzy, 2016] and [Clements et al., 2023] consider procedures based on moments and cumulants.

In contrast, [Hardiman and Bouchaud, 2014] introduce a method that does not require estimation of the kernel. It is based on the ratio of mean and variance of the point process. Its main practical advantage is that it is very fast as there is no optimization procedure. However, their method is not robust to the time-dependent baseline. In this paper, we propose an extension of their estimation method. More specifically, we replace their estimator of mean and variance on $[0, T]$ by empirical average and variance of local point process estimates. See also [Achab et al., 2018a] and [Achab et al., 2018b], who consider a multidimensional extension.

In addition, we develop a test for a branching ratio value. We consider a Wald test, which is based on the estimation of the branching ratio. It extends the test for the absence of a Hawkes component in [Potiron et al., 2025]. Moreover, we propose a Wald test for a branching ratio value that does not depend on n against the branching ratio depends on n and tends to unity as $n \to \infty$. More specifically, we consider an alternative where the branching ratio depends on n of the form $BR = \widetilde{b}_n$, where $\widetilde{b}_n < 1$ and $\widetilde{b}_n \to 1$. [Jaisson and Rosenbaum, 2015] introduce such processes and they call them nearly unstable Hawkes processes. It is due to the fact that the Hawkes processes are unstable when the branching ratio is equal to unity (see [Brémaud and Massoulié, 2001]). As far as the authors know, these two tests are novel to the literature.

Our main result (see Theorem 1) is a feasible CLT for estimation of the branching ratio. It extends Section III in [Hardiman and Bouchaud, 2014], who show the consistency of the estimation procedure when the baseline is constant. It extends Theorem 4.1 in [Potiron et al., 2025]. We also give the limit theory of the Wald test for a branching ratio value (see Corollary 1). It extends Corollary 5.4 in [Potiron et al., 2025]. Moreover, we give the limit theory of the Wald test for a branching ratio value that does not depend on n (see Theorem 2). Under the alternative hypothesis, this extends Theorem 2.2 in [Jaisson and Rosenbaum, 2015]. The main novelty in the proofs is to divide the intensity by $1 - \widetilde{b}_n$ to deal with a time-dependent baseline and the branching ratio which both explode as the number of observations n increases.

Simulation studies corroborate the asymptotic theory. An empirical application on high-frequency data of the E-mini S&P500 future contracts shows that the branching ratio is around 0.7 and 0.8, and that the tests are ??. BLABLABLA.

The remainder of this paper is organized as follows. We provide the setting in Section 2, and we introduce the estimation and testing strategy in Section 3. We give the theoretical results in Section 4. In Section 5, we carry out numerical studies, which corroborate the asymptotic theory. In Section 6, an empirical application on high-frequency data of the E-mini S&P500 future contracts is presented. Finally, we provide concluding remarks in Section 7. All the proofs are gathered in the supplementary materials.

2 Setting

In this section, we introduce Hawkes self-exciting processes with a baseline driven by an Itô semi-martingale with possible jumps when the horizon T is finite. We also introduce the definition of the branching ratio.

For any space S such that $0 \in S$, we define the space without zero as S_*. For any space S, we denote by $\mathcal{B}(S)$ the Borel σ-algebra on the space S. In what follows, we introduce the point process N_t, which counts the number of events between 0 and t. We define N_t as a simple point process on the space of positive real numbers $\mathbb{R}^+$, i.e., a family $\{N(C)\}_{C \in \mathcal{B}(\mathbb{R}^+)}$ of random variables with values in the space of natural integers $\mathbb{N}$. Moreover, $N(C) = \sum_{k \in \mathbb{N}} \mathbf{1}_C(T_k)$ and $\{T_k\}_{k \in \mathbb{N}}$ is a sequence of event times, which are $\mathbb{R}^+$-valued and random. We assume that the first time is equal to 0 and the following times are increasing a.s., i.e., $\mathbb{P}\big(T_0 = 0 \text{ and } T_k < T_{k+1} \text{ for } k \in \mathbb{N}_*\big) = 1$. Let $\mathcal{B} = (\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \in \mathbb{R}^+}, \mathbb{P})$ be a filtered probability space which satisfies the usual conditions. For any $t \in \mathbb{R}$, we denote the filtration generated by some stochastic process X as $\mathcal{F}_t^X = \sigma\{X_s : 0 \leq s \leq t\}$. We assume that, for any $t \in \mathbb{R}^+$, the filtration generated by the point process N_t is included in the main filtration, i.e., $\mathcal{F}_t^N \subset \mathcal{F}_t$. Any nonnegative $\mathcal{F}_t$-progressively measurable process $\{\lambda_t\}_{t \in \mathbb{R}^+}$, which satisfies

$\mathbb{E}[N((a,b]) \mid \mathcal{F}_a] = \mathbb{E}\big[\int_a^b \lambda_s ds \,\big|\, \mathcal{F}_a\big]$ a.s. for all intervals $(a,b]$, is called an $\mathcal{F}_t$-intensity of N_t. Intuitively, the intensity corresponds to the expected number of events given the past information, i.e., $\lambda_t = \lim_{u\to 0} \mathbb{E}\Big[\frac{N_{t+u}-N_t}{u}\Big|\mathcal{F}_t\Big]$ a.s.. For background on point processes, the reader can consult [Jacod, 1975], [Jacod and Shiryaev, 2003], [Daley and Vere-Jones, 2003], and [Daley and Vere-Jones, 2008].

The present work is concerned with simple point processes N_t admitting an $\mathcal{F}_t$-intensity of the form

$$\lambda_t = \mu_t + \int_0^t \phi(t-s)\, dN_s. \tag{2}$$

Here, we have that μ_t is the $\widetilde{\mathcal{F}}_t$-Itô semimartingale baseline process with $\widetilde{\mathcal{F}}_t \subset \mathcal{F}_t$ and ϕ is the nonnegative exciting kernel, i.e., $\phi(t) \geq 0$ for any $t \in \mathbb{R}^+$. Since μ_t follows an $\widetilde{\mathcal{F}}_t$-Itô semimartingale, then we can construct a filtered extension $\overline{\mathcal{B}} = (\overline{\Omega}, \overline{\mathcal{F}}, \{\overline{\mathcal{F}}_t\}_{t\in[0,T]}, \overline{\mathbb{P}})$ on which are defined a standard Brownian motion W and a Poisson random measure $\underline{\mu}$ on $\mathbb{R}_+ \times E$, which is compensated by $\underline{\nu}(dt,dx) = dt \otimes F_t(dx)$. Here, we assume that E is an auxiliary Polish space and that F_t is σ-finite, infinite, and optional measure, having no atom. Then, the baseline μ_t has the Grigelionis representation of the form

$$\begin{aligned}\mu_t = \mu_0 + \int_0^t b_s\, ds + \int_0^t \sigma_s\, dW_s + \int_0^t \int_E (\delta(s,z)\mathbf{1}_{\{|\delta(s,z)|\leq 1\}})(\underline{\mu}-\underline{\nu})(ds,dz) \\ + \int_0^t \int_E (\delta(s,z)\mathbf{1}_{\{|\delta(s,z)|>1\}})\underline{\mu}(ds,dz).\end{aligned} \tag{3}$$

Here, we have μ_0 is $\mathcal{F}_0$-measurable. We also have the drift b_t is an $\mathbb{R}$-valued predictable process on the filtered probability space $(\Omega, \mathcal{F}, \{\widetilde{\mathcal{F}}_t\}_{t\in[0,T]}, \mathbb{P})$ such that its integral defined in Equation (3) is well-defined. Moreover, we have the variance σ_t^2 is $\mathbb{R}^+$-valued predictable process on $(\Omega, \mathcal{F}, \{\widetilde{\mathcal{F}}_t\}_{t\in[0,T]}, \mathbb{P})$ such that its integral defined in Equation (3) is well-defined. Finally, δ is a $\mathbb{R}$-valued predictable function on $\Omega \times \mathbb{R}^+ \times E$ such that both integral defined in Equation (3) are well-defined. Although we have extended the filtered space, in the sequel we keep the original space $\mathcal{B} = (\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{P})$. We pretend that the Grigelionis form above is defined on the original space, to avoid more complicated notation. For further details of definitions and notations, see Section 1.4.3 (pp. 47-49) in [Aït-Sahalia and Jacod, 2014]. The baseline model (3) is general in the sense that it is a slightly restricted version of a semimartingale. The semimartingale class is the most general class of "stochastic integrator" (see [Jacod and Shiryaev, 2003]).

The goal of this paper is the estimation of the branching ratio

$$\mathrm{BR} = \|\phi\|_1 = \int_0^\infty \phi(t)\, dt. \tag{4}$$

In terms of population interpretation, this branching ratio is the average number of children of an individual. From a financial perspective, the branching ratio can be interpreted as the average proportion of endogenous events. It is used as an empirical measure of the degree of endogeneity in the market.

We also develop a test for a branching ratio value. We consider a Wald test, which is based on the estimation of the branching ratio. For a branching ratio value $b \in [0, 1)$, we define the null hypothesis and the alternative hypothesis as $H_0 : \{\mathrm{BR} = b\}$ and $H_1 : \{\mathrm{BR} \neq b\}$. It extends the test for the absence of a Hawkes component in [Potiron et al., 2025]. Moreover, we propose a Wald test for a branching ratio value that does not depend on n against the branching ratio depends on n and tends to unity as $n \to \infty$. More specifically, we consider an alternative where the branching ratio depends on n of the form $BR = \widetilde{b}_n$, where $\widetilde{b}_n < 1$ and $\widetilde{b}_n \to 1$. We define the null hypothesis and the alternative hypothesis as $\overline{H}_0 : \{BR = b\}$ and $\overline{H}_1 : \{BR = \widetilde{b}_n\}$. As far as the authors know, all these tests are novel to the literature.

3 Estimation

In this section, we introduce the in-fill asymptotics. We also introduce empirical average and variance of local Poisson estimates. Finally, we introduce the branching ratio estimator, the test statistic for a branching ratio value and the test statistic for near criticality.

We prefer most of the time not to write explicitly the dependence on n, and any limit theorem refers to the convergence when $n \to \infty$. For inference purposes, we consider in-fill asymptotics with joint boosting of the baseline and the kernel, i.e.,

$$\lambda_t = n\mu_t + \int_0^t n\phi(n(t-s))\, dN_s.$$

Here, in-fill asymptotics are based on random observation times of order n within the time interval $[0, T]$ for a finite horizon time T. These in-fill asymptotics, also based on joint boosting, are used in [Clinet and Potiron, 2018], [Kwan et al., 2023], [Potiron and Volkov, 2025] but also available in [Erdemlioglu et al., 2025]. There are also compatible with the in-fill asymptotics used in the paper [Christensen and Kolokolov, 2024]. They are different from [Chen and Hall, 2013] in-fill asymptotics which considers no boosting of the kernel. Here, in-fill asymptotics are desirable because we can incorporate random features of the baseline into asymptotic variances in the CLT.

For a finite horizon T, we consider $M = \lfloor T/\Delta_n \rfloor$ intervals with equal length Δ_n such that $\bigcup_{i=1}^{M}[(i-1)\Delta_n, i\Delta_n) \subset [0,T)$, where $\lfloor \cdot \rfloor$ denotes the floor function. For $i = 1, \dots, M$, we define an estimator for local Poisson estimates on the i-th interval $[(i-1)\Delta_n, i\Delta_n)$ as

$$\widehat{\lambda}_i = \frac{1}{\Delta_n}(N_{i\Delta_n -} - N_{(i-1)\Delta_n}). \tag{5}$$

Then, we propose an estimator for empirical average and two estimators for empirical variance of local Poisson estimates as

$$\widehat{\text{Mean}} = \Delta_n \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} \widehat{\lambda}_i = N_T, \tag{6}$$

$$\widehat{\text{Var}}_1 = \sum_{i=2}^{\lfloor T/\Delta_n \rfloor} \left(\Delta_i \widehat{\lambda}\right)^2 \mathbf{1}_{\{|\Delta_i \widehat{\lambda}| \leq \alpha \Delta_n^{-\overline{\omega}}\}}, \tag{7}$$

$$\widehat{\text{Var}}_2 = \sum_{i=2}^{\lfloor T/(2\Delta_n) \rfloor} \left(\frac{\Delta_{2i-2}\widehat{\lambda} + \Delta_{2i-1}\widehat{\lambda}}{2}\right)^2 \mathbf{1}_{\{|(\Delta_{2i-1}\widehat{\lambda} + \Delta_{2i}\widehat{\lambda})/2| \leq \alpha \Delta_n^{-\overline{\omega}}\}}. \tag{8}$$

Here, we have that $\Delta_i \widehat{\lambda} = \widehat{\lambda}_i - \widehat{\lambda}_{i-1}$. We also have that $\alpha > 0$ and $\overline{\omega}$ are truncation parameters. Since the intensity explodes asymptotically, the three aforementioned estimators also diverge to infinity. The two variance estimators with a different scale are requested for the estimation of the branching ratio. We consider a truncation in our variance estimators since they would be contaminated by the jumps otherwise.

We define the diverging target values as

$$\text{Mean} = n \frac{1}{1 - BR} \int_0^T \mu_t \, dt, \tag{9}$$

$$\text{Var}_1 = n^2 \frac{1}{(1 - BR)^2} \int_0^T \left(\frac{2}{3}\sigma_t^2 + \frac{1}{1 - BR}\frac{2}{c}\mu_t\right) dt, \tag{10}$$

$$\text{Var}_2 = n^2 \frac{1}{(1 - BR)^2} \int_0^T \left(\frac{2}{3}\sigma_t^2 + \frac{1}{1 - BR}\frac{1}{2c}\mu_t\right) dt. \tag{11}$$

In practice, the order of observation number n is unknown. Thus, the length of intervals Δ_n cannot be chosen directly. Instead, we can estimate it as

$$\Delta_n = \frac{\sqrt{cT}}{\sqrt{N_T}}. \tag{12}$$

We use $c = 0.5$, which works the best in our numerical studies.

We can now estimate the branching ratio as

$$\widehat{BR} = 1 - \sqrt{\frac{3\widehat{\text{Mean}}}{2\Delta_n^2\big(\widehat{\text{Var}}_1 - \widehat{\text{Var}}_2\big)}}. \tag{13}$$

In Equation (13), we use the estimator from [Hardiman and Bouchaud, 2014] and we replace their variance by $\frac{2}{3}\big(\widehat{\text{Var}}_1 - \widehat{\text{Var}}_2\big)$ since we have a time-dependent baseline in this paper. This method does not require estimation of the kernel. It is based on the ratio of mean and variance of the empirical average and variance of local point process estimates. Its main practical advantage is that it is very fast as there is no optimization procedure.

For any $t \in [0,T]$, we define $\breve{\vartheta}_t$ as $\breve{\vartheta}_t = \frac{\mu_t}{c(1-BR)^3}$, $\breve{\sigma}_t$ as $\breve{\sigma}_t = \frac{\sigma_t}{1-BR}$. We also define the non diverging asymptotic variance of $(\text{Mean}, \text{Var}_1, \text{Var}_2)$ as

$$\Sigma = \int_0^T \begin{pmatrix} \frac{\mu_t}{1-BR} & 0 & 0 \\ 0 & \breve{\sigma}_t^4 + 4\breve{\sigma}_t^2\breve{\vartheta}_t + 12\breve{\vartheta}_t^2 & \frac{29}{24}\breve{\sigma}_t^4 + \frac{3}{2}\breve{\sigma}_t^2\breve{\vartheta}_t + \frac{3}{2}\breve{\vartheta}_t^2 \\ 0 & \frac{29}{24}\breve{\sigma}_t^4 + \frac{3}{2}\breve{\sigma}_t^2\breve{\vartheta}_t + \frac{3}{2}\breve{\vartheta}_t^2 & 2\breve{\sigma}_t^4 + 2\breve{\sigma}_t^2\breve{\vartheta}_t + \frac{3}{2}\breve{\vartheta}_t^2 \end{pmatrix} dt. \tag{14}$$

Then, we define the estimator of the non diverging asymptotic variance as

$$\widehat{\Sigma} = \begin{pmatrix} \frac{\Delta_n^2}{c}\widehat{\Sigma}_{11} & 0 & 0 \\ 0 & \frac{\Delta_n^4}{c^2}\widehat{\Sigma}_{22} & \frac{\Delta_n^4}{c^2}\widehat{\Sigma}_{23} \\ 0 & \frac{\Delta_n^4}{c^2}\widehat{\Sigma}_{23} & \frac{\Delta_n^4}{c^2}\widehat{\Sigma}_{33} \end{pmatrix}. \tag{15}$$

Here, the components of the matrix are defined as $\widehat{\Sigma}_{11} = \widehat{\text{Mean}}$, $\widehat{\Sigma}_{22} = \frac{3}{4}\widehat{\kappa}_{4,1} - 3\widehat{\eta}\widehat{\kappa}_{3,1} + 9\widehat{\eta}^2\widehat{\kappa}_{2,1}$, $\widehat{\Sigma}_{23} = \frac{29}{32}\widehat{\kappa}_{4,1} - \frac{69}{8}\widehat{\eta}\widehat{\kappa}_{3,1} + \frac{63}{8}\widehat{\eta}^2\widehat{\kappa}_{2,1}$, and $\widehat{\Sigma}_{33} = \frac{3}{2}\widehat{\kappa}_{4,2} - 6\widehat{\eta}\widehat{\kappa}_{3,2} + 18\widehat{\eta}^2\widehat{\kappa}_{2,2}$. We also define $\widehat{\kappa}_{2,1}$ as $\widehat{\kappa}_{2,1} = \Delta_n^{-3}\sum_{i=2}^{\lfloor T/\Delta_n\rfloor}\widehat{\lambda}_i^2\mathbf{1}_{\{|\Delta_i\widehat{\lambda}|\leq n\varpi_i\}}$, and $\widehat{\kappa}_{2,2}$ as $\widehat{\kappa}_{2,2} = \Delta_n^{-3}\sum_{i=2}^{\lfloor T/(2\Delta_n)\rfloor}(\frac{\widehat{\lambda}_{i-1}+\widehat{\lambda}_i}{2})^2\mathbf{1}_{\{|(\Delta_{2i-1}\widehat{\lambda}+\Delta_{2i}\widehat{\lambda})/2|\leq n\varpi_i\}}$. Additionally, we can define $\widehat{\kappa}_{3,1}$ as $\widehat{\kappa}_{3,1} = \Delta_n^{-2}\sum_{i=2}^{\lfloor T/\Delta_n\rfloor}\widehat{\lambda}_i(\Delta_i\widehat{\lambda})^2\mathbf{1}_{\{|\Delta_i\widehat{\lambda}|\leq n\varpi_i\}}$. Moreover, $\widehat{\kappa}_{3,2}$ is defined as

$$\widehat{\kappa}_{3,2} = \Delta_n^{-2}\sum_{i=2}^{\lfloor T/(2\Delta_n)\rfloor}\frac{\widehat{\lambda}_{i-1}+\widehat{\lambda}_i}{2}\Big(\frac{\Delta_{2i-2}\widehat{\lambda}+\Delta_{2i-1}\widehat{\lambda}}{2}\Big)^2\mathbf{1}_{\{|(\Delta_{2i-1}\widehat{\lambda}+\Delta_{2i}\widehat{\lambda})/2|\leq n\varpi_i\}}.$$

In addition, we define $\widehat{\kappa}_{4,1}$ as $\widehat{\kappa}_{4,1} = \Delta_n^{-1} \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} (\Delta_i \widehat{\lambda})^4 \mathbf{1}_{\{|\Delta_i \widehat{\lambda}| \leq n\varpi_i\}}$. Finally, $\widehat{\kappa}_{4,2}$ is defined as $\widehat{\kappa}_{4,2} = \Delta_n^{-1} \sum_{i=1}^{\lfloor T/(2\Delta_n) \rfloor} (\frac{\Delta_{2i-2}\widehat{\lambda} + \Delta_{2i-1}\widehat{\lambda}}{2})^4 \mathbf{1}_{\{|(\Delta_{2i-1}\widehat{\lambda} + \Delta_{2i}\widehat{\lambda})/2| \leq n\varpi_i\}}$, and $\widehat{\eta}$ as $\widehat{\eta} = \frac{2}{3} \frac{\Delta_n^2 (\widehat{\mathrm{Var}}_1 - \widehat{\mathrm{Var}}_2)}{\widehat{\mathrm{Mean}}}$.

We denote the gradient of a function f by ∇f. We introduce $x = \big(0, \mathrm{Var}_1, \mathrm{Var}_2\big)^T$ and the function $f_1(x) = 1 - \sqrt{\frac{3n\mathrm{Mean}}{2c(x_2 - x_3)}}$. We define the asymptotic variance for estimation of branching ratio as

$$AVar = \nabla f_1(x)^T \Sigma \nabla f_1(x). \tag{16}$$

We introduce $\widehat{x} = [0, \widehat{\mathrm{Var}}_1, \widehat{\mathrm{Var}}_2]^T$ and the function $f_2(x) = 1 - \sqrt{\frac{3\widehat{\mathrm{Mean}}}{2\Delta_n^2 (x_2 - x_3)}}$. Then, we define the estimator of the asymptotic variance for estimation of branching ratio as

$$\widehat{AVar} = \nabla f_2(\widehat{x})^T \widehat{\Sigma} \nabla f_2(\widehat{x}). \tag{17}$$

Moreover, we introduce our Wald test statistic for a branching ratio value, i.e.,

$$S = \frac{\Delta_n^{-1} (\widehat{\mathrm{BR}} - b)^2}{\widehat{AVar}}. \tag{18}$$

We introduce the function $f_3(x) = \frac{2c(x_2 - x_3)}{3n\mathrm{Mean}}$. We define the asymptotic variance for the Wald test statistic where the branching ratio value does not depend on n as

$$\overline{AVar} = \nabla f_3(x)^T \Sigma \nabla f_3(x). \tag{19}$$

We introduce the function $f_4(x) = \frac{2c(x_2 - x_3)}{3n\widehat{\mathrm{Mean}}}$. Then, we define the estimator of the asymptotic variance for the Wald test statistic where the branching ratio value does not depend on n as

$$\widehat{\overline{AVar}} = \nabla f_4(\widehat{x})^T \widehat{\Sigma} \nabla f_4(\widehat{x}). \tag{20}$$

Furthermore, we introduce our Wald test statistic for a branching ratio value that does not depend on n, i.e.,

$$\overline{S} = \frac{\Delta_n^{-1}}{\widehat{\overline{AVar}}} \left(\frac{2\Delta_n^2 (\widehat{\mathrm{Var}}_1 - \widehat{\mathrm{Var}}_2)}{3\widehat{\mathrm{Mean}}} - \frac{1}{(1-b)^2} \right)^2. \tag{21}$$

The reason why we consider a Wald test statistic $\overline{S}$ based on estimation of $(1-b)^{-2}$ rather than b itself is that $(1 - \widetilde{b}_n)^{-2}$ explodes. Thus, this test statistics is more likely to explode under the alternative hypothesis $\overline{H}_1$.

4 Theory

In this section, our main result characterizes feasible statistics induced by CLT for estimation of branching ratio. We also give the limit theory of the test for a branching ratio value and the test for a branching ratio value that does not depend on n. The results are obtained with in-fill asymptotics when $n \to \infty$ and T is fixed.

Let us introduce a set of conditions required for the existence of Hawkes processes with a time-dependent baseline driven by an Itô semimartingale.

Condition 1. (a) The baseline is positive a.e. a.s., i.e., $\mathbb{P}(\mu_t > 0 \forall t \in [0,T]) = 1$.

(b) The baseline is integrable a.s., i.e., $\mathbb{P}(\int_0^T \mu_s \, ds < \infty) = 1$.

(c) For any $0 \leq t \leq T$, we have $\mathcal{F}_t = \widetilde{\mathcal{F}}_t \vee \mathcal{F}_t^N$, where the filtration $\widetilde{\mathcal{F}}_t$ is independent from the other filtration $\mathcal{F}_t^N$. We also have $\underline{N}_t$ is a 2-dimensional $\mathcal{F}_t$-adapted Poisson process of intensity 1 that generates N_t, i.e., $N_t = \int_{[0,t]\times\mathbb{R}} \mathbf{1}_{[0,\lambda_s]}(x)\underline{N}(ds \times dx)$.

(d) The branching ratio is strictly less than one, i.e., $BR < 1$.

Condition 1 (a) implies that the point process is well-defined. Condition 1 (b) is already a condition in the simpler case of heterogeneous Poisson processes without a kernel (see [Daley and Vere-Jones, 2003]). Condition 1 (c) corresponds to Poisson imbedding ([Brémaud and Massoulié, 1996], Section 3, pp. 1571-1572) and assumes independence between $\widetilde{\mathcal{F}}_t$ and $\mathcal{F}_t^N$. In particular, N_t is defined as the point process counting the points of $\underline{N}$ below the curve $t \to \lambda_t$. Finally, Condition 1 (d) is necessary to obtain a stationary intensity with finite first moment (see Lemma 1 (p. 495) in [Hawkes and Oakes, 1974] and Theorem 1 (p. 1567) in [Brémaud and Massoulié, 1996]). Condition 1 corresponds exactly to Assumption 1 in [Potiron et al., 2025].

For any positive function f, we define its L^1 norm as $\|f\|_1 = \int_0^\infty f(t)dt$. We define an alternative drift as $b'_t = b_t - \int_E \delta(t,z)\mathbf{1}_{\{|\delta(t,z)|\leq 1\}}F_t(dz)$ for any $t \in [0,T]$. Finally, we define $V_a^b(f)$ as the total variation of f from a to b. Let us introduce a set of conditions required for estimation of branching ratio.

Condition 2. (a) The kernel satisfies the *short-range condition*, i.e., $\int_0^\infty t\phi(t)dt < \infty$.

(b) For any $k \in \mathbb{N}$ with $k \geq 2$, the L^1 norm of ϕ^k is finite, i.e., $\|\phi^k\|_1 < \infty$.

(c) There exists a $c > 0$ such that $n\Delta_n^2 \xrightarrow{\mathbb{P}} c$.

(d) There exists a $\beta \in [0,1)$ such that $\sup_{0\leq t\leq T} \int \min(|x|^r, 1)F_t(dx)$ is a.s. finite.

(e) The truncation level satisfies $\overline{\omega} \in (0, 1/(4-2\beta))$.

(f) For any $k \in \mathbb{N}_*$ and any $t \in [0,T]$, we have $|b_t|^k < \infty, |\sigma_t|^k < \infty$.

(g) We have that $\left[\exp\left(\frac{1}{2}\int_0^T \frac{(b_s')^2}{\sigma_s^2}ds\right)\right] < \infty$.

(h) The volatility process is a semimartingale, i.e., $\sigma_t^2 = A_t + M_t^{(\sigma)}$, where A_t is a $\mathcal{F}_t$-adapted cadlag process with finite variation and $M^{(\sigma)}$ is a square-integrable martingale. Moreover, $\mathbb{E}|V_0^T(A)|^k < \infty$ and $\mathbb{E}|\sigma_t - \sigma_s|^k \leq C(t-s)^{k\gamma}$ for a $\gamma > 0$ and any $k \in \mathbb{N}_*$.

Conditions 2 (a) and (b) put some restrictions on the kernel shape. Condition 2 (c) is natural for local estimation. Conditions 2 (d) and (e) are due to the presence of jumps. Condition 2 (f) is used in the proof of jumps and also in the proof of the CLT. Condition 2 (g) is required to apply the Girsanov theorem. Condition 2 (h) is used in the proof of the CLT. Condition 2 corresponds exactly to Assumption 2 in [Potiron et al., 2025].

We define $\mathcal{N}(0,1)$ as a standard normal variable. We denote $\xrightarrow{\mathcal{D}-s}$ as the $\mathcal{F}_t$-stable weak convergence for the Skorokhod topology on $\mathbb{D}([0,T],\mathbb{R})$. Moreover, we introduce the local asymptotic variance $w_t^2 > 0$ which satisfies

$$w_t^2 = \nabla f_1(x)^T \begin{pmatrix} \frac{\mu_t}{1-BR} & 0 & 0 \\ 0 & \breve{\sigma}_t^4 + 4\breve{\sigma}_t^2\breve{\vartheta}_t + 12\breve{\vartheta}_t^2 & \frac{29}{24}\breve{\sigma}_t^4 + \frac{3}{2}\breve{\sigma}_t^2\breve{\vartheta}_t + \frac{3}{2}\breve{\vartheta}_t^2 \\ 0 & \frac{29}{24}\breve{\sigma}_t^4 + \frac{3}{2}\breve{\sigma}_t^2\breve{\vartheta}_t + \frac{3}{2}\breve{\vartheta}_t^2 & 2\breve{\sigma}_t^4 + 2\breve{\sigma}_t^2\breve{\vartheta}_t + \frac{3}{2}\breve{\vartheta}_t^2 \end{pmatrix} \nabla f_1(x). \tag{22}$$

We provide now the CLT for estimation of branching ratio. This extends [Hardiman and Bouchaud, 2014] (see Section III), who only show the consistency of the estimation procedure when the baseline is constant. This extends Theorem 4.1 in [Potiron et al., 2025]. The theorem is obtained with in-fill asymptotics when $n \to \infty$ and T is fixed. The convergence rate is $\Delta_n^{-1/2}$, which by Condition 2 (c) is asymptotically equivalent to $n^{1/4}$.

Theorem 1. *Under Conditions 1 and 2, there is a standard Brownian extension of $\mathcal{B}$, with the canonical standard Brownian motion $\widetilde{W}_t$ such that*

$$\Delta_n^{-\frac{1}{2}}(\widehat{\mathrm{BR}} - \mathrm{BR}) \xrightarrow{\mathcal{D}-s} \int_0^T w_t d\widetilde{W}_t. \tag{23}$$

Moreover, we have the normalized CLT with feasible variance, i.e.,

$$\frac{\Delta_n^{-\frac{1}{2}}(\widehat{\mathrm{BR}} - \mathrm{BR})}{\sqrt{\widehat{AVar}}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{24}$$

We define $q(u)$ as the quantile function of the chi-squared distribution with one degree of freedom. The following corollary gives the limit theory of the Wald test for a branching ratio value. This is an application of Theorem 1. This extends Corollary 5.4 in [Potiron et al., 2025]. The corollary is obtained with in-fill asymptotics when $n \to \infty$ and T is fixed.

Corollary 1. *Under Conditions 1 and 2, the test statistic S converges in distribution to a chi-squared random variable with one degree of freedom under the null hypothesis H_0. Moreover, the test statistic S is consistent under the alternative hypothesis H_1, i.e., we have $\mathbb{P}(S > q(u) \mid H_1) \to 1$ for any $0 < u < 1$.*

We denote the kernel under the alternative hypothesis $\overline{H}_1$ by $\widetilde{\phi}$. Let $\psi(t) = \sum_{k=1}^{\infty} (\widetilde{\phi})^{*k}(t)$ where $(\widetilde{\phi})^{*1} = \widetilde{\phi}$ and $(\widetilde{\phi})^{*k}$ denotes the convolution product of $(\widetilde{\phi})^{k-1}$ with the function $\widetilde{\phi}$ for $k \geq 2$. Finally, we define $\rho(t)$ as $\rho(t) = \psi(t)/\|\psi\|_1$ for $t \in \mathbb{R}^+$. In what follows, we provide a set of conditions required for the consistency under the alternative hypothesis of the Wald test statistic for a branching ratio value that does not depend on n.

Condition 3. (a) The intensity is equal to

$$\lambda_t = n\mu_t + \widetilde{b}_n \int_0^t n\widetilde{\phi}(n(t-s))\, dN_s.$$

(b) There exists $\lambda > 0$ such that the sequence $\widetilde{b}_n$ satisfies $n(1 - \widetilde{b}_n) \to \lambda$.

(c) The non-negative measurable function $\widetilde{\phi}\colon \mathbb{R}^+ \to \mathbb{R}^+$ satisfies $\|\widetilde{\phi}'\|_1 < \infty$ and

$$\int_0^\infty t\widetilde{\phi}(t)dt < \infty. \tag{25}$$

Moreover, $\widetilde{\phi}$ is differentiable with derivative $\widetilde{\phi}'$ such that $\sup_{t\in\mathbb{R}^+} |\widetilde{\phi}'(t)| < \infty$ and $\|\widetilde{\phi}'\|_1 < \infty$.

(d) We have that the function ρ is bounded uniformly, i.e. $\sup_{t\in\mathbb{R}^+, n\in\mathbb{N}} |\rho(t)| < \infty$.

Condition 3 (a) extends Section 2.3 from [Jaisson and Rosenbaum, 2015], who considers time-invariant baseline and $T \to \infty$ for inference purposes, to the time-dependent baseline and in-fill asymptotics case. Condition 3 (b) is necessary to obtain the existence of the limit and can be compared to Equation (3) in [Jaisson and Rosenbaum, 2015]. Condition 3 (c) corresponds exactly to Assumption (1) in [Jaisson and Rosenbaum, 2015]. Condition 3 (d) extends Assumption (2) in [Jaisson and Rosenbaum, 2015].

The following theorem gives the limit theory of the Wald test statistic for a branching ratio value that does not depend on n. Under the alternative hypothesis $\overline{H}_1$, this extends Theorem 2.2 in [Jaisson and Rosenbaum, 2015]. The theorem is obtained with in-fill asymptotics when $n \to \infty$ and T is fixed.

Theorem 2. *We assume that Conditions 1 and 2 hold. Then, the test statistic $\overline{S}$ converges in distribution to a chi-squared random variable with one degree of freedom under the null hypothesis $\overline{H}_0$. If we also assume Condition 3, the test statistic $\overline{S}$ is consistent under the alternative hypothesis $\overline{H}_1$, i.e., we have $\mathbb{P}(\overline{S} > q(u) \mid \overline{H}_1) \to 1$ for any $0 < u < 1$.*

5 Simulation studies

In this section, we conduct simulation studies to document how the estimator of branching ratio and two tests behave. We also show that we improve branching ratio estimation for this more realistic baseline.

We consider the following simulation design to be as close as possible from the data application in finance. We set $T = 1$, i.e., 6.5 hour long day of trading. The order of the observation number n varies from 50,000 to 1,000,000. With these realistic values, the simulation design allows for both less traded and highly traded stocks. The number of replications is equal to 1,000. We use the python package `tick` from [Bacry et al., 2017] for the generation of the point process.

We define the intensity process as

$$\lambda_t \;=\; n(1 - BR)\big(\mu_t^{\mathrm{C}} + \mu_t^{\mathrm{B}}\big) + \int_0^t n\phi(n(t-s))dN_s. \tag{26}$$

Here, the component of the baseline μ_t^{C} satisfies a square root process (SRP)

$$d\mu_t^{\mathrm{C}} \;=\; 30(b_t - \mu_t^{\mathrm{C}})dt + 3\sqrt{\mu_t^{\mathrm{C}}}dW_t. \tag{27}$$

Here, b_t is a solution of the ordinary differential $dr_t = 30(b_t - r_t)dt$ with inverse J-shape r_t defined as

$$r_t = 20\left((t - 0.53)^4 + \frac{1}{24}\right), \tag{28}$$

and $\mu_0^{\mathrm{C}} = r_0$. We have that the drift term in Equation (27) ensures mean reversion of μ_t^{C} to the process b_t. Moreover, b_t pushes μ_t^{C} to follow the inverse J-shape nonrandom term r_t. In Equation (27), the diffusion term $\sqrt{\mu_t^{\mathrm{C}}}dW_t$ is the random fluctuation. The Feller condition (see [Feller, 1951]) is satisfied with $30 \times b_t \geq 3^2$ for any $t \in [0, T]$, thus μ_t^{C} is positive.

In Equation (27), μ_t^{B} are the intensity jumps (see [Rambaldi et al., 2018]). They are defined as a sudden occurrence of a big number of exogenous points for a short period of time, i.e., around one second. The arrival time of jumps z_i is sampled from an homogeneous Poisson process with rate $2/T$. The size of the bursts Z_i are drawn from $\max(\mathcal{N}(200n, (50n)^2), 50n)$. The intensity bursts have the form

$$\mu_t^{\mathrm{B}} = \sum_{z_i \leq t} Z_i \mathbf{1}_{\{(t - z_i) \in [0, 1/(3600 \times 6.5)]\}}. \tag{29}$$

The parameter values come from our empirical application and the results from [Rambaldi et al., 2018] (p. 6), where the authors report an average number of jumps between 1.95–3.25 for a 6.5-hour period.

In Equation (26), we consider an exponential kernel defined as $\phi_e(t) = 1.6e^{-2t}$ and a power kernel defined as $\phi_p(t) = 1.6(1+t)^{-3}$. With these kernel values, the branching ratio is equal to $BR = 0.8$, which is the average value that we obtain in our own empirical application and in the results of [Filimonov and Sornette, 2012]. Finally, we set the truncation parameters as $\alpha = 1$ and

$$\varpi = \Delta_n^{-\frac{1}{4}} \sqrt{\frac{1}{\lfloor T/\Delta_n \rfloor} \sum_{i=1}^{\lfloor T/\Delta_n \rfloor} (\Delta_i \widehat{\lambda})^2}. \tag{30}$$

We consider the following models to disentangle the effects. First, we set Model 1 as a null kernel and a constant baseline, i.e., $\lambda_t = n$. Second, we set Model 2 as a null kernel and a J-shape baseline, i.e., $\lambda_t = 20\big((t-0.53)^4 + \frac{1}{24}\big)n$. Third, we set Model 3 as a null kernel and a J-shape + SRP + jump baseline, i.e., $\lambda_t = n(\mu_t^{\mathrm{C}} + \mu_t^{\mathrm{B}})$. Then, we set Model 4 as an exponential kernel and a constant baseline, i.e., $\lambda_t = n + \int_0^t n\phi_e(n(t-s))dN_s$. We also set Model 5 as an exponential kernel and a J-shape baseline, i.e., $\lambda_t = n\mu_t + \int_0^t n\phi_e(n(t-s))dN_s$ where $\mu_t = 20(1-BR)\big((t-0.53)^4 + \frac{1}{24}\big)$. We set Model 6 as an exponential kernel and a J-shape + SRP + burst baseline, i.e., $\lambda_t = n(1-BR)(\mu_t^{\mathrm{C}} + \mu_t^{\mathrm{B}}) + \int_0^t n\phi_e(n(t-s))dN_s$. We set Model 7 as a power kernel, and a constant baseline as $\lambda_t = n + \int_0^t n\phi_p(n(t-s))dN_s$. We set Model 8 as a Power kernel and a J-shape baseline, i.e., $\lambda_t = n\mu_t + \int_0^t n\phi_p(n(t-s))dN_s, \mu_t = 20(1-BR)\big((t-0.53)^4 + \frac{1}{24}\big)$. Moreover, we set Model 9 as a power kernel and a J-shape + SRP + burst baseline, i.e., $\lambda_t = n(1-BR)(\mu_t^{\mathrm{C}} + \mu_t^{\mathrm{B}}) + \int_0^t n\phi_p(n(t-s))dN_s$. Finally, we set Model 10 as nearly unstable Hawkes with constant baseline, i.e., $\lambda_t = \sqrt{n} + \int_0^t \widetilde{b}_n \phi^{\sqrt{n}}(t-s)dN_s$, where $\phi^{\sqrt{n}}(t) = 2\sqrt{n}e^{-2\sqrt{n}t}$ and $\widetilde{b}_n = 1 - n^{-1/2}$ with the branching ratio $BR = \widetilde{b}_n$ which goes to unity. These models are summarized in Table 1.

In general, the intensity jumps μ^{B} follow Equation (29). In the case of power kernel, we first generate points without the jump and then add points whose intensity follows $(1-BR)^{-1}\mu^{\mathrm{B}}$. It is due to the implemented function in the package `tick` taking over a day to generate points when there is a jump. However, it does not give any significant differences in the results.

Figure 1 provides a comparison between simulated intensity with Model 9 (left panel) and intensity based on AAPL (Apple) data on April 1st 2016 (right panel). The intensity is obtained from one-

Table 1: Summary of models.

	Baseline Model (μ_t)		
Kernel	Constant	J-shape	J-shape + SRP + jump
Null	Model 1	Model 2	Model 3
Exponential	Model 4	Model 5	Model 6
Power	Model 7	Model 8	Model 9
Nearly unstable	Model 10		

minute intervals. The simulated process captures the U-shaped pattern and intensity jump well. It also exhibits some random fluctuation of the baseline intensity. These patterns can also be seen in the data that justify our simulation design being realistic.

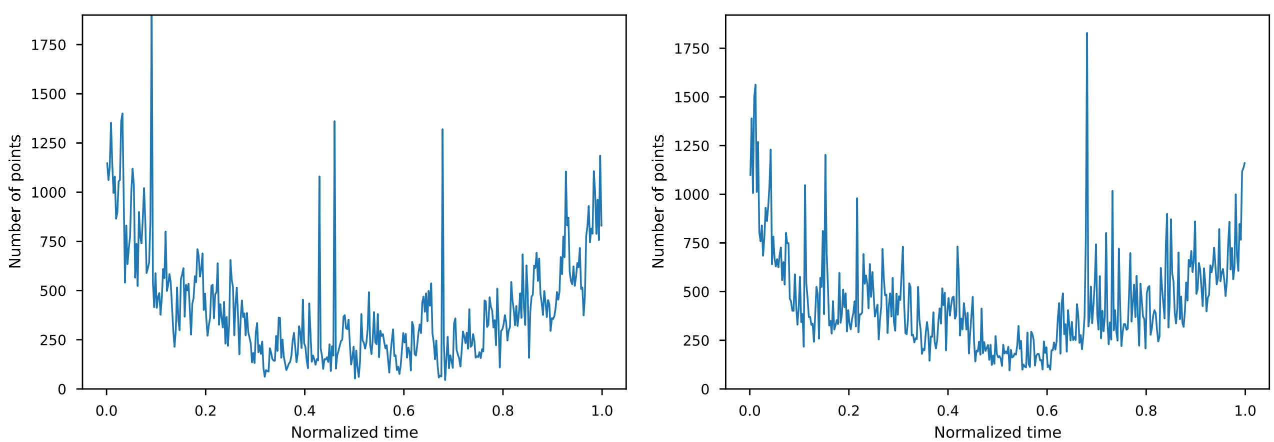

Figure 1: Comparison between simulated intensity with Model 9 (left panel) and intensity based on AAPL data on April 1st 2016 (right panel).

Table 2 and Figure 2 report the summary statistics and the histogram for the branching ratio with Models 1-10. The order of the observation number n is 150,000 and 1,000,000. The absolute value of the mean ranges from 0% to 28%, with an average of 10%. It has an average of 7% for the statistics with unfeasible variance, and an average of 13% for the statistics with feasible variance. Overall, the statistics are slightly biased, especially when the variance is unfeasible. However, the bias gets smaller when n increases. The variance ranges from 100% to 110%, with an average of 103%. It has an average of 101% for the statistics with unfeasible variance, and an average of 105% for the statistics with feasible variance. Overall, the variance is close to unity.

Table 2: Summary statistics for estimation of branching ratio with Models 1-10. The order of the observation number n is 150,000 and 1,000,000, and the number of replications is 1,000.

n	150,000				1,000,000			
Variance	Unfeasible		Feasible		Unfeasible		Feasible	
Model	Mean	Variance	Mean	Variance	Mean	Variance	Mean	Variance
Model 1	0.0096	1.0332	0.0178	1.0574	-0.0424	1.0108	-0.0840	1.0186
Model 2	-0.0281	1.0210	-0.0650	1.0358	0.0163	1.0347	0.0310	1.0567
Model 3	0.0706	1.0256	0.1526	1.0595	0.0402	1.0058	0.0697	1.0078
Model 4	-0.0834	1.0144	-0.1552	1.0289	-0.0243	1.0461	-0.0550	1.0809
Model 5	-0.0542	1.0609	-0.1353	1.1020	-0.0636	1.0184	-0.1127	1.0228
Model 6	-0.0465	1.0240	-0.0873	1.0544	0.0224	1.0066	0.0371	1.0113
Model 7	-0.1474	1.0019	-0.2808	1.0028	-0.0823	1.0086	-0.1563	1.0044
Model 8	-0.1313	1.0255	-0.2689	1.0324	0.0823	1.0121	-0.1295	1.0210
Model 9	-0.0924	1.0339	-0.2094	1.0982	-0.0214	1.0208	0.1247	1.0550
Model 10	-0.0120	0.9991	-0.0092	0.9978	0.0004	1.0115	0.0016	1.0175

We compare our branching ratio estimator to several branching ratio estimators from the literature. The competitors are Hardiman-Bouchaud estimator (H&B), a local average of the Hardiman-Bouchaud estimator with 2 hour-long intervals (H&B(2h)), MLE with exponential kernel implemented in the package `tick`, EM implemented in `tick`, and EM(O&H&A) by [Omi et al., 2017]. In all these methods,

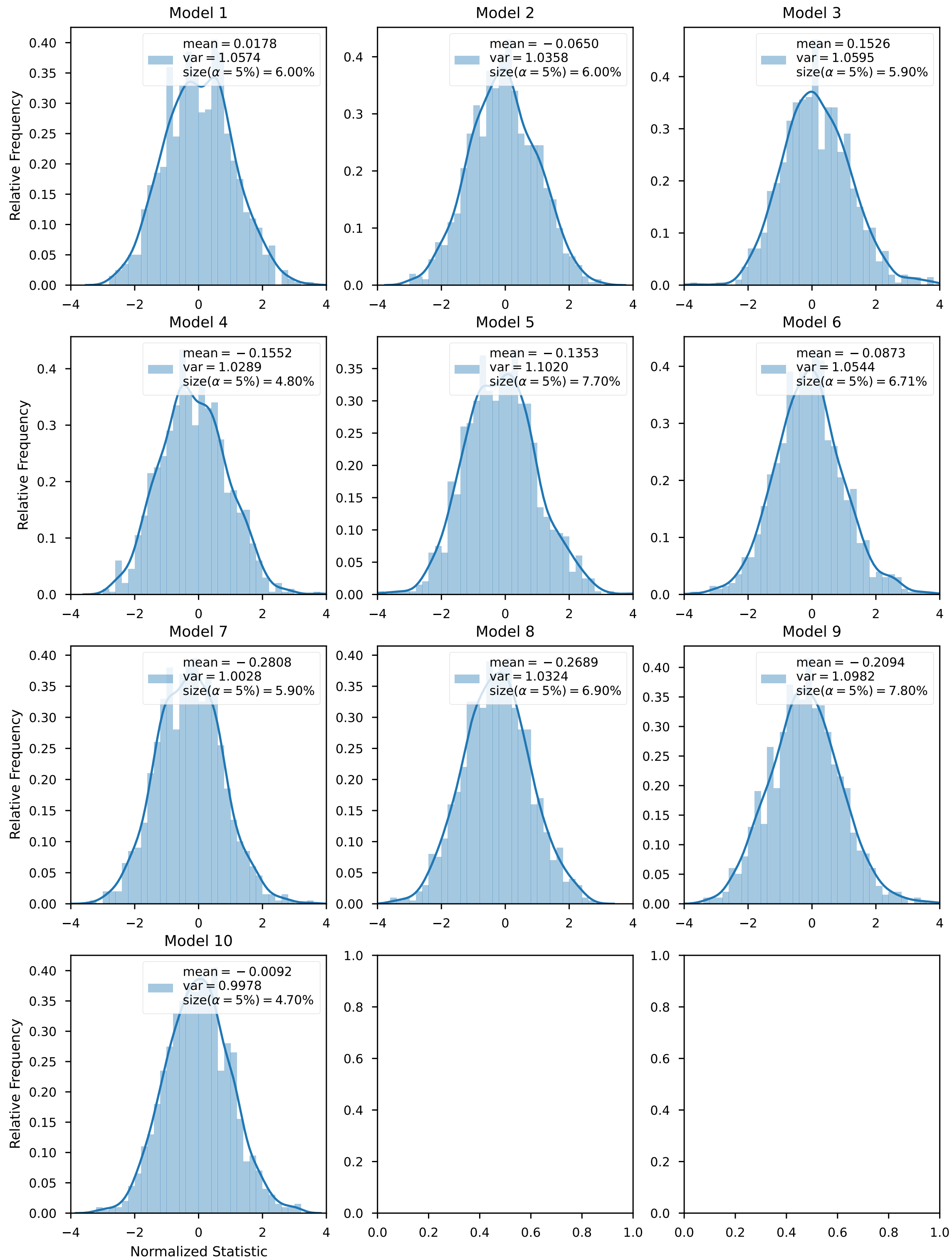

Figure 2: Histogram of the normalized CLT with feasible variance (24) for branching ratio estimation with Models 1-10. The order of the observation number n is 150,000, and the number of replications is 1,000.

only EM(O&H&A) is adapted for a time-dependent baseline.

Table 3: Computation time for estimation of branching ratio with Model 6

Method	n			
	1000	4000	10000	50000
Our method	0.002 sec	0.003 sec	0.003 sec	0.007 sec
H&B	0.001 sec	0.001 sec	0.001 sec	0.002 sec
H&B(2h)	0.001 sec	0.002 sec	0.002 sec	0.003 sec
MLE(exp)	0.049 sec	0.062 sec	0.049 sec	0.184 sec
EM	0.016 sec	0.061 sec	0.146 sec	0.851 sec
EM(O&H&A)	18.752 sec	67.487 sec	105.432 sec	723.410 sec

Table 3 reports the computation time for several estimation methods. Our method and Hardiman-Bouchaud method achieve a millisecond computation time even for $n = 50{,}000$, while the other methods require a longer time ranging from 20 times bigger for MLE(exp) to 10^5 times bigger for EM(O&H&A). This illustrates well that Hardiman-Bouchaud and our method do not require any optimization procedure. As the computational time for EM(O&H&A) is too big, we unfortunately have to drop it in what follows.

Figure 3 gives the histogram of several branching ratio estimation methods with Models 1–9. For the simple Model 1, most estimation methods behave properly. However, most estimation methods are severely biased when the baseline is time-dependent, and this is the most pronounced for Model 3. On the contrary, our method gives estimation results with a very small bias and relatively small variance under all the models. These results are also visible in Table 4, which reports the mean absolute error (MAE) and mean squared error (MSE) of the estimation results. This seems to indicate that our estimation method is adequate.

Table 5 reports the percentage of rejections at the 5% level of the null hypothesis for the two tests

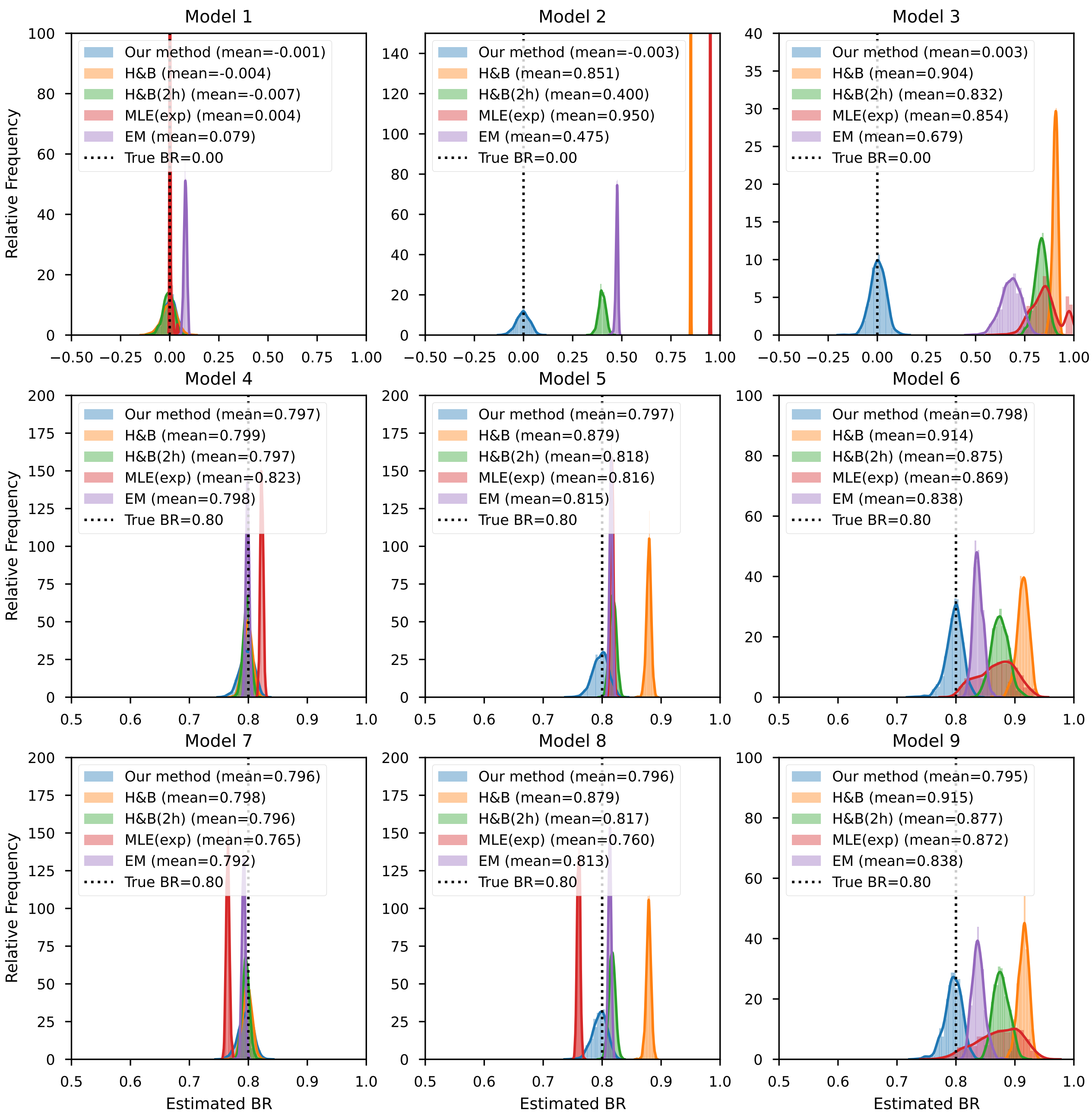

Figure 3: Histogram of several branching ratio estimation methods with Models 1–9. The scaling parameter n is set to 150,000, and the number of replications is 1,000.

Table 4: Mean absolute error (MAE) and mean squared error (MSE) of several branching ratio estimation methods with Models 1–9. The scaling parameter n is set to 150,000, and the number of replications is 1,000.

Methods	Models								
	1	2	3	4	5	6	7	8	9
	MAE($\times 10^2$)								
Our method	2.668	2.673	3.105	0.981	1.055	1.095	0.991	1.027	1.193
H&B	2.926	85.074	90.443	0.627	7.908	11.358	0.650	7.889	11.464
H&B(2h)	2.201	39.971	83.190	0.528	1.820	7.476	0.595	1.694	7.691
MLE(exp)	0.376	95.017	85.425	2.268	1.649	6.916	3.485	3.987	7.299
EM	7.945	47.459	67.906	0.256	1.490	3.767	0.804	1.280	3.794
	MSE($\times 10^2$)								
Our method	0.108	0.112	0.152	0.015	0.018	0.021	0.016	0.017	0.024
H&B	0.138	72.376	81.817	0.006	0.627	1.300	0.007	0.624	1.322
H&B(2h)	0.078	16.009	69.298	0.004	0.036	0.579	0.005	0.032	0.608
MLE(exp)	0.008	90.283	73.485	0.052	0.028	0.575	0.122	0.160	0.657
EM	0.637	22.527	46.372	0.001	0.023	0.149	0.007	0.017	0.154

with Models 1-10. The order of the observation number n is 50,000, 150,000 and 1,000,000. The size ranges from 4.5% to 6.1%, with an average of 5.6%. It has an average of 5.8% with the test for a branching ratio value, and an average of 5.0% with the test for near criticality. Overall, the test for a branching ratio value is slightly oversized while the size of the test for near criticality is adequate. The power is always equal to 100%, and thus is also adequate.

Table 5: Percentage of rejections at the 5% level of the null hypothesis for the two tests with Models 1-10. The order of the observation number n is 50,000, 150,000 and 1,000,000, and the number of replications is 1,000.

	Test for a branching ratio value $BR = 0.8$									
	Size						Power			
n	4	5	6	7	8	9	1	2	3	
50,000	5.4	5.8	5.9	6.1	6.0	5.8	100	100	99.9	
150,000	5.8	5.7	5.6	5.9	6.1	5.5	100	100	100	
1,000,000	5.3	6.0	5.7	5.7	5.4	6.1	100	100	100	
	Test for a branching ratio value that does not depend on n									
	Size									Power
n	1	2	3	4	5	6	7	8	9	10
50,000										
150,000										
1,000,000										

6 Empirical application

Our empirical application focuses on the S&P500 E-mini futures. They are liquid contracts traded on the Chicago Mercantile Exchange. We obtain the 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 Tick-datamarket database. The data set covers the period from January 2020 to December 2021. All index quotes are considered during normal trading hours.

Now, we turn to testing the two hypotheses formulated in Section **??**, i.e., branching ratio value and near criticality. For each day in the sample, we perform the tests following Corollary 1 and Theorem 2. Figure 5 shows corresponding test statistics revealing rejection of the null hypothesis in both cases.

`AOS/Intensity.pdf`

Figure 4: Estimated intensity of the S&P500 E-mini futures quotes. The number of the mid-quote price changes in millions per minute is shown.

To verify that our test results are not distorted due to a multiple statistical inference problem, we implement the sequential Bonferroni procedure of [?] for all p-values. The adjusted p-values computed at

AOS/test_empirics1.pdf

Figure 5: Test statistics for the null hypothesis in the two tests from Sections **??** and **??** with the 5% critical value.

the 1% level provide identical conclusions about all hypotheses, confirming the statistical robustness of our results. Another robustness check of our test results is conducted following [Bajgrowicz et al., 2016] and the results are in agreement with the Bonferroni corrected tests.

7 Conclusion

In this paper, we have studied Hawkes processes where the baseline is driven by an Itô semimartingale. We have considered estimation of the branching ratio. The estimation procedure was based on empirical average and variance of local point process estimates. We gave the CLT for the estimation procedure. We have also developed a test for any branching ratio value, and another one for near criticality or the branching ratio is null. The results are obtained with in-fill asymptotics. Simulation studies have corroborated asymptotic theory. An empirical application on high-frequency data of the E-mini S&P500 future contracts showed that the branching ratio is around 0.7 and 0.8, and that the tests for a branching ratio tending to unity or equal to zero are rejected.

There are some follow-up questions, which are left for future work. First, the case of a multidimensional branching ratio matrix could have more interesting applications in finance. Second, we could consider, under the alternative, the fractional Brownian motion (see [Jaisson and Rosenbaum, 2016]) under the heavy-tailed condition (i.e. $1 - \int_0^x \phi(t)\, dt \sim Cx^{-\alpha}$ for some $\alpha \in (0, 1)$). The fractional Brownian motion does not satisfy the short-range condition, thus it would be theoretically challenging to explore this problem.

The code is available at `https://github.com/SeunghyeonTonyYu/TSRV2Hawkes`.

Supplementary materials

All proofs of the theory can be found in the supplementary materials. These proofs are based on [Jaisson and Rosenbaum, 2015] and [Potiron et al., 2025].

Funding

Financial support from the Japanese Society for the Promotion of Science under grant 23H00807 (Potiron) is gratefully acknowledged.

References

[Achab et al., 2018a] Achab, M., Bacry, E., Gaïffas, S., Mastromatteo, I., and Muzy, J.-F. (2018a). Uncovering causality from multivariate Hawkes integrated cumulants. *Journal of Machine Learning Research*, 18(192):1–28.

[Achab et al., 2018b] Achab, M., Bacry, E., Muzy, J.-F., and Rambaldi, M. (2018b). Analysis of order book flows using a non-parametric estimation of the branching ratio matrix. *Quantitative Finance*, 18(2):199–212.

[Aït-Sahalia et al., 2015] Aït-Sahalia, Y., Cacho-Diaz, J., and Laeven, R. (2015). Modeling financial contagion using mutually exciting jump processes. *Journal of Financial Economics*, 117:585–606.

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

[Aït-Sahalia et al., 2014] Aït-Sahalia, Y., Laeven, R., and Pelizzon, L. (2014). Mutual excitation in eurozone sovereign CDS. *Journal of Econometrics*, 183:151–167.

[Bacry et al., 2017] Bacry, E., Bompaire, M., Deegan, P., Gaïffas, S., and Poulsen, S. V. (2017). tick: A python library for statistical learning, with an emphasis on Hawkes processes and time-dependent models. *The Journal of Machine Learning Research*, 18(1):7937–7941.

[Bacry et al., 2013] Bacry, E., Delattre, S., Hoffmann, M., and Muzy, J.-F. (2013). Some limit theorems for Hawkes processes and application to financial statistics. *Stochastic Processes and their Applications*, 123(7):2475–2499.

[Bacry and Muzy, 2016] Bacry, E. and Muzy, J.-F. (2016). First-and second-order statistics characterization of Hawkes processes and non-parametric estimation. *IEEE Transactions on Information Theory*, 62(4):2184–2202.

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

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

[Brémaud and Massoulié, 1996] Brémaud, P. and Massoulié, L. (1996). Stability of nonlinear Hawkes processes. *Annals of Probability*, pages 1563–1588.

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

[Cavaliere et al., 2023] Cavaliere, G., Lu, Y., Rahbek, A., and Stærk-Østergaard, J. (2023). Bootstrap inference for Hawkes and general point processes. *Journal of Econometrics*, 235:133–165.

[Chavez-Demoulin et al., 2005] Chavez-Demoulin, V., Davison, A., and McNeil, A. (2005). Estimating value-at-risk: a point process approach. *Quantitative Finance*, 5:227–234.

[Chen and Hall, 2013] 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(4):1006–1024.

[Cheysson and Lang, 2022] Cheysson, F. and Lang, G. (2022). Spectral estimation of Hawkes processes from count data. *Annals of Statistics*, 50(3):1722–1746.

[Christensen and Kolokolov, 2024] Christensen, K. and Kolokolov, A. (2024). An unbounded intensity model for point processes. *Journal of Econometrics*, 244(1):105840.

[Clements et al., 2023] Clements, A., Hurn, A., Lindsay, K., and Volkov, V. (2023). Estimating a non-parametric memory kernel for mutually-exciting point processes. *Journal of Financial Econometrics*, 21(5):1759–1790.

[Clinet and Potiron, 2018] Clinet, S. and Potiron, Y. (2018). Statistical inference for the doubly stochastic self-exciting process. *Bernoulli*, 24(4B):3469–3493.

[Clinet and Yoshida, 2017] 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 et al., 2020] 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 and Vere-Jones, 2003] Daley, D. J. and Vere-Jones, D. (2003). *An introduction to the theory of point processes: Elementary theory and methods*, volume 1. Springer New York, NY, 2nd edition.

[Daley and Vere-Jones, 2008] Daley, D. J. and Vere-Jones, D. (2008). *An introduction to the theory of point processes: General theory and structure*, volume 2. Springer New York, NY, 2nd edition.

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

[Erdemlioglu et al., 2025] Erdemlioglu, D., Potiron, Y., and Volkov, V. (2025). Estimation of time-dependent latency with locally stationary Hawkes processes. *Working paper available at https://www.fbc.keio.ac.jp/$\sim$ potiron/Erdemlioglu2025workingpaper.pdf*.

[Feller, 1951] Feller, W. (1951). Two singular diffusion problems. *Annals of Mathematics*, pages 173–182.

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

[Filimonov and Sornette, 2015] Filimonov, V. and Sornette, D. (2015). Apparent criticality and calibration issues in the Hawkes self-excited point process model: application to high-frequency financial data. *Quantitative Finance*, 15(8):1293–1314.

[Fulop et al., 2015] Fulop, A., Li, J., and Yu, J. (2015). Self-exciting jumps, learning, and asset pricing implications. *The Review of Financial Studies*, 28(3):876–912.

[Hardiman et al., 2013] Hardiman, S. J., Bercot, N., and Bouchaud, J.-P. (2013). Critical reflexivity in financial markets: a Hawkes process analysis. *The European Physical Journal B*, 86(10):1–9.

[Hardiman and Bouchaud, 2014] Hardiman, S. J. and Bouchaud, J.-P. (2014). Branching-ratio approximation for the self-exciting Hawkes process. *Physical Review E*, 90(6):062807.

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

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

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

[Hawkes and Oakes, 1974] Hawkes, A. G. and Oakes, D. (1974). A cluster process representation of a self-exciting process. *Journal of Applied Probability*, 11(3):493–503.

[Ikefuji et al., 2022] Ikefuji, M., Laeven, R., Magnus, J., 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, 1975] Jacod, J. (1975). Multivariate point processes: predictable projection, Radon-Nikodym derivatives, representation of martingales. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 31(3):235–253.

[Jacod and Shiryaev, 2003] Jacod, J. and Shiryaev, A. (2003). *Limit theorems for stochastic processes.* Springer Berlin, Heidelberg, 2nd edition.

[Jaisson and Rosenbaum, 2015] Jaisson, T. and Rosenbaum, M. (2015). Limit theorems for nearly unstable Hawkes processes. *Annals of Applied Probability*, 25(2):600–631.

[Jaisson and Rosenbaum, 2016] Jaisson, T. and Rosenbaum, M. (2016). Rough fractional diffusions as scaling limits of nearly unstable heavy tailed Hawkes processes. *Annals of Applied Probability*, 26(5):2860–2882.

[Kwan, 2023] Kwan, T. (2023). *Asymptotic analysis and ergodicity of the Hawkes process and its extensions.* PhD thesis, UNSW Sydney.

[Kwan et al., 2023] Kwan, T., Chen, F., and Dunsmuir, W. (2023). Alternative asymptotic inference theory for a nonstationary Hawkes process. *Journal of Statistical Planning and Inference*, 227:75–90.

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

[Lewis and Mohler, 2011] Lewis, E. and Mohler, G. (2011). A nonparametric EM algorithm for multiscale Hawkes processes. *Journal of Nonparametric Statistics*, 1(1):1–20.

[Liniger, 2009] Liniger, T. (2009). *Multivariate Hawkes processes.* PhD thesis, ETH Zurich.

[Luo et al., 2024] Luo, Y., Sornette, D., and Lera, S. (2024). Quantification of the self-excited emotion dynamics in online interactions. *SFI Working paper 2024.78.*

[Mammen and Müller, 2023] Mammen, E. and Müller, M. (2023). Nonparametric estimation of locally stationary Hawkes processes. *Bernoulli*, 29(3):2062–2083.

[Marsan and Lengline, 2008] Marsan, D. and Lengline, O. (2008). Extending earthquakes' reach through cascading. *Science*, 319(5866):1076–1079.

[Morariu-Patrichi and Pakkanen, 2022] Morariu-Patrichi, M. and Pakkanen, M. S. (2022). State-dependent Hawkes processes and their application to limit order book modelling. *Quantitative Finance*, 22(3):563–583.

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

[Ogata, 1988] Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83(401):9–27.

[Omi et al., 2017] Omi, T., Hirata, Y., and Aihara, K. (2017). Hawkes process model with a time-dependent background rate and its application to high-frequency financial data. *Physical Review E*, 96(1):012303.

[Ozaki, 1979] Ozaki, T. (1979). Maximum likelihood estimation of Hawkes' self-exciting point processes. *Annals of the Institute of Statistical Mathematics*, 31(1):145–155.

[Potiron, 2025] Potiron, Y. (2025). Inference for Hawkes processes with a general kernel. *Working paper available at https://www.fbc.keio.ac.jp/~ potiron/Potiron2025inferenceworkingpaper.pdf*.

[Potiron et al., 2025] Potiron, Y., Scaillet, O., Volkov, V., and Yu, S. (2025). High-frequency estimation of Itô semimartingale baseline for Hawkes processes. *Working paper available at https://www.fbc.keio.ac.jp/~ potiron/Potiron2025workingpaper.pdf*.

[Potiron and Volkov, 2025] Potiron, Y. and Volkov, V. (2025+). Mutually exciting point processes with latency. *To appear in Journal of the American Statistical Association.*

[Rambaldi et al., 2018] Rambaldi, M., Filimonov, V., and Lillo, F. (2018). Detection of intensity bursts using Hawkes processes: An application to high-frequency financial data. *Physical Review E*, 97(3):032318.

[Rambaldi et al., 2015] Rambaldi, M., Pennesi, P., and Lillo, F. (2015). Modeling foreign exchange market activity around macroeconomic news: Hawkes-process approach. *Physical Review E*, 91(1):012819.

[Roueff and Von Sachs, 2019] Roueff, F. and Von Sachs, R. (2019). Time-frequency analysis of locally stationary Hawkes processes. *Bernoulli*, 25(2):1355–1385.

[Roueff et al., 2016] Roueff, F., von Sachs, R., and Sansonnet, L. (2016). Locally stationary Hawkes processes. *Stochastic Processes and their Applications*, 126(6):1710–1743.

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

[Veen and Schoenberg, 2008] Veen, A. and Schoenberg, F. P. (2008). Estimation of space–time branching process models in seismology using an EM–type algorithm. *Journal of the American Statistical Association*, 103(482):614–624.

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

[Vere-Jones and Ozaki, 1982] 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.

[Yu, 2004] Yu, J. (2004). Empirical characteristic function estimation and its applications. *Econometric Reviews*, 23(2):93–123.

Supplementary materials

This part corresponds to the supplementary materials of "Estimation of branching ratio for Hawkes processes with Ito semimartingale baseline" by Yoann Potiron, Olivier Scaillet, Vladimir Volkov and Seunghyeon Yu submitted to the Journal of American Statistical Association. All the proofs of the theory can be found in Section 8.

8 Proofs

In what follows, the constant C refers to a generic constant, which can differ from line to line. We say that $X = o_{\mathbb{P}}(Y)$ when $\frac{X}{Y\mathbf{1}_{\{Y \neq 0\}}} \xrightarrow{\mathbb{P}} 0$. We denote the Hessian matrix of a function f by H_f. If x is a real number, a vector or a matrix, we define the sum of the absolute values of its components as $|x| = \sum_i |x|_i$.

In the lemma that follows, we show that estimation for asymptotic variance of the branching ratio is consistent. This extends Theorem 4.1 in [Potiron et al., 2025].

Lemma 1. *Under Conditions 1 and 2, we have the consistency of the asymptotic variance estimator, i.e.* $\widehat{AVar} \xrightarrow{\mathbb{P}} AVar$.

Proof of Lemma 1. By Definition (17), we have

$$\widehat{AVar} = \nabla f_2(\widehat{x})^T \widehat{\Sigma} \nabla f_2(\widehat{x}). \tag{31}$$

By definition of f_2, we get

$$\widehat{AVar} = \nabla \left(1 - \sqrt{\frac{3\widehat{\text{Mean}}}{2\Delta_n^2(x_2 - x_3)}}\right)(\widehat{x})^T \widehat{\Sigma} \nabla \left(1 - \sqrt{\frac{3\widehat{\text{Mean}}}{2\Delta_n^2(x_2 - x_3)}}\right)(\widehat{x}). \tag{32}$$

By definition of $\widehat{x}$, we can deduce that

$$\widehat{AVar} = \nabla \left(1 - \sqrt{\frac{3\widehat{\text{Mean}}}{2\Delta_n^2(\widehat{\text{Var}}_1 - \widehat{\text{Var}}_2)}}\right)^T \widehat{\Sigma} \nabla \left(1 - \sqrt{\frac{3\widehat{\text{Mean}}}{2\Delta_n^2(\widehat{\text{Var}}_1 - \widehat{\text{Var}}_2)}}\right). \tag{33}$$

By Condition 2 (c), we obtain

$$\widehat{AVar} = \nabla \left(1 - \sqrt{\frac{3n\widehat{\text{Mean}}}{2c(\widehat{\text{Var}}_1 - \widehat{\text{Var}}_2)}}\right)^T \widehat{\Sigma} \nabla \left(1 - \sqrt{\frac{3n\widehat{\text{Mean}}}{2c(\widehat{\text{Var}}_1 - \widehat{\text{Var}}_2)}}\right) + o_{\mathbb{P}}(1). \tag{34}$$

We have that $(n^{-1}\widehat{\text{Mean}}, n^{-2}\widehat{\text{Var}}_1, n^{-2}\widehat{\text{Var}}_2) - (n^{-1}\text{Mean}, n^{-2}\text{Var}_1, n^{-2}\text{Var}_2) \xrightarrow{\mathbb{P}} (0,0)$ by Theorem 4.1 in [Potiron et al., 2025]. Thus, we can deduce that

$$\widehat{AVar} = \nabla\left(1 - \sqrt{\frac{3n\text{Mean}}{2c(\text{Var}_1 - \text{Var}_2)}}\right)^T \widehat{\Sigma}\nabla\left(1 - \sqrt{\frac{3n\text{Mean}}{2c(\text{Var}_1 - \text{Var}_2)}}\right) + o_{\mathbb{P}}(1). \tag{35}$$

By definitions (14) and (15), we get

$$\widehat{AVar} = \nabla\left(1 - \sqrt{\frac{3n\text{Mean}}{2c(\text{Var}_1 - \text{Var}_2)}}\right)^T \Sigma\nabla\left(1 - \sqrt{\frac{3n\text{Mean}}{2c(\text{Var}_1 - \text{Var}_2)}}\right) + o_{\mathbb{P}}(1). \tag{36}$$

Finally, we have by Definition (16) that

$$Avar = \nabla f_1(x)^T \Sigma \nabla f_1(x). \tag{37}$$

We can prove the lemma with Expressions (36) and (37).

□

We now give the proof of Theorem 1. The proof is based on Theorem 4.1 in [Potiron et al., 2025].

Proof of Theorem 1. By Definition (4), we can deduce that

$$\Delta_n^{-\frac{1}{2}}(\widehat{\text{BR}} - \text{BR}) = \Delta_n^{-\frac{1}{2}}\left(1 - \sqrt{\frac{3\widehat{\text{Mean}}}{2\Delta_n^2\big(\widehat{\text{Var}}_1 - \widehat{\text{Var}}_2\big)}} - BR\right). \tag{38}$$

We introduce

$$h(x_1, x_2, x_3) = 1 - \sqrt{\frac{3x_1}{2\big(x_2 - x_3\big)}}.$$

Then, we can rewrite Equation (38) as

$$\Delta_n^{-\frac{1}{2}}(\widehat{\text{BR}} - \text{BR}) = \Delta_n^{-\frac{1}{2}}\Big(h\big(\Delta_n^2\widehat{\text{Mean}}, \Delta_n^4\widehat{\text{Var}}_1, \Delta_n^4\widehat{\text{Var}}_2\big) - BR\Big). \tag{39}$$

We have by Theorem 4.1 in [Potiron et al., 2025] that the convergence of mean estimation is faster than $\Delta_n^{-\frac{1}{2}}$. Thus, we obtain

$$\Delta_n^{-\frac{1}{2}}\Big(h\big(\Delta_n^2\widehat{\text{Mean}}, \Delta_n^4\widehat{\text{Var}}_1, \Delta_n^4\widehat{\text{Var}}_2\big) - BR\Big) = \Delta_n^{-\frac{1}{2}}\Big(h\big(\Delta_n^2\text{Mean}, \Delta_n^4\widehat{\text{Var}}_1, \Delta_n^4\widehat{\text{Var}}_2\big) - BR\Big) \tag{40}$$
$$+ o_{\mathbb{P}}(1).$$

We have that $h(\Delta_n^2\text{Mean}, \Delta_n^4 x_2, \Delta_n^4 x_3)$ is continuously differentiable twice in a neighborhood of

$$(\Delta_n^2\text{Mean}, \Delta_n^4\text{Var}_1, \Delta_n^4\text{Var}_2).$$

We have that $(\Delta_n^4\widehat{\text{Var}}_1, \Delta_n^4\widehat{\text{Var}}_2) - (\Delta_n^4\text{Var}_1, \Delta_n^4\text{Var}_2) \xrightarrow{\mathbb{P}} (0,0)$ by Theorem 4.1 in [Potiron et al., 2025]. Thus, we can apply a Taylor expansion. We obtain that a.s.

$$\begin{aligned} h\big(\Delta_n^2\text{Mean}, \Delta_n^4\widehat{\text{Var}}_1, \Delta_n^4\widehat{\text{Var}}_2\big) = {} & h\big(\Delta_n^2\text{Mean}, \Delta_n^4\text{Var}_1, \Delta_n^4\text{Var}_2\big) \\ & + \nabla h\big(\Delta_n^2\text{Mean}, \Delta_n^4(\widehat{\text{Var}}_1 - \text{Var}_1), \Delta_n^4(\widehat{\text{Var}}_2 - \text{Var}_2)\big) \\ & + \big(\Delta_n^2\text{Mean}, \Delta_n^4(\widehat{\text{Var}}_1 - \text{Var}_1), \Delta_n^4(\widehat{\text{Var}}_2 - \text{Var}_2)\big)^T H_h(\zeta) \\ & * \big(\Delta_n^2\text{Mean}, \Delta_n^4(\widehat{\text{Var}}_1 - \text{Var}_1), \Delta_n^4(\widehat{\text{Var}}_2 - \text{Var}_2)\big). \end{aligned} \tag{41}$$

Here, we have that the random vector $\zeta = (\Delta_n^2\text{Mean}, \zeta_2, \zeta_3)$ is such that ζ_2 is between $\Delta_n^4\text{Var}_1$ and $\Delta_n^4\widehat{\text{Var}}_1$, and ζ_3 is between $\Delta_n^4\text{Var}_2$ and $\Delta_n^4\widehat{\text{Var}}_2$. With the use of Equations (39), (40) and (41), we get

$$\begin{aligned} \Delta_n^{-\frac{1}{2}}(\widehat{\text{BR}} - \text{BR}) = {} & \Delta_n^{-\frac{1}{2}}\Big(h\big(\Delta_n^2\text{Mean}, \Delta_n^4\text{Var}_1, \Delta_n^4\text{Var}_2\big) \\ & + \nabla h\big(\Delta_n^2\text{Mean}, \Delta_n^4(\widehat{\text{Var}}_1 - \text{Var}_1), \Delta_n^4(\widehat{\text{Var}}_2 - \text{Var}_2)\big) \\ & + \big(\Delta_n^2\text{Mean}, \Delta_n^4(\widehat{\text{Var}}_1 - \text{Var}_1), \Delta_n^4(\widehat{\text{Var}}_2 - \text{Var}_2)\big)^T H_h(\zeta) \\ & * \big(\Delta_n^2\text{Mean}, \Delta_n^4(\widehat{\text{Var}}_1 - \text{Var}_1), \Delta_n^4(\widehat{\text{Var}}_2 - \text{Var}_2)\big) - BR\Big) + o_{\mathbb{P}}(1). \end{aligned} \tag{42}$$

By Definitions (9), (10) and (11), we obtain

$$\begin{aligned} & \Delta_n^{-\frac{1}{2}}\Big|h\big(\Delta_n^2\text{Mean}, \Delta_n^4\text{Var}_1, \Delta_n^4\text{Var}_2\big) - BR\Big| \\ & = \Delta_n^{-\frac{1}{2}}\left|1 - \sqrt{\frac{3n\frac{1}{1-BR}\int_0^T \mu_t\,dt}{2\Delta_n^2 n^2 \frac{1}{(1-BR)^2}\int_0^T \Big(\frac{1}{1-BR}\big(\frac{2}{c} - \frac{1}{2c}\big)\mu_t\Big)\,dt}} - BR\right|. \end{aligned} \tag{43}$$

By an algebraic manipulation, this can be rewritten as

$$\begin{aligned} & \Delta_n^{-\frac{1}{2}}\left|1 - \sqrt{\frac{3n\frac{1}{1-BR}\int_0^T \mu_t\,dt}{2\Delta_n^2 n^2 \frac{1}{(1-BR)^2}\int_0^T \Big(\frac{1}{1-BR}\big(\frac{2}{c} - \frac{1}{2c}\big)\mu_t\Big)\,dt}} - BR\right| \\ & = \Delta_n^{-\frac{1}{2}}\left|1 - (1-BR)\sqrt{\frac{c}{\Delta_n^2 n}} - BR\right|. \end{aligned} \tag{44}$$

By Condition 2 (c), we can deduce that

$$\Delta_n^{-\frac{1}{2}}\left|1-(1-BR)\sqrt{\frac{c}{\Delta_n^2 n}}-BR\right| \xrightarrow{\mathbb{P}} 0. \tag{45}$$

By Expressions (43), (44) and (45), we get

$$\Delta_n^{-\frac{1}{2}}\left|h\big(\Delta_n^2\mathrm{Mean},\Delta_n^4\mathrm{Var}_1,\Delta_n^4\mathrm{Var}_2\big)-BR\right| \xrightarrow{\mathbb{P}} 0. \tag{46}$$

By an application of Theorem 4.1 in [Potiron et al., 2025] with Slutsky's lemma, we obtain

$$\Delta_n^{-\frac{1}{2}}\nabla h\big(\Delta_n^2\mathrm{Mean},\Delta_n^4(\widehat{\mathrm{Var}}_1-\mathrm{Var}_1),\Delta_n^4(\widehat{\mathrm{Var}}_2-\mathrm{Var}_2)\big) \xrightarrow{\mathcal{D}-s} \int_0^T w_t d\widetilde{W}_t. \tag{47}$$

By norm properties, we have

$$\begin{aligned}
&\Big|\Delta_n^{-\frac{1}{2}}\big(\Delta_n^2\mathrm{Mean},\Delta_n^4(\widehat{\mathrm{Var}}_1-\mathrm{Var}_1),\Delta_n^4(\widehat{\mathrm{Var}}_2-\mathrm{Var}_2)\big)^T H_h(\zeta) \\
&\quad *\big(\Delta_n^2\mathrm{Mean},\Delta_n^4(\widehat{\mathrm{Var}}_1-\mathrm{Var}_1),\Delta_n^4(\widehat{\mathrm{Var}}_2-\mathrm{Var}_2)\big)\Big| \\
&\leq \Delta_n^{-\frac{1}{2}}\Big|\big(\Delta_n^2\mathrm{Mean},\Delta_n^4(\widehat{\mathrm{Var}}_1-\mathrm{Var}_1),\Delta_n^4(\widehat{\mathrm{Var}}_2-\mathrm{Var}_2)\big)\Big|^2\Big|H_h(\zeta)\Big|.
\end{aligned} \tag{48}$$

As H_h is continuous around the point $\big(\Delta_n^2\mathrm{Mean},\Delta_n^4\mathrm{Var}_1,\Delta_n^4\mathrm{Var}_2\big)$ and we have that $(\Delta_n^4\widehat{\mathrm{Var}}_1,\Delta_n^4\widehat{\mathrm{Var}}_2)-(\Delta_n^4\mathrm{Var}_1,\Delta_n^4\mathrm{Var}_2)\xrightarrow{\mathbb{P}} 0$ by Theorem 4.1 in [Potiron et al., 2025], we obtain

$$\begin{aligned}
&\Delta_n^{-\frac{1}{2}}\Big|\big(\Delta_n^2\mathrm{Mean},\Delta_n^4(\widehat{\mathrm{Var}}_1-\mathrm{Var}_1),\Delta_n^4(\widehat{\mathrm{Var}}_2-\mathrm{Var}_2)\big)\Big|^2\Big|H_h(\zeta)\Big| \\
&\leq C\Delta_n^{-\frac{1}{2}}\Big|\big(\Delta_n^2\mathrm{Mean},\Delta_n^4(\widehat{\mathrm{Var}}_1-\mathrm{Var}_1),\Delta_n^4(\widehat{\mathrm{Var}}_2-\mathrm{Var}_2)\big)\Big|^2.
\end{aligned} \tag{49}$$

By Theorem 4.1 in [Potiron et al., 2025], we get

$$C\Delta_n^{-\frac{1}{2}}\Big|\big(\Delta_n^2\mathrm{Mean},\Delta_n^4(\widehat{\mathrm{Var}}_1-\mathrm{Var}_1),\Delta_n^4(\widehat{\mathrm{Var}}_2-\mathrm{Var}_2)\big)\Big|^2 \xrightarrow{\mathbb{P}} 0. \tag{50}$$

By Equations (48), (49) and (50), we can deduce that

$$\begin{aligned}
&\Big|\Delta_n^{-\frac{1}{2}}\big(\Delta_n^2\mathrm{Mean},\Delta_n^4(\widehat{\mathrm{Var}}_1-\mathrm{Var}_1),\Delta_n^4(\widehat{\mathrm{Var}}_2-\mathrm{Var}_2)\big)^T H_h(\zeta) \\
&\quad *\big(\Delta_n^2\mathrm{Mean},\Delta_n^4(\widehat{\mathrm{Var}}_1-\mathrm{Var}_1),\Delta_n^4(\widehat{\mathrm{Var}}_2-\mathrm{Var}_2)\big)\Big| \xrightarrow{\mathbb{P}} 0.
\end{aligned} \tag{51}$$

Finally, we can deduce Equation (23) for the theorem by Equations (42),(46), (47) and (51). By Slutsky's lemma with Lemma 1, we obtain Equation (24) for the theorem. □

We now give the proof of Corollary 1. The proof is an application of Theorem 1.

Proof of Corollary 1. By Definition (18), the test statistic is equal to

$$S = \frac{\Delta_n^{-1}(\widehat{\mathrm{BR}} - b)^2}{\widehat{AVar}}. \tag{52}$$

Under the null hypothesis H_0, we obtain by an application of the delta method to Expression (23) from Theorem 1 that

$$\frac{\Delta_n^{-1}(\widehat{\mathrm{BR}} - b)^2}{AVar} \xrightarrow{\mathcal{D}} \chi^2. \tag{53}$$

Here, χ^2 is a chi-squared random variable with one degree of freedom. By Expression (53) with Lemma 1 and Slutsky's lemma, we get

$$\frac{\Delta_n^{-1}(\widehat{\mathrm{BR}} - b)^2}{\widehat{AVar}} \xrightarrow{\mathcal{D}} \chi^2. \tag{54}$$

By Expressions (52) and (54), we can deduce that the test statistic S converges in distribution to a chi-squared random variable with one degree of freedom under the null hypothesis H_0.

Under the alternative hypothesis H_1, we introduce the branching ratio $\widetilde{b}$. We have

$$\frac{\Delta_n^{-1}(\widehat{\mathrm{BR}} - b)^2}{\widehat{AVar}} = \frac{\Delta_n^{-1}(\widehat{\mathrm{BR}} - \widetilde{b} + \widetilde{b} - b)^2}{\widehat{AVar}}. \tag{55}$$

An algebraic manipulation yields

$$\frac{\Delta_n^{-1}(\widehat{\mathrm{BR}} - \widetilde{b} + \widetilde{b} - b)^2}{\widehat{AVar}} = \frac{\Delta_n^{-1}((\widehat{\mathrm{BR}} - \widetilde{b})^2 + (\widetilde{b} - b)^2 + 2(\widehat{\mathrm{BR}} - \widetilde{b})(\widetilde{b} - b))}{\widehat{AVar}}. \tag{56}$$

By Expression (54), we get

$$\frac{\Delta_n^{-1}(\widehat{\mathrm{BR}} - \widetilde{b})^2}{\widehat{AVar}} \xrightarrow{\mathcal{D}} \chi^2. \tag{57}$$

We can deduce by Lemma 1 that

$$\frac{\Delta_n^{-1}((\widetilde{b} - b)^2}{\widehat{AVar}} \xrightarrow{\mathbb{P}} \infty. \tag{58}$$

Finally, we obtain by Theorem 1 that

$$\frac{\Delta_n^{-1}(2(\widehat{\mathrm{BR}} - \widetilde{b})(\widetilde{b} - b))}{\widehat{AVar}} \xrightarrow{\mathbb{P}} \infty. \tag{59}$$

By Expressions (57), (58) and (59), we can deduce

$$\frac{\Delta_n^{-1}(\widehat{\mathrm{BR}} - \widetilde{b} + \widetilde{b} - b)^2}{\widehat{AVar}} \xrightarrow{\mathbb{P}} \infty. \tag{60}$$

By Expressions (52),(55) and (60), we can deduce

$$S \xrightarrow{\mathbb{P}} \infty. \tag{61}$$

Thus, the test statistic S is consistent under the alternative hypothesis H_1. $\square$

In the lemma that follows, we show that estimation of asymptotic variance for the Wald test statistic where the branching ratio value does not depend on n is consistent. The lemma extends Theorem 4.1 in [Potiron et al., 2025]. Its proof is based on Theorem 4.1 in [Potiron et al., 2025].

Lemma 2. *Under Conditions 1 and 2, we have the consistency of the asymptotic variance estimator, i.e.* $\widehat{AVar} \xrightarrow{\mathbb{P}} \overline{AVar}$.

Proof of Lemma 2. By Definition (20), we have

$$\widehat{AVar} = \nabla f_4(\widehat{x})^T \widehat{\Sigma} \nabla f_4(\widehat{x}). \tag{62}$$

By definition of f_4, we get

$$\widehat{AVar} = \nabla \left(\frac{2c(x_2 - x_3)}{3n\widehat{\mathrm{Mean}}} \right)(\widehat{x})^T \widehat{\Sigma} \nabla \left(\frac{2c(x_2 - x_3)}{3n\widehat{\mathrm{Mean}}} \right)(\widehat{x}). \tag{63}$$

By definition of $\widehat{x}$, we can deduce that

$$\widehat{AVar} = \nabla \left(\frac{2c(\widehat{\mathrm{Var}}_1 - \widehat{\mathrm{Var}}_2)}{3n\widehat{\mathrm{Mean}}} \right)^T \widehat{\Sigma} \nabla \left(\frac{2c(\widehat{\mathrm{Var}}_1 - \widehat{\mathrm{Var}}_2)}{3n\widehat{\mathrm{Mean}}} \right). \tag{64}$$

We have that $(n^{-1}\widehat{\mathrm{Mean}}, n^{-2}\widehat{\mathrm{Var}}_1, n^{-2}\widehat{\mathrm{Var}}_2) - (n^{-1}\mathrm{Mean}, n^{-2}\mathrm{Var}_1, n^{-2}\mathrm{Var}_2) \xrightarrow{\mathbb{P}} (0,0)$ by Theorem 4.1 in [Potiron et al., 2025]. Thus, we can deduce that

$$\widehat{AVar} = \nabla \left(\frac{2c(\mathrm{Var}_1 - \mathrm{Var}_2)}{3n\mathrm{Mean}} \right)^T \widehat{\Sigma} \nabla \left(\frac{2c(\mathrm{Var}_1 - \mathrm{Var}_2)}{3n\mathrm{Mean}} \right) + o_{\mathbb{P}}(1). \tag{65}$$

By definitions (14) and (15), we get

$$\widehat{AVar} = \nabla \left(\frac{2c(\mathrm{Var}_1 - \mathrm{Var}_2)}{3n\mathrm{Mean}} \right)^T \Sigma \nabla \left(\frac{2c(\mathrm{Var}_1 - \mathrm{Var}_2)}{3n\mathrm{Mean}} \right) + o_{\mathbb{P}}(1). \tag{66}$$

Finally, we have by Definition (19) that

$$\overline{Avar} = \nabla f_3(x)^T \Sigma \nabla f_3(x). \tag{67}$$

We can prove the lemma with Expressions (66) and (67). $\square$

We introduce the local asymptotic variance $\overline{w}_t^2 > 0$ which satisfies

$$\overline{w}_t^2 = \nabla f_3(x)^T \begin{pmatrix} \frac{\mu_t}{1-BR} & 0 & 0 \\ 0 & \breve{\sigma}_t^4 + 4\breve{\sigma}_t^2 \breve{\vartheta}_t + 12\breve{\vartheta}_t^2 & \frac{29}{24}\breve{\sigma}_t^4 + \frac{3}{2}\breve{\sigma}_t^2\breve{\vartheta}_t + \frac{3}{2}\breve{\vartheta}_t^2 \\ 0 & \frac{29}{24}\breve{\sigma}_t^4 + \frac{3}{2}\breve{\sigma}_t^2\breve{\vartheta}_t + \frac{3}{2}\breve{\vartheta}_t^2 & 2\breve{\sigma}_t^4 + 2\breve{\sigma}_t^2\breve{\vartheta}_t + \frac{3}{2}\breve{\vartheta}_t^2 \end{pmatrix} \nabla f_3(x). \tag{68}$$

In the lemma that follows, we show the CLT for estimation of the Wald test statistic for a branching ratio value that does not depend on n under the null hypothesis $\overline{H}_0$. The lemma extends Theorem 4.1 in [Potiron et al., 2025]. Its proof is based on Theorem 4.1 in [Potiron et al., 2025].

Lemma 3. *Under Conditions 1 and 2, we have the CLT for estimation of the Wald test statistic for a branching ratio value that does not depend on n under the null hypothesis $\overline{H}_0$. More specifically, there is a standard Brownian extension of $\mathcal{B}$, with the canonical standard Brownian motion $\widetilde{\overline{W}}_t$ such that*

$$\Delta_n^{-\frac{1}{2}} \left(\frac{2\Delta_n^2(\widehat{\mathrm{Var}}_1 - \widehat{\mathrm{Var}}_2)}{3\widehat{\mathrm{Mean}}} - \frac{1}{(1-BR)^2} \right) \xrightarrow{\mathcal{D}-s} \int_0^T \overline{w}_t d\widetilde{\overline{W}}_t. \tag{69}$$

Proof of Lemma 3. We introduce

$$h_2(x_1, x_2, x_3) = \frac{2(x_2 - x_3)}{3x_1}.$$

Then, we have

$$\Delta_n^{-\frac{1}{2}} \left(\frac{2\Delta_n^2(\widehat{\mathrm{Var}}_1 - \widehat{\mathrm{Var}}_2)}{3\widehat{\mathrm{Mean}}} - \frac{1}{(1-BR)^2} \right) = \Delta_n^{-\frac{1}{2}} \Big(h_2\big(\Delta_n^2 \widehat{\mathrm{Mean}}, \Delta_n^4 \widehat{\mathrm{Var}}_1, \Delta_n^4 \widehat{\mathrm{Var}}_2\big) - \frac{1}{(1-BR)^2} \Big). \tag{70}$$

We have by Theorem 4.1 in [Potiron et al., 2025] that the convergence of mean estimation is faster than $\Delta_n^{-\frac{1}{2}}$. Thus, we obtain

$$\begin{aligned} &\Delta_n^{-\frac{1}{2}} \Big(h_2\big(\Delta_n^2 \widehat{\mathrm{Mean}}, \Delta_n^4 \widehat{\mathrm{Var}}_1, \Delta_n^4 \widehat{\mathrm{Var}}_2\big) - \frac{1}{(1-BR)^2} \Big) \\ &= \Delta_n^{-\frac{1}{2}} \Big(h_2\big(\Delta_n^2 \mathrm{Mean}, \Delta_n^4 \widehat{\mathrm{Var}}_1, \Delta_n^4 \widehat{\mathrm{Var}}_2\big) - \frac{1}{(1-BR)^2} \Big) + o_{\mathbb{P}}(1). \end{aligned} \tag{71}$$

We have that $h_2(\Delta_n^2\text{Mean}, \Delta_n^4 x_2, \Delta_n^4 x_3)$ is continuously differentiable twice in a neighborhood of

$$(\Delta_n^2\text{Mean}, \Delta_n^4\text{Var}_1, \Delta_n^4\text{Var}_2).$$

We have that $(\Delta_n^4\widehat{\text{Var}_1}, \Delta_n^4\widehat{\text{Var}_2}) - (\Delta_n^4\text{Var}_1, \Delta_n^4\text{Var}_2) \xrightarrow{\mathbb{P}} (0,0)$ by Theorem 4.1 in [Potiron et al., 2025]. Thus, we can apply a Taylor expansion. We obtain that a.s.

$$\begin{aligned} h_2\big(\Delta_n^2\text{Mean}, \Delta_n^4\widehat{\text{Var}_1}, \Delta_n^4\widehat{\text{Var}_2}\big) &= h_2\big(\Delta_n^2\text{Mean}, \Delta_n^4\text{Var}_1, \Delta_n^4\text{Var}_2\big) \\ &\quad + \nabla h_2\big(\Delta_n^2\text{Mean}, \Delta_n^4(\widehat{\text{Var}_1} - \text{Var}_1), \Delta_n^4(\widehat{\text{Var}_2} - \text{Var}_2)\big) \\ &\quad + \big(\Delta_n^2\text{Mean}, \Delta_n^4(\widehat{\text{Var}_1} - \text{Var}_1), \Delta_n^4(\widehat{\text{Var}_2} - \text{Var}_2)\big)^T H_{h_2}(\zeta_2) \\ &\quad * \big(\Delta_n^2\text{Mean}, \Delta_n^4(\widehat{\text{Var}_1} - \text{Var}_1), \Delta_n^4(\widehat{\text{Var}_2} - \text{Var}_2)\big). \end{aligned} \tag{72}$$

Here, we have that the random vector $\zeta_2 = (\Delta_n^2\text{Mean}, \zeta_{2,2}, \zeta_{2,3})$ is such that $\zeta_{2,2}$ is between $\Delta_n^4\text{Var}_1$ and $\Delta_n^4\widehat{\text{Var}_1}$, and $\zeta_{2,3}$ is between $\Delta_n^4\text{Var}_2$ and $\Delta_n^4\widehat{\text{Var}_2}$. With the use of Equations (70), (71) and (72), we get

$$\begin{aligned} \Delta_n^{-\frac{1}{2}}\left(\frac{2\Delta_n^2(\widehat{\text{Var}_1} - \widehat{\text{Var}_2})}{3\widehat{\text{Mean}}} - \frac{1}{(1-BR)^2}\right) &= \Delta_n^{-\frac{1}{2}}\Big(h_2\big(\Delta_n^2\text{Mean}, \Delta_n^4\text{Var}_1, \Delta_n^4\text{Var}_2\big) \\ &\quad + \nabla h_2\big(\Delta_n^2\text{Mean}, \Delta_n^4(\widehat{\text{Var}_1} - \text{Var}_1), \Delta_n^4(\widehat{\text{Var}_2} - \text{Var}_2)\big) \\ &\quad + \big(\Delta_n^2\text{Mean}, \Delta_n^4(\widehat{\text{Var}_1} - \text{Var}_1), \Delta_n^4(\widehat{\text{Var}_2} - \text{Var}_2)\big)^T H_{h_2}(\zeta_2) \\ &\quad * \big(\Delta_n^2\text{Mean}, \Delta_n^4(\widehat{\text{Var}_1} - \text{Var}_1), \Delta_n^4(\widehat{\text{Var}_2} - \text{Var}_2)\big) \\ &\quad - \frac{1}{(1-BR)^2}\Big) + o_{\mathbb{P}}(1). \end{aligned} \tag{73}$$

By Definitions (9), (10) and (11), we obtain

$$\begin{aligned} &\Delta_n^{-\frac{1}{2}}\left|h_2\big(\Delta_n^2\text{Mean}, \Delta_n^4\text{Var}_1, \Delta_n^4\text{Var}_2\big) - \frac{1}{(1-BR)^2}\right| \\ &= \Delta_n^{-\frac{1}{2}}\left|\frac{2\Delta_n^2 n^2 \frac{1}{(1-BR)^2}\int_0^T \left(\frac{1}{1-BR}\big(\frac{2}{c} - \frac{1}{2c}\big)\mu_t\right)dt}{3n\frac{1}{1-BR}\int_0^T \mu_t\, dt} - \frac{1}{(1-BR)^2}\right|. \end{aligned} \tag{74}$$

By an algebraic manipulation, this can be rewritten as

$$\begin{aligned} &\Delta_n^{-\frac{1}{2}}\left|\frac{2\Delta_n^2 n^2 \frac{1}{(1-BR)^2}\int_0^T \left(\frac{1}{1-BR}\big(\frac{2}{c} - \frac{1}{2c}\big)\mu_t\right)dt}{3n\frac{1}{1-BR}\int_0^T \mu_t\, dt} - \frac{1}{(1-BR)^2}\right| \\ &= \Delta_n^{-\frac{1}{2}}\left|\frac{1}{(1-BR)^2}\frac{\Delta_n^2 n}{c} - \frac{1}{(1-BR)^2}\right|. \end{aligned} \tag{75}$$

By Condition 2 (c), we can deduce that

$$\Delta_n^{-\frac{1}{2}}\left|\frac{1}{(1-BR)^2}\frac{\Delta_n^2 n}{c}-\frac{1}{(1-BR)^2}\right|\xrightarrow{\mathbb{P}}0. \tag{76}$$

By Expressions (74), (75) and (76), we get

$$\Delta_n^{-\frac{1}{2}}\left|h_2\big(\Delta_n^2\mathrm{Mean},\Delta_n^4\mathrm{Var}_1,\Delta_n^4\mathrm{Var}_2\big)-\frac{1}{(1-BR)^2}\right|\xrightarrow{\mathbb{P}}0. \tag{77}$$

By an application of Theorem 4.1 in [Potiron et al., 2025] with Slutsky's lemma, we obtain

$$\Delta_n^{-\frac{1}{2}}\nabla h_2\big(\Delta_n^2\mathrm{Mean},\Delta_n^4(\widehat{\mathrm{Var}}_1-\mathrm{Var}_1),\Delta_n^4(\widehat{\mathrm{Var}}_2-\mathrm{Var}_2)\big)\xrightarrow{\mathcal{D}-s}\int_0^T\overline{w}_t d\widetilde{\overline{W}}_t. \tag{78}$$

By norm properties, we have

$$\begin{aligned}\Big|\Delta_n^{-\frac{1}{2}}&\big(\Delta_n^2\mathrm{Mean},\Delta_n^4(\widehat{\mathrm{Var}}_1-\mathrm{Var}_1),\Delta_n^4(\widehat{\mathrm{Var}}_2-\mathrm{Var}_2)\big)^T H_{h_2}(\zeta_2)\\&*\big(\Delta_n^2\mathrm{Mean},\Delta_n^4(\widehat{\mathrm{Var}}_1-\mathrm{Var}_1),\Delta_n^4(\widehat{\mathrm{Var}}_2-\mathrm{Var}_2)\big)\Big|\\\leq\Delta_n^{-\frac{1}{2}}&\Big|\big(\Delta_n^2\mathrm{Mean},\Delta_n^4(\widehat{\mathrm{Var}}_1-\mathrm{Var}_1),\Delta_n^4(\widehat{\mathrm{Var}}_2-\mathrm{Var}_2)\big)\Big|^2\Big|H_{h_2}(\zeta_2)\Big|.\end{aligned} \tag{79}$$

As H_{h_2} is continuous around the point $\big(\Delta_n^2\mathrm{Mean},\Delta_n^4\mathrm{Var}_1,\Delta_n^4\mathrm{Var}_2\big)$ and we have that $(\Delta_n^4\widehat{\mathrm{Var}}_1,\Delta_n^4\widehat{\mathrm{Var}}_2)-(\Delta_n^4\mathrm{Var}_1,\Delta_n^4\mathrm{Var}_2)\xrightarrow{\mathbb{P}}0$ by Theorem 4.1 in [Potiron et al., 2025], we obtain

$$\begin{aligned}\Delta_n^{-\frac{1}{2}}&\Big|\big(\Delta_n^2\mathrm{Mean},\Delta_n^4(\widehat{\mathrm{Var}}_1-\mathrm{Var}_1),\Delta_n^4(\widehat{\mathrm{Var}}_2-\mathrm{Var}_2)\big)\Big|^2\Big|H_{h_2}(\zeta_2)\Big|\\&\leq C\Delta_n^{-\frac{1}{2}}\Big|\big(\Delta_n^2\mathrm{Mean},\Delta_n^4(\widehat{\mathrm{Var}}_1-\mathrm{Var}_1),\Delta_n^4(\widehat{\mathrm{Var}}_2-\mathrm{Var}_2)\big)\Big|^2.\end{aligned} \tag{80}$$

By Theorem 4.1 in [Potiron et al., 2025], we get

$$C\Delta_n^{-\frac{1}{2}}\Big|\big(\Delta_n^2\mathrm{Mean},\Delta_n^4(\widehat{\mathrm{Var}}_1-\mathrm{Var}_1),\Delta_n^4(\widehat{\mathrm{Var}}_2-\mathrm{Var}_2)\big)\Big|^2\xrightarrow{\mathbb{P}}0. \tag{81}$$

By Equations (79), (80) and (81), we can deduce that

$$\begin{aligned}\Big|\Delta_n^{-\frac{1}{2}}&\big(\Delta_n^2\mathrm{Mean},\Delta_n^4(\widehat{\mathrm{Var}}_1-\mathrm{Var}_1),\Delta_n^4(\widehat{\mathrm{Var}}_2-\mathrm{Var}_2)\big)^T H_{h_2}(\zeta_2)\\&*\big(\Delta_n^2\mathrm{Mean},\Delta_n^4(\widehat{\mathrm{Var}}_1-\mathrm{Var}_1),\Delta_n^4(\widehat{\mathrm{Var}}_2-\mathrm{Var}_2)\big)\Big|\xrightarrow{\mathbb{P}}0.\end{aligned} \tag{82}$$

Finally, we can deduce the theorem by Equations (73),(77), (78) and (82). □

In the proposition that follows, we show the Wald test statistic for a branching ratio value that does not depend on n is consistent under the alternative hypothesis $\overline{H}_1$. This proposition extends Theorem 2.2 in [Jaisson and Rosenbaum, 2015], who considers time-invariant baseline and $T \to \infty$ for inference purposes, to the time-dependent baseline and in-fill asymptotics case. The proof also extends the arguments from the proof of Theorem 2.2 in [Jaisson and Rosenbaum, 2015]. The main novelty in the proofs is to divide the intensity by $1-\widetilde{b}_n$ to deal with a time-dependent baseline and the branching ratio which both explode as the number of observations n increases.

Proposition 1. *We assume that Conditions 1, 2 and 3 hold. Then, the test statistic $\overline{S}$ is consistent under the alternative hypothesis $\overline{H}_1$, i.e., we have $\mathbb{P}(\overline{S} > q(u) \mid \overline{H}_1) \to 1$ for any $0 < u < 1$.*

Proof of Proposition 1. By Definition (21), the test statistic is equal to

$$\overline{S} = \frac{\Delta_n^{-1}}{\widehat{\widehat{AVar}}}\left(\frac{2\Delta_n^2(\widehat{\mathrm{Var}}_1 - \widehat{\mathrm{Var}}_2)}{3\widehat{\mathrm{Mean}}} - \frac{1}{(1-b)^2}\right)^2. \tag{83}$$

Under the alternative hypothesis $\overline{H}_1$, we have

$$\begin{aligned}\frac{\Delta_n^{-1}}{\widehat{\widehat{AVar}}}\left(\frac{2\Delta_n^2(\widehat{\mathrm{Var}}_1 - \widehat{\mathrm{Var}}_2)}{3\widehat{\mathrm{Mean}}} - \frac{1}{(1-b)^2}\right)^2 = \frac{\Delta_n^{-1}}{\widehat{\widehat{AVar}}}\Bigg(&\frac{2\Delta_n^2(\widehat{\mathrm{Var}}_1 - \widehat{\mathrm{Var}}_2)}{3\widehat{\mathrm{Mean}}} - \frac{1}{(1-\widetilde{b}_n)^2} \\ &+\frac{1}{(1-\widetilde{b}_n)^2} - \frac{1}{(1-b)^2}\Bigg)^2.\end{aligned} \tag{84}$$

An algebraic manipulation yields

$$\begin{aligned}&\frac{\Delta_n^{-1}}{\widehat{\widehat{AVar}}}\left(\frac{2\Delta_n^2(\widehat{\mathrm{Var}}_1 - \widehat{\mathrm{Var}}_2)}{3\widehat{\mathrm{Mean}}} - \frac{1}{(1-\widetilde{b}_n)^2} + \frac{1}{(1-\widetilde{b}_n)^2} - \frac{1}{(1-b)^2}\right)^2 \\ &= \frac{\Delta_n^{-1}}{\widehat{\widehat{AVar}}}\left(\frac{2\Delta_n^2(\widehat{\mathrm{Var}}_1 - \widehat{\mathrm{Var}}_2)}{3\widehat{\mathrm{Mean}}} - \frac{1}{(1-\widetilde{b}_n)^2}\right)^2 + \frac{\Delta_n^{-1}}{\widehat{\widehat{AVar}}}\left(\frac{1}{(1-\widetilde{b}_n)^2} - \frac{1}{(1-b)^2}\right)^2 \\ &\quad + \frac{2\Delta_n^{-1}}{\widehat{\widehat{AVar}}}\left(\frac{2\Delta_n^2(\widehat{\mathrm{Var}}_1 - \widehat{\mathrm{Var}}_2)}{3\widehat{\mathrm{Mean}}} - \frac{1}{(1-\widetilde{b}_n)^2}\right)\left(\frac{1}{(1-\widetilde{b}_n)^2} - \frac{1}{(1-b)^2}\right).\end{aligned} \tag{85}$$

By Lemma 2, we get

$$\frac{\Delta_n^{-1}}{\widehat{\widehat{AVar}}}\left(\frac{2\Delta_n^2(\widehat{\mathrm{Var}}_1 - \widehat{\mathrm{Var}}_2)}{3\widehat{\mathrm{Mean}}} - \frac{1}{(1-\widetilde{b}_n)^2}\right)^2 = \frac{\Delta_n^{-1}}{\overline{AVar}}\left(\frac{2\Delta_n^2(\widehat{\mathrm{Var}}_1 - \widehat{\mathrm{Var}}_2)}{3\widehat{\mathrm{Mean}}} - \frac{1}{(1-\widetilde{b}_n)^2}\right)^2 + o_{\mathbb{P}}(1). \tag{86}$$

By the use of an extension of Theorem 2.2 in [Jaisson and Rosenbaum, 2015] with Condition 3, we have

$$\frac{2\Delta_n^2(\widehat{\mathrm{Var}}_1 - \widehat{\mathrm{Var}}_2)}{3\widehat{\mathrm{Mean}}} = o_{\mathbb{P}}\Big(\frac{1}{(1-\widetilde{b}_n)^2}\Big). \tag{87}$$

Then, we can deduce that

$$\frac{2\Delta_n^2(\widehat{\mathrm{Var}}_1 - \widehat{\mathrm{Var}}_2)}{3\widehat{\mathrm{Mean}}} - \frac{1}{(1-\widetilde{b}_n)^2} = o_{\mathbb{P}}\Big(\frac{1}{(1-\widetilde{b}_n)^2}\Big). \tag{88}$$

As $\frac{1}{(1-\widetilde{b}_n)^2} \to \infty$ under the alternative hypothesis $\overline{H}_1$, we obtain

$$\frac{2\Delta_n^2(\widehat{\mathrm{Var}}_1 - \widehat{\mathrm{Var}}_2)}{3\widehat{\mathrm{Mean}}} - \frac{1}{(1-\widetilde{b}_n)^2} \xrightarrow{\mathbb{P}} \infty. \tag{89}$$

By monotonicity of the square function, we get

$$\left(\frac{2\Delta_n^2(\widehat{\mathrm{Var}}_1 - \widehat{\mathrm{Var}}_2)}{3\widehat{\mathrm{Mean}}} - \frac{1}{(1-\widetilde{b}_n)^2}\right)^2 \xrightarrow{\mathbb{P}} \infty. \tag{90}$$

As $\Delta_n^{-1} \to \infty$, we can deduce that

$$\frac{\Delta_n^{-1}}{\overline{AVar}}\left(\frac{2\Delta_n^2(\widehat{\mathrm{Var}}_1 - \widehat{\mathrm{Var}}_2)}{3\widehat{\mathrm{Mean}}} - \frac{1}{(1-\widetilde{b}_n)^2}\right)^2 \xrightarrow{\mathbb{P}} \infty. \tag{91}$$

We get by Expressions (86) and (91) that

$$\frac{\Delta_n^{-1}}{\widehat{\overline{AVar}}}\left(\frac{2\Delta_n^2(\widehat{\mathrm{Var}}_1 - \widehat{\mathrm{Var}}_2)}{3\widehat{\mathrm{Mean}}} - \frac{1}{(1-\widetilde{b}_n)^2}\right)^2 \xrightarrow{\mathbb{P}} \infty. \tag{92}$$

We can deduce by Lemma 2, the fact that $\widetilde{b}_n \to 1$ and $b < 1$ that

$$\frac{\Delta_n^{-1}}{\widehat{\overline{AVar}}}\left(\frac{1}{(1-\widetilde{b}_n)^2} - \frac{1}{(1-b)^2}\right)^2 \xrightarrow{\mathbb{P}} \infty. \tag{93}$$

Finally, we obtain by Expression (89), the fact that $\widetilde{b}_n \to 1$ and $b < 1$, that

$$\frac{2\Delta_n^{-1}}{\widehat{\overline{AVar}}}\left(\frac{2\Delta_n^2(\widehat{\mathrm{Var}}_1 - \widehat{\mathrm{Var}}_2)}{3\widehat{\mathrm{Mean}}} - \frac{1}{(1-\widetilde{b}_n)^2}\right)\left(\frac{1}{(1-\widetilde{b}_n)^2} - \frac{1}{(1-b)^2}\right) \xrightarrow{\mathbb{P}} \infty. \tag{94}$$

By Expressions (85), (92), (93) and (94), we can deduce

$$\frac{\Delta_n^{-1}}{\widehat{\overline{AVar}}}\left(\frac{2\Delta_n^2(\widehat{\mathrm{Var}}_1 - \widehat{\mathrm{Var}}_2)}{3\widehat{\mathrm{Mean}}} - \frac{1}{(1-\widetilde{b}_n)^2} + \frac{1}{(1-\widetilde{b}_n)^2} - \frac{1}{(1-b)^2}\right)^2 \xrightarrow{\mathbb{P}} \infty. \tag{95}$$

By Expressions (83),(84) and (95), we can deduce

$$\overline{S} \xrightarrow{\mathbb{P}} \infty. \tag{96}$$

Thus, the test statistic $\overline{S}$ is consistent under the alternative hypothesis $\overline{H}_1$. $\square$

We now give the proof of Theorem 2. The proof is an application of Lemma 3 and Proposition 1.

Proof of Theorem 2. By Definition (21), the test statistic is equal to

$$\overline{S} = \frac{\Delta_n^{-1}}{\widehat{\overline{AVar}}} \left(\frac{2\Delta_n^2(\widehat{\mathrm{Var}}_1 - \widehat{\mathrm{Var}}_2)}{3\widehat{\mathrm{Mean}}} - \frac{1}{(1-b)^2} \right)^2. \tag{97}$$

Under the null hypothesis $\overline{H}_0$, we obtain by an application of the delta method to Lemma 3 that

$$\frac{\Delta_n^{-1}}{\overline{\overline{AVar}}} \left(\frac{2\Delta_n^2(\widehat{\mathrm{Var}}_1 - \widehat{\mathrm{Var}}_2)}{3\widehat{\mathrm{Mean}}} - \frac{1}{(1-b)^2} \right)^2 \xrightarrow{\mathcal{D}} \chi^2. \tag{98}$$

By Expression (98) with Lemma 2 and Slutsky's lemma, we get

$$\frac{\Delta_n^{-1}}{\widehat{\overline{AVar}}} \left(\frac{2\Delta_n^2(\widehat{\mathrm{Var}}_1 - \widehat{\mathrm{Var}}_2)}{3\widehat{\mathrm{Mean}}} - \frac{1}{(1-b)^2} \right)^2 \xrightarrow{\mathcal{D}} \chi^2. \tag{99}$$

By Expressions (97) and (99), we can deduce that the test statistic $\overline{S}$ converges in distribution to a chi-squared random variable with one degree of freedom under the null hypothesis $\overline{H}_0$. Finally, the test statistic $\overline{S}$ is consistent under the alternative hypothesis $\overline{H}_1$ by Proposition 1. $\square$