Goodness-of-Fit Test for General Counting Processes

Yuhao Li LIYUHAO.ECON@WHU.EDU.CN
Economics and Management School
Wuhan University

Xiaojun Song SXJ@GSM.PKU.EDU.CN
Guanghua School of Management
Peking University

Rui Cui RCUI@FEM.ECNU.EDU.CN
Faculty of Economics and Management
East China Normal University

Draft Time: November, 2022

Abstract

We propose an omnibus goodness-of-fit test for general counting processes. We show our test is consistent against (almost) any deviation, and can detect local alternatives tending to the null at a $\sqrt{n}$ rate. We contribute to the literature in the following aspects. First, the test statistic is constructed based on an empirical process rather than a sequence of transformed event times. Second, we explicitly take the estimation effect into consideration when bootstrapping the critical value. Third, the proposed framework is valid for both the one-observational counting process as well as the n-observational process. Monte Carlo experiments results suggest good size and power properties of our test, and a simple empirical application is also studied.

JEL Classification: C12
Keywords: Counting Process, Goodness-of-Fit, Multiplier Bootstrap

1. Introduction

Let $S_1, S_2, \ldots$ be random variables that represent event times, a counting process $N(s)$:

$$N(s) = \sum_{j=1}^{\infty} \mathbb{I}\{S_j \leq s\}$$

counts the number of S_j that fall below s. A counting process is uniquely characterized by its conditional random intensity $\lambda(t)$ defined as :

$$\lambda(s) = \frac{\mathbb{E}(dN(s)|\mathcal{F}(s-))}{ds}$$

where $\mathcal{F}(s-)$ represents a filtration that contains relevant information up to time s. One might interpret the intensity function as the conditional expectation of the jump size $dN(s)$ at time s given observed history $\mathcal{F}(s-)$. The counting process has been widely used in finance and economics, e.g., Bacry et al. (2015); Bowsher (2007); Chavez-Demoulin et al. (2005); Stabile and Torrisi (2010); Swishchuk et al. (2021), in seismology, see Zhuang et al. (2002) and in criminology, see Mohler et al. (2012).

In this paper, we aim to provide an omnibus goodness-of-fit test for general counting process models. These include models for one-observational process where researchers only have data on one observational unit, and models for multiple-observational process where data from more than one units is available. Conventional goodness-of-fit tests of a counting process rely on the *Random Time Change* theorem, where one uses the cumulative intensity to transform the original event times to another sequence of event times to form a homogeneous Poisson process with unit intensity. The time transformation is given by:

$$\begin{aligned} Z_j(\theta_0) &= \int_{s_{(j-1)}}^{s_j} \lambda(t; \theta_0 \mid \mathcal{F}(t-)) dt \\ &= \Lambda(s_j; \theta_0 | \mathcal{F}(s_j-)) - \Lambda(s_{(j-1)}; \theta_0 | \mathcal{F}(s_{(j-1)}-)) \end{aligned} \tag{1}$$

where $\Lambda(s; \theta_0|\mathcal{F}(s-)) = \int_0^s \lambda(u; \theta_0|\mathcal{F}(u-))du$ is the cumulative intensity function. $\{Z_j\}$ are i.i.d EXP(1) random variables if the model parameters are evaluated at true values: $\theta = \theta_0$, $\theta \in \Theta \subseteq \mathbb{R}^q$. The basic ideas of these goodness-of-fit tests are to check if the estimated rescaled duration fit the EXP(1) distribution, or if the counting process consisted by the rescaled duration fit the standard Poisson process.

Daley and Vere-Jones (2007) summarizes one classical algorithm of such tests. The basic procedures consist of

1. Form the transformed time sequence: $\tilde{S}_i = \sum_{j=1}^{i} Z_j(\hat{\theta}_n)$, where $\hat{\theta}_n$ is a vector of consistent estimators.

2. Plot the cumulative step function $Y(x)$ through the points $(x_i, y_i) = (\tilde{S}_i/\tau, i/N(\tau))$ in the unit square $0 \leq x, y \leq 1$, where τ is a terminal time.

3. Plot confidence lines $y = x \pm C_{1-\alpha}/\sqrt{\tau}$, where with Φ denoting the standard normal distribution, $\Phi(C_p) = p$.

4. Implement an approximate $100(1-\alpha)\%$ test of the hypothesis that the $\{\tilde{S}_i\}$ come from a unit-rate Poisson process by observing whether the empirical process $Y(x)$ falls outside the confidence band drawn in step 3.

At step 4, this procedure uses the maximum deviation from the expected rate curve to check for departures. It is analogous to the Kolmogorov-Smirnov test in this context. There are two sources of approximation. First, it is a large sample test, based on the Brownian motion approximation to the Poisson process. Second, and more importantly, it does not consider for the effect of estimating the parameters from the same data as are used to check the model. The bias resulting from the latter in moderate-sized data sets may be considerable and even more severe when the process has strong time-dependence features that reduce the effective amount of information available in the data, see Schoenberg (2002).

In this study, we build our goodness-of-fit test procedures based on the Doob-Meyer decomposition:

$$N(s) = \Lambda\left(s; \theta_0 | \mathcal{F}(s-)\right) + M(s; \theta_0) \tag{2}$$

where $M(s; \theta_0)$ is a (local) martingale with mean zero:

$$M_0(s; \theta_0) = \mathbb{E} M(s; \theta_0) = 0$$

The Doob-Meyer decomposition implies an empirical process:

$$\sqrt{n} \bar{M}_n(s; \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(N_i(s) - \Lambda_i(s; \theta_0 | \mathcal{F}_i(s-))\right)$$

where the subscript i denotes an observed unit. Our test statistic is based on measuring the distance to zero of this empirical process. Under the null that the model is correctly specified, this empirical process will converge weakly to a centered Gaussian process $B_\Gamma(t)$ with covariance structure given by $\Gamma(t_1, t_2) = \mathbb{E}\left(M_1(t_1; \theta_0) M_1(t_2; \theta_0)\right)$.

For one-observational process models, we notice that a general counting process consists of a summation of single-event counting processes $\{N_j\}_{j \geq 1}$. Let $T_j = S_j - s_{(j-1)}$ be a random variable that represent the waiting time for $j-$th event, here $s_{(j-1)}$ denotes a realization of $S_{(j-1)}$. To avoid confusion, we will use $T(t)$ to denote the waiting time and $S(s)$ to denote the event time. The single-event counting process is defined as:

$$N_j(s - s_{(j-1)}) = \mathbb{I}\{T_j \leq s - s_{(j-1)}\}$$

and let $\lambda_j(t; \theta_0)$ ($\Lambda_j(t; \theta_0)$) denote the corresponding (cumulative) intensity function. A counting process can be re-written as:

$$N(s) = \sum_{j=1}^{\infty} N_j(s - s_{(j-1)}) = \sum_{j=1}^{\infty} \mathbb{I}\{T_j \leq s - s_{(j-1)}\}$$

Suppose the underlying process is strongly mixing (i.e., the waiting times $\{T_j\}_{j \geq 1}$ are also strongly mixing), one can construct a similar empirical process based:

$$\sqrt{n} \bar{M}_n(t; \theta_0) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left(N_j(t) - \Lambda_j(t; \theta_0 | \mathcal{F}(t-))\right)$$

The main challenge here is to write explicitly the estimation effect in $\sqrt{n} \bar{M}_n(t; \hat{\theta}_n)$, where $\hat{\theta}_n$ is any root-n consistent estimator. In this study, we focus on (1) maximum likelihood estimators and (2) minimum distance estimators. Maximum likelihood (ML) based methods are standard in one-observational models, while for multiple-observational models, minimum distance (MD) based methods should be employed as ML methods may be invaild. The MD method is based on a continuum of moment restrictions $\mathbb{E}M_1(t) = 0, \forall t \in (0, \tau]$.

Commonly used test procedures, like Kolmogorov-Smirnov and Cramer-von Mises, are based on proper functionals of the empirical process. In practice, there are three ways of obtaining critical values of a test: (1) through resampling; (2) through transforming the empirical process to an appropriate martingale, and (3) through the orthogonal components in the spectral representation of the underlying empirical process. In this study, we adopt the first approach and use a multiplier bootstrap procedure to obtain critical values.

The paper is structured as follows. Section 2 studies the empirical process in detail. Specifically, for one-observational process models, we show how to construct the empirical process from an intensity function. Construction of an empirical process in the $n-$observation case is straightforward. Section 3 discusses the proposed test and its asymptotic properties. Section 4 proposes the multiplier bootstrap procedure for estimating critical values, and Section 5 presents Monte Carlo studies. In section 6, we illustrate the test with a financial application. Finally Section 7 concludes the whole paper.

2. The Empirical Process

2.1 Empirical Process of One-Observation Process

In general, we could specify a parametric counting process model via its intensity function. For example, in finance, economics and seismology, the Hawkes process (Hawkes, 1971) is well understood and widely applied. Its intensity can be written as:

$$\begin{aligned}\lambda(s \mid \mathcal{F}(s-)) &= \lambda_0 + \int_0^s g(s-z)dN(z) \\ &= \lambda_0 + \sum_{j:s_j<s} g(s-s_j)\end{aligned}$$

One classical specification for the exciting kernel is $g(s) = \alpha \exp(-\mu s)$. In most cases, researchers would study an *one-observation counting process*. For example, in finance, it is bid or ask times for one stock; in seismology, it is earthquake occurrence times in a region; in criminology, it is crime events in one area; and in insurance, it is the ruin of one insurance company. However, this one-observational process can be understood as a summation of multiple single-event counting processes. As explained in the introduction section, single-event processes are related to waiting times (or durations). Furthermore, the intensity function $\lambda_j(t|\mathcal{F}_j(t-))$ of a single-event process $N_j(t)$ is a product:

$$\lambda_j(t|\mathcal{F}_j(t-)) = h_j(t)\mathbb{I}\{T_j > t\}$$

of the hazard rate $h_j(t)$ and a random process $\mathbb{I}\{T_j > t\}$ indicating whether $j-$th event is at risk just before t.

By the construction of $N(s)$, one might argue that its intensity function should be:

$$\lambda(s|\mathcal{F}(s-)) = \sum_{j=1}^{\infty} h_j(s - s_{(j-1)})\mathbb{I}\{S_j > s > s_{(j-1)}\}$$

where $s - s_{(j-1)}$ is the duration of $j-$th event. The following theorem states that this is indeed the correct specification.

Theorem 1 *Given a filtration $\mathcal{F}(s-)$ such that $\sigma(N(u) : u < s) \subseteq \mathcal{F}(s-)$, the cumulative intensity function $\Lambda(s|\mathcal{F}(s-))$ of $N(s)$ is given by:*

$$\begin{aligned}\Lambda(s|\mathcal{F}(s-)) &= \Lambda(s_{(j-1)}|\mathcal{F}(s_{(j-1)}-)) \\ &\quad + \int_{s_{(j-1)}}^{s} h_j(u - s_{(j-1)})\mathbb{I}\{S_j > u > s_{(j-1)}\}du\end{aligned}$$

Proof See Appendix A ■

Thus, the cumulative intensity functions of single-event processes can be derived by:

$$\Lambda_j(t|\mathcal{F}_j(t-)) = \Lambda(s|\mathcal{F}(s-)) - \Lambda(s_{(j-1)}|\mathcal{F}(s_{(j-1)}-))$$

Let $\mathcal{F}_a^b = \mathcal{F}((a, b]) = \sigma(N(s) : a < s \leq b)$ be the $\sigma-$algebra generated the counting process on the interval $(a, b]$. Define

$$\alpha_N(r) = \sup_{s\in\mathbb{R}} \alpha(\mathcal{F}_{-\infty}^{s}, \mathcal{F}_{s+r}^{+\infty})$$

where

$$\alpha(\mathcal{A}, \mathcal{B}) = \sup\{|\Pr(A \cap B) - \Pr(A)\Pr(B)|; A \in \mathcal{A}, B \in \mathcal{B}\}$$

is the Rosenblatt's strong mixing coefficient. A counting process $N(s)$ is said to be strongly mixing if $\alpha_N(r) \to 0$ as $r \to \infty$. Intuitivly, the strong mixing condition conveys that the dependence between past and future events decreases uniformly to zero as the time gap between them increases.

Theorem 2 *Suppose the underlying counting process is strongly mixing, then the empirical process,*

$$\sqrt{n}\bar{M}_n(t; \theta_0) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left(N_j(t) - \Lambda_j(t; \theta_0|\mathcal{F}_{s_{j-1}}^{s-})\right)$$

converges weakly to a centered Gaussian process $B_\Gamma(t)$ with covariance structure described as

$$\Gamma(t_1, t_2) = \mathbb{E}\left(M_1(t_1; \theta_0)M_1(t_2; \theta_0)\right)$$

Proof See Ivanoff (1982). ■

Particularly, the strong mixing properties for a Hawkes process is studied in Cheysson and Lang (2020).

2.2 Empirical Process for n-Observation Processes

In many empirical studies, an individual's decision or activity records can be presented as a counting process. For example, the individual outpatient activities might consist of a counting process, where each point in this process represents a date of an outpatient consumption. In these situations, a researcher observe a dataset of durations $\{t_{ij}\}$, $i = 1, \ldots, n$ and for each individual i, $j = 1, \ldots, n_i$, within a fixed time interval $\mathcal{T} = (0, \tau]$. Let $y_i = \{t_{ij}\}_{j=1,\ldots,n_i}$ be the observations for an individual, we assume that y_i are i.i.d copies of Y in the sense that individuals are independent and ideneticaly distributed according to the finite dimensional distribution $\mathbb{P}^{Y}_{t_1,\ldots,t_k}(A) = \mathbb{P}\{\omega \in \Omega | (y_{t_1}(\omega), \ldots, y_{t_k}(\omega)) \in A\}$ for $k \in \mathbb{N}$.

For an individual i, let $\{S_{ij} = \sum_{k=1}^{j} T_{ik}\}$ be random variables that represent the event times. a counting process is constructed as:

$$N_i(s) = \sum_{j=1}^{\infty} \mathbb{I}\{S_{ij} \leq s\}$$

The corresponding empirical process for this n-observation processes data is

$$\sqrt{n}\bar{M}_n(s; \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (N_i(s) - \Lambda_i(s; \theta_0 | \mathcal{F}_i(s-)))$$

where $\sigma(N_i(z) : z < s) \subseteq \mathcal{F}_i(s-)$ is individual i's filtration up to time s.

A few words on the dataset $\{t_{ij}\}_{i,j}$ are in order. For an individual, both durations t_{ij} and the number of events n_i are random. Although this data structure is similar to that of an unbalanced panel data, the source of this unbalanceness is different. None of the conventional reasons, like rotating, randomly missing data, pooling cross-sectional and time-series data, nonresponsive, censoring or selection bias, see Baltagi and Song (2006); Hsiao (2014) for references, is the source for the randomness of n_i. Rather, it is the underlying data generating process (described by $\Lambda_i(s)$) that makes n_i vary across individuals.

3. Goodness-of-Fit Test of Counting Process

We are interested in testing the null hypothesis

$$H_0 : M_0(s; \theta_0) = 0, \quad s \in \mathcal{T}, a.s., \tag{3}$$

against the alternative:

$$H_1 : P\left(M_0(s;\theta_0)\right) < 1 \tag{4}$$

We propose a martingale-based test statistic

$$T_n = n\int_{s\in\mathcal{T}}\left(\bar{N}_n(s) - \bar{\Lambda}_n(s;\hat{\theta}_n)\right)^2 \bar{N}_n(ds) \tag{5}$$

We will show in this section that tests based on T_n is omnibus and takes into account the estimation effect. However, depending on estimators, the asymptotic behaviors of T_n are different.

3.1 ML Estimators

For a stationary counting process, Rubin (1972) has written its log-likelihood function as:

$$\log L(s_1, s_2, \ldots, s_n;\theta) = -\int_{\mathcal{T}}\lambda(s;\theta)ds + \int_{\mathcal{T}}\log\lambda(s;\theta)dN(s) \tag{6}$$

Note that the Rubin's log-likelihood is defined under the assumption that the occurrence times $s_1, s_2, \ldots, s_n$ are observed from the beginning of the process, i.e. the time zero, and the log-likelihood is given at the time $\tau \geq s_n$. However, in most applications, only $s_1, s_2, \ldots, s_n$ are given and τ is not specified. Thus, we assume $\tau = s_n$ in the rest of the paper. Thus,

$$\log L(s_1, s_2, \ldots, s_n;\theta) = \sum_{j=1}^{n}\log b_j(\theta)$$

where $\log b_j(\theta) = \log\lambda(s_j;\theta) - \int_{s_{j-1}}^{s_j}\lambda(u;\theta)du$

The linear representation of the ML estimator $\tilde{\theta}_n$ is

$$\begin{aligned}\sqrt{n}\left(\tilde{\theta}_n - \theta_0\right) &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(-\frac{1}{n}\sum_{j=1}^{n}\frac{\partial^2}{\partial\theta\partial\theta^\top}\log b_j(\theta)|_{\theta=\theta_0}\right)^{-1}\log\dot{b}_j(\theta_0) + o_p(1)\\ &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n} l(s_j;\theta_0) + o_p(1)\end{aligned} \tag{7}$$

where

- $\log\dot{b}_j(\theta) = \frac{\partial}{\partial\theta}\log b_j(\theta)|_{\theta=\theta}$
- $l_0 = \mathbb{E}l(S_1;\theta_0) = 0$
- $\Omega = \mathbb{E}(l(S_1;\theta_0)l(S_1;\theta_0)^\top)$ exists.

One can then expand the empirical process as:

$$\sqrt{n}\bar{M}_n(t;\tilde{\theta}_n) = \sqrt{n}\bar{M}_n(t;\theta_0) + \dot{M}_n(t;\theta_0)^\top\frac{1}{\sqrt{n}}\sum_{j=1}^{n} l(s_j;\theta_0) + o_p(1)$$

where

$$\dot{M}_n(t;\theta_0) = \frac{\partial}{\partial\theta}\bar{M}_n(t;\theta)\ |_{\theta=\theta_0}$$

3.1.1 Behavior of T_n Under the Null Hypothesis

We first study the behavior of $\sqrt{n}\bar{M}_n(t;\tilde{\theta}_n)$ under the null. By standard functional central limit theorem, we have $\sqrt{n}\bar{M}_n(t;\theta_0) \Rightarrow B_\Gamma(t)$, where B_Γ denotes a centered Gaussian process with covariance structure given by

$$\Gamma(t_1,t_2) = \mathbb{E}(M_1(t_1;\theta_0)M_1(t_2;\theta_0))$$

In addition, by standard central limit theorem, $\frac{1}{\sqrt{n}}\sum_{j=1}^n l(s_j;\theta_0)$ will converge in distribution to a normal distribution with covariance matrix Ω. Thus, we have the following proposition:

Proposition 3 *Under the null hypothesis and regular conditions,*

$$\sqrt{n}\bar{M}_n(t;\tilde{\theta}_n) \Rightarrow B_\Phi(t)$$

where $B_\Phi(t)$ is a centred Gaussian process with covariance structure:

$$\begin{aligned}\Phi(t_1,t_2) = \Gamma(t_1,t_2) &+ \dot{M}_0(t_1;\theta_0)\mathbb{E}\left(M_1(t_1;\theta_0)l(S_1;\theta_0)\right)\\ &+ \dot{M}_0(t_2;\theta_0)\mathbb{E}\left(M_1(t_2;\theta_0)l(S_1;\theta_0)\right)\\ &+ \dot{M}_0(t_1;\theta_0)^\top\Omega\dot{M}_0(t_2;\theta_0)\end{aligned}$$

and

$$\dot{M}_0(t;\theta_0) = \partial M_0(t;\theta)/\partial\theta|_{\theta=\theta_0}$$

is a $k \times 1$ vector.

Corollary 4 *Under the null hypothesis and regular conditions,*

$$T_n \Rightarrow \int_{\mathcal{T}} B_\Phi(u)^2\mathbb{E}\Lambda_1(du;\theta_0)$$

3.1.2 Behavior of T_n Under the Alternative Hypothesis

In this subsection, we consider both fixed alternatives and local alternatives. For fixed alternatives,

$$H_1 : P(M_0(t;\theta) = 0) < 1 \tag{8}$$

some assumptions are needed.

- B1. $\tilde{\theta}_n \overset{a.s}{\to} \theta_1$.
- B2. $M_i(t;\theta), i = 1, 2, \ldots$ is continuous at θ_1 for each s and either $B2.1$ or $B2.2$ holds: $\forall\theta \in \mathcal{N}_1$ neighborhood of θ_1
  - B2.1 $\exists k(\cdot)$ such that $|M_i(s;\theta)| < k(s)$ and $\mathbb{E}k(S) < \infty$.

– B2.2 $\exists k(\cdot)$ such that $|M_i(s;\theta) - M_i(s;\theta_1)| \leq k(s)|\theta - \theta_1|$ and $\mathbb{E}k(S) < \infty$.

Assumptions B1-B2 are similar to Assumption B of Domínguez and Lobato (2015). Under these assumptions and H_1, it is straightforward to show that the proposed test is consistent.

We can then state trivially the following proposition that establish the consistency of the proposed test.

Proposition 5 *Under the alternatives H_1 and assumptions B1-B2, we have*

$$\frac{T_n}{n} \xrightarrow{p} \int_{\mathcal{T}} M_0(u;\theta_1)^2 \mathbb{E}\Lambda_1(du;\theta_1) \neq 0$$

and consequently, as $n \to \infty$,

$$P(T_n > \varpi) \to 1$$

, for all $\varpi \in \mathbb{R}_+$.

Next, we consider local alternatives:

$$H_{1,n} : M_0(t;\theta_0) = \frac{\mathbb{E}g_1(t)}{\sqrt{\tau}} = \frac{g_0(t)}{\sqrt{\tau}}$$

or equivalently

$$H_{1,n} : \mathbb{E}\Lambda_1(t;\theta_0) = \mathbb{E}\Lambda_1^o(t;\theta_0) + \frac{g_0(t)}{\sqrt{\tau}}$$

$$H_{1,n} : \Lambda(s;\theta_0) = \Lambda^o(s_{j-1};\theta_0) + \int_{s_{j-1}}^{s} \lambda^o(u;\theta_0)du + \frac{\sum_{k=1}^{j-1} g_k(t_k)}{\sqrt{\tau}} + \frac{g_j(s - s_{j-1})}{\sqrt{\tau}}$$

where $g_0(t)$ is a nonstochastic differentiable function, and $\Lambda_1^o(t;\theta_0)$ and $\Lambda^o(s;\theta_0)$ are the cumulative intensities of the single-event process and the whole counting process under the null. Furthermore, we impose that

- $\lim_{\tau\to\infty} \frac{\sqrt{n}}{\sqrt{\tau}} = \zeta$

Under $H_{1,n}$ the linear representation of $\sqrt{n}(\tilde{\theta}_n - \theta_0)$ now has the component

$$l^*(s_j;\theta_0) = \left(-\frac{1}{n}\sum_{j=1}^{n} \frac{\partial^2}{\partial\theta\partial\theta^\top} \log b_j^*(\theta)|_{\theta=\theta_0} \right)^{-1} \log \dot{b}_j^*(\theta_0)$$

where

$$b_j^*(\theta_0) = \left(\lambda^o(s_j;\theta_0) + \frac{\sum_{k=1}^{j} G_k(t_k)}{\sqrt{\tau}} \right) \exp\left(-\left(\Lambda^o(s_j) - \Lambda^o(s_{j-1}) + \frac{G_j(t_j)}{\sqrt{\tau}} \right) \right)$$

and

$$\log b_j^*(\theta_0) = \log b_j(\theta_0) + q_{j,\delta}(\theta_0)$$

with $G_k(t) = dg_k(t)/dt$, and where

$$\log \dot{b}_j^*(\theta_0) = \frac{\partial}{\partial\theta} \log b_j^*(\theta)|_{\theta=\theta_0}$$

$$q_{j,\delta}(\theta_0) = \frac{\sum_{k=1}^{j} G_k(t_k)}{\delta\left(\lambda^o(s_j;\theta_0) + \frac{\sum_{k=1}^{j} G_k(t_k)}{\sqrt{\tau}}\right) + (1-\delta)\lambda^o(s_k;\theta_0)} - \frac{G_j(t_j)}{\sqrt{\tau}}, \quad \delta \in (0,1)$$

Note that

$$\log \dot{b}_j^*(\theta_0) = \log \dot{b}_j(\theta_0) + \dot{q}_{j,\delta}(\theta_0)$$

where $\dot{q}_{j,\delta}(\theta_0) = \partial q_{j,\delta}(\theta)/\partial\theta|_{\theta=\theta_0}$

Since the estimator $\tilde{\theta}_n$ that maximizes $\sum_{j=1}^{n} \log b_j(\theta)$ also maximizes $\sum_{j=1}^{n} \log b_j^*(\theta)$. We have,

Proposition 6 *Under $H_{1,n}$,*

$$\tilde{\theta}_n \xrightarrow{p} \theta_0$$

Proposition 7 *Under $H_{1,n}$,*

$$\sqrt{n}\left(\tilde{\theta}_n - \theta_0\right) \xrightarrow{d} N((\mathcal{I}^*)^{-1}\dot{q}_{0,\delta}(\theta_0), (\mathcal{I}^*)^{-1})$$

where

$$\dot{q}_{0,\delta}(\theta_0) = \mathbb{E}\dot{q}_{1,\delta}(\theta_0)$$

and

$$\mathcal{I}^* = \mathbb{E}\left(-\frac{\partial^2}{\partial\theta\partial\theta^\top} \log b_j^*(\theta)|_{\theta=\theta_0}\right)$$

Finally, let $V_j(t;\theta) = M_i(t;\theta) - g_j(t)/\sqrt{\tau}$, $\bar{g}_n(t) = \sum_{j=1}^{n} g_j(t)/n$, then under $H_{1,n}$,

$$\sqrt{n}\bar{M}_n(t;\tilde{\theta}_n) = \sqrt{n}\bar{V}_n(t;\theta_0) + \frac{\sqrt{n}}{\sqrt{\tau}}\bar{g}_n(t) + \dot{M}_n(t;\theta_0)^\top \frac{1}{\sqrt{n}} \sum_{j=1}^{n} l^*(s_j;\theta_0) + o_p(1)$$

and $\sqrt{n}\bar{V}_n(t;\theta_0) \Rightarrow B_\Gamma(t)$.

Proposition 8 *Under $H_{1,n}$,*

$$\begin{aligned}\sqrt{n}\bar{M}_n(t;\tilde{\theta}_n) &\Rightarrow B_{\Phi^*}(t) + \dot{M}_0(t;\theta_0)^\top (\mathcal{I}^*)^{-1} \dot{q}_{0,\delta}(\theta_0) + \zeta g_0(t) \\ &= B_{\Phi^*}(t) + C(t)\end{aligned}$$

$$T_n \Rightarrow \int_{\mathcal{T}} (B_{\Phi^*}(u) + C(u))^2 \, \mathbb{E}\Lambda_1(du;\theta_0)$$

where $B_{\Phi^*}(s)$ is a centred Gaussian process with covariance structure:

$$\begin{aligned}\Phi^*(t_1,t_2) = \Gamma(t_1,t_2) &+ \dot{M}_0(t_1;\theta_0)^\top \mathbb{E}\left(V_1(t_1;\theta_0)l^*(S_1;\theta_0)\right)\\ &+ \dot{M}_0(t_2;\theta_0)^\top \mathbb{E}\left(V_1(t_2;\theta_0)l^*(S_1;\theta_0)\right)\\ &+ \dot{M}_0(t_1;\theta_0)^\top \Omega^* \dot{M}_0(t_2;\theta_0)\end{aligned}$$

with $\Omega^* = \mathbb{E}(l^*(S_1;\theta_0)l^*(S_1;\theta_0)^\top)$

Proposition 8 points out that T_n converges to a different limit under $H_{1,n}$ except when $C=0$. In such case, the limit distribution under H_0 and $H_{1,n}$ is the same and T_n cannot detect the deviation from H_0. The case of $C=0$ occurs only when $\dot{M}_0(t;\theta_0)^\top \left(\mathcal{I}^*\right)^{-1} \dot{q}_{0,\delta}(\theta_0)+\zeta g_0(t)=0$.

3.2 MD Estimators

Maximum likelihood methods might be invalid when one is given multiple-observational processes data (See Li (2022) for a detailed discussion). The Doob-Meyer decomposition, on the other hand, suggests a continuum of moment restrictions:

$$M_0(s;\theta_0) = 0, \quad a.s$$

for an unique value $\theta_0 = \Theta$, where $\Theta \in \mathbb{R}^k$ is the parametric space. θ_0 is also the unique value minimizing

$$Q(\theta) = \int_{\mathcal{T}} M_0(s;\theta)^2 \mathbb{E}\Lambda_1(ds)$$

Hence, $\theta_0 = \arg\min_\theta Q(\theta)$ is a minimum distance parameter. For a sample of size n, let the sample analogs of $M_0(s;\theta)$ and $Q(\theta)$ be

$$\bar{M}_n(s;\theta) = \bar{N}_n(s) - \bar{\Lambda}_n(s;\theta)$$

and

$$Q_n(\theta) = \int_{\mathcal{T}} \bar{M}_n(s;\theta)^2 \bar{N}_n(ds)$$

The minimum distance estimator is defined by

$$\hat{\theta}_n = \arg\min_{\theta\in\Theta} Q_n(\theta)$$

Under Assumptions A1-A6 (see Appendix B), Kopperschmidt and Stute (2013); Li (2022) have shown that

$$\hat{\theta}_n \xrightarrow{p} \theta_0$$

and

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \left(\int_{\mathcal{T}} \dot{M}_0(s;\theta_0)\dot{M}_0(s;\theta_0)^\top \mathbb{E}\Lambda_1(ds;\theta_0)\right)^{-1} \int_{\mathcal{T}} \dot{M}_0(s;\theta_0) B_\Gamma(s) \mathbb{E}\Lambda_1(ds;\theta_0)$$

3.2.1 Behavior of T_n Under the Null Hypothesis

In order to derive the asymptotic null distribution of T_n, we consider first the behavior of $\bar{M}_n(s;\hat{\theta}_n)$ under H_0.

Decomposing $\sqrt{n}\bar{M}_n(s;\hat{\theta}_n)$ we have:

$$\sqrt{n}\bar{M}_n(s;\hat{\theta}_n) = \sqrt{n}\bar{M}_n(s;\theta_0) - \sqrt{n}\left(\bar{\Lambda}_n(s;\hat{\theta}_n) - \bar{\Lambda}_n(s;\theta_0)\right)$$

where

$$\begin{aligned}
&\sqrt{n}\left(\bar{\Lambda}_n(s;\hat{\theta}_n) - \bar{\Lambda}_n(s;\theta_0)\right) = -\left(\frac{\partial\bar{\Lambda}_n(s;\theta_0)}{\partial\theta}|_{\theta=\theta_0}\right)^\top \sqrt{n}(\hat{\theta}_n - \theta_0) + o_p(1)\\
&= \dot{M}_n(s;\theta_0)^\top\sqrt{n}(\hat{\theta}_n - \theta_0) + o_p(1)\\
&= \dot{M}_n(s;\theta_0)^\top\left(\int_{\mathcal{T}}\dot{M}_0(s;\theta_0)\dot{M}_0(s;\theta_0)^\top\mathbb{E}N(ds)\right)^{-1}\sqrt{n}\int_{\mathcal{T}}\bar{M}_n(s;\theta_0)\dot{M}_0(s;\theta_0)\mathbb{E}N(ds) + o_p(1)\\
&= \dot{M}_n(s;\theta_0)^\top\left(\frac{1}{n}\dot{M}_n(\theta_0)^\top\dot{M}_n(\theta_0)\right)^{-1}\left(\frac{1}{\sqrt{n}}\dot{M}_n(\theta_0)^\top\bar{M}_n(\theta_0)\right) + o_p(1)
\end{aligned}$$

Let $N_n = n\bar{N}_n(\tau)$ denote the number of the pooled events. $\dot{M}_n(\theta_0)$ is a $N_n \times k$ matrix whose j-th row is the $1 \times k$ vector $\dot{M}_n(s_j;\theta_0)^\top$. Similarly, $\bar{M}_n(\theta_0)$ is a $N_n \times 1$ vector consists of $\bar{M}_n(s_1;\theta_0),\ldots,\bar{M}_n(s_{N_n};\theta_0)$. The third equality comes from Theorems 2 and the linear representation of $\sqrt{n}(\hat{\theta}_n - \theta_0)$. The last equality arises from the law of large number.

Let W_n be an operator such that

$$W_n r(s) = r(s) - \dot{M}_n(s;\theta_0)^\top\left(\frac{1}{N_n}\dot{M}_n(\theta_0)^\top\dot{M}_n(\theta_0)\right)^{-1}\left(\frac{1}{N_n}\sum_{j=1}^{N_n}\dot{M}_n(s_j;\theta_0)^\top r(s_j)\right)$$

Note W_n projects $r(s)$ into the orthogonal space of $\dot{M}_n$ and hence, it is a continuous operator. We could re-write $\sqrt{n}\bar{M}_n(s;\hat{\theta}_n)$ as

$$\sqrt{n}\bar{M}_n(s;\hat{\theta}_n) = W_n\sqrt{n}\bar{M}_n(s;\theta_0)$$

Since, $\sqrt{n}\bar{M}_n(s;\theta_0) \xrightarrow{d} B_\Gamma(s)$, and $W_n \xrightarrow{p} W$, where W is the projection operator defined by

$$Wr(s) = r(s) - \dot{M}_0(s;\theta_0)^\top\left(\int_{\mathcal{T}}\dot{M}_0(s;\theta_0)\dot{M}_0(s;\theta_0)^\top\mathbb{E}\Lambda_1(ds;\theta_0)\right)^{-1}\int_{\mathcal{T}}\dot{M}_0(s;\theta_0)^\top r(s)\mathbb{E}\Lambda(ds;\theta_0)$$

Then

Proposition 9 *Under the null hypothesis and Assumptions A1-A6,*

$$\sqrt{n}\bar{M}_n(s;\hat{\theta}_n) \xrightarrow{d} WB_\Gamma(s) = B_\Psi(s)$$

The corresponding covariance structure is:

$$
\begin{aligned}
\Psi(s_1, s_2) = \Gamma(s_1, s_2) - \dot{M}_0(s_1;\theta_0)^\top \Sigma_{\dot{M}\dot{M}^\top}^{-1} \Sigma_{\dot{M}^\top\Gamma}^\top(s_2) - \Sigma_{\dot{M}^\top\Gamma}(s_1)\Sigma_{\dot{M}\dot{M}^\top}^{-1}\dot{M}_0(s_2;\theta_0) \\
+ \dot{M}_0(s_1;\theta_0)^\top \Sigma_{\dot{M}\dot{M}^\top}^{-1} \Sigma \Sigma_{\dot{M}\dot{M}^\top}^{-1} \dot{M}_0(s_2;\theta_0)
\end{aligned}
$$

where

$$
\begin{aligned}
\Sigma_{\dot{M}\dot{M}^\top} &= \int_{\mathcal{T}} \dot{M}_0(s;\theta_0)\dot{M}_0(s;\theta_0)^\top \mathbb{E}\Lambda_1(ds;\theta_0) \\
\Sigma_{\dot{M}^\top\Gamma}(s) &= \int_{\mathcal{T}} \dot{M}_0(u;\theta_0)^\top \Gamma(u,s)\mathbb{E}\Lambda_1(du;\theta_0) \\
\Sigma &= \int_{\mathcal{T}} \dot{M}_0(u;\theta_0)\Gamma(u,v)\dot{M}_0(v;\theta_0)^\top \mathbb{E}\Lambda_1(du;\theta_0)\mathbb{E}\Lambda_1(dv;\theta_0)
\end{aligned}
$$

Corollary 10 *Under the null hypothesis and Assumptions A1-A6,*

$$
T_n \xrightarrow{d} T = \int_{s\in\mathcal{T}} B_\Psi(s)^2 \mathbb{E}\Lambda_1(ds;\theta_0)
$$

3.2.2 Behavior of T_n Under the Alternative Hypothesis

We next discuss the asymptotic behavior of T_n under different alternative hypotheses. We first study its behavior under any fixed alternatives.

Proposition 11 *Under Assumptions H_1, B1 and B2, and as $n \to \infty$,*

$$
P(T_n > \varpi) \to 1
$$

, for all $\varpi \in \mathbb{R}$.

Next, we consider the behavior under local alternatives. Specifically, we consider this sequence of local alternatives:

$$
H_{1,n} : M_i(s;\theta_0) - \frac{g_i(s)}{\sqrt{n}} = 0, \forall s \in \mathcal{T}, i = 1, 2, \ldots, n
$$

We assume that

- C1: The nonstochastic function $g(t) = \mathbb{E}g_1(s) < \infty, \forall s \in \mathcal{T}$.

This assumption is standard in the literature, e.g., Domínguez and Lobato (2015).

Let $\bar{g}(s) = 1/n\sum_{i=1}^n g_i(s)$, and $\bar{V}_n(s;\theta) = \bar{M}_n(s;\theta) - \bar{g}(s)/\sqrt{n}$, then under $H_{1,n}$, we have

$$
\bar{M}_n(s;\theta) = \bar{M}_n(s;\theta) - \bar{M}_n(s;\theta_0) + \bar{V}_n(s;\theta_0) + \frac{\bar{g}(s)}{\sqrt{n}}
$$

By Assumptions A1-A6 and C1, as well as the Glivenko-Cantelli argument, we have

$$
\bar{M}_n(s;\theta) \xrightarrow{a.s} M(s;\theta) - M(s;\theta_0)
$$

Hence,

$$\int_{s\in\mathcal{T}}\left(\bar{N}_n(s)-\bar{\Lambda}_n(s;\hat{\theta}_n)\right)^2\bar{N}_n(ds)\overset{a.s}{\to}\int_{s\in\mathcal{T}}\left(M(s;\theta)-M(s;\theta_0)\right)^2\mathbb{E}\Lambda(ds;\theta_0) \tag{9}$$

Notice that θ_0 still minimizes the r.h.s of equation 9. Thus

Proposition 12 *Under $H_{1,n}$ and assumptions A1-A3, and C1*

$$\hat{\theta}_n\overset{a.s}{\to}\theta_0$$

We then study the asymptotic normality property of $\hat{\theta}_n$ under $H_{1,n}$. Notice that the local linear asymptotic expansion of $\sqrt{n}(\hat{\theta}_n-\theta_0)$ now becomes:

$$\begin{aligned}\sqrt{n}(\hat{\theta}_n-\theta_0)=&-\left(\int_{t\in\mathcal{T}}\dot{M}_n(s;\hat{\theta}_n)\dot{M}_n(s;\theta_n^*)^\top\bar{N}_n(ds)\right)^{-1}\sqrt{n}\int_{t\in\mathcal{T}}\dot{M}_n(s;\hat{\theta}_n)\bar{M}_n(s;\theta_0)\bar{N}_n(ds)\\ =&-\left(\int_{t\in\mathcal{T}}\dot{M}_n(s;\hat{\theta}_n)\dot{M}_n(s;\theta_n^*)^\top\bar{N}_n(ds)\right)^{-1}\sqrt{n}\int_{t\in\mathcal{T}}\dot{M}_n(s;\hat{\theta}_n)\bar{V}_n(s;\theta_0)\bar{N}_n(ds)\\ &-\left(\int_{t\in\mathcal{T}}\dot{M}_n(s;\hat{\theta}_n)\dot{M}_n(s;\theta_n^*)^\top\bar{N}_n(ds)\right)^{-1}\int_{t\in\mathcal{T}}\dot{M}_n(s;\hat{\theta}_n)\bar{g}(s)\bar{N}_n(ds)\end{aligned}$$

where $\theta_n^*=\gamma\theta_0+(1-\gamma)\hat{\theta}_n, 0<\gamma<1$. Since $\mathbb{E}\bar{V}_n(s;\theta_0)=0$, $\bar{V}_n(s;\theta_0)$ is a martingale difference sequence. We can apply the functional central limit theorem to $\bar{V}_n(s;\theta_0)$, just as we apply it to $\bar{M}_n(s;\theta_0)$, getting:

$$\bar{V}_n\overset{d}{\to}B_\Gamma$$

where the covariance structure Γ is defined as before.

Proposition 13 *Under $H_{1,n}$, Assumptions A1-A6 and C1, we have*

$$\sqrt{n}(\hat{\theta}_n-\theta_0)\overset{d}{\to}N(\varphi,\Omega')$$

where

$$\varphi=-\Sigma_{\dot{M}\dot{M}^\top}^{-1}\int_{s\in\mathcal{T}}\dot{M}_0(s;\theta_0)g(s)\mathbb{E}\Lambda_1(ds;\theta_0)$$

and

$$\Omega'=\Sigma_{\dot{M}\dot{M}^\top}^{-1}\Sigma\Sigma_{\dot{M}\dot{M}^\top}^{-1}$$

Finally, we discuss the asymptotic behavior of T_n under $H_{1,n}$. First, notice that

$$\sqrt{n}\bar{M}_n(s;\hat{\theta}_n)=W_n\sqrt{n}\bar{M}_n(s;\theta_0)$$

where W_n is the empirical projection operator, defined in previous sections. Replacing $\sqrt{n}\bar{M}_n(s;\theta)$ by $\sqrt{n}\bar{V}_n(s;\theta)+g(s)$, we have

$$\sqrt{n}\bar{M}_n(s;\theta)=\sqrt{n}\bar{V}_n(s;\hat{\theta}_n)+g(s)=W_n\sqrt{n}\bar{V}_n(s;\theta_0)+W_ng(s)$$

The first element is the linear approximation to $M_n(t;\theta_0)$ under H_0, the second term introduces a mean in the process $M_n(t;\theta_0)$. Thus, the asymptotic distribution of T_n under $H_{1,n}$ is the same Gaussian process obtained under the null, but centered at the function C where

$$C(s) = Wg(s)$$

Proposition 14 *Under $H_{1,n}$, Assumptions A1-A6, and C1, we have*

$$T_n \stackrel{d}{\to} T = \int_{s\in\mathcal{T}} (B_\Psi(s) + C(s))^2 \mathbb{E}\Lambda(ds;\theta_0)$$

Since $C(s)$ is the projection of $g(s)$ onto the orthogonal space of $\dot{M}(s;\theta_0)$, the case of $C(s) = 0$ corresponds to when $g(s)$ is a linear combination of $\dot{M}(s;\theta_0)$. However, in a general situation, the case $g(s) = \beta^\top \dot{M}(s;\theta_0)$ is unlikely to occur since $\dot{M}(s;\theta_0)$ will depend on θ_0, which is unknown.

3.2.3 Matrix Representation of the Test Statistic

Since the asymptotic distribution of the test statistic T_n is not distribution free, the critical values have to be obtained by bootstrap. In this subsection, we represent T_n in terms of matrices to gain insights on the bootstrap.

Let $\bar{M}_n(\theta)$ is the $N_n \times 1$ vector introduced before. The test statistic now can be written as

$$T_n = \bar{M}_n(\hat{\theta}_n)^\top \bar{M}_n(\hat{\theta}_n)$$

Let $\dot{M}_n(\theta)$ be the $N_n \times k$ matrix defined before, we have

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \sqrt{n}\left(\dot{M}_n(\theta_0)^\top \dot{M}_n(\theta_0)\right)^{-1} \dot{M}_n(\theta_0)^\top \bar{M}_n(\theta_0) + o_p(1) \tag{10}$$

Equation 10 is similar to the linearization of the nonlinear least square estimator.

A straightforward asymptotic expansion of the objective function yields:

$$\bar{M}_n(\theta_0)^\top \bar{M}_n(\theta_0) = \bar{M}_n(\hat{\theta}_n)^\top \bar{M}_n(\hat{\theta}_n) + (\hat{\theta}_n - \theta_0)^\top \left(\dot{M}_n(\hat{\theta}_n)^\top \dot{M}_n(\hat{\theta}_n)\right)(\hat{\theta}_n - \theta_0) + o_p(1) \tag{11}$$

Equation 11 shows that T_n underestimates $\bar{M}_n(\theta_0)^\top \bar{M}_n(\theta_0)$. We could further rewrite the proposed test statistic as:

$$T_n = \bar{M}_n(\theta_0)^\top \bar{M}_n(\theta_0) - \bar{M}_n(\theta_0)^\top \dot{M}_n(\theta_0)\left(\dot{M}_n(\theta_0)^\top \dot{M}_n(\theta_0)\right)^{-1} \dot{M}_n(\theta_0)^\top \bar{M}_n(\theta_0) + o_p(1) \tag{12}$$

Define

$$\mathbb{P}_n(\theta_0) = \dot{M}_n(\theta_0)\left(\dot{M}_n(\theta_0)^\top \dot{M}_n(\theta_0)\right)^{-1} \dot{M}_n(\theta_0)^\top$$

$$\mathbb{W}_n(\theta_0) = \mathbb{I} - \mathbb{P}_n(\theta_0)$$

Then

$$T_n = \bar{M}_n(\theta_0)^\top \mathbb{W}_n(\theta_0)\bar{M}_n(\theta_0) + o_p(1) \tag{13}$$

Equation 13 indicates that T_n can be understood as an overidentification test, analogous to Hansen's J test or Sargan test. Furthermore, Compare to Hansen or Sargan tests, our test can not have an asymptotic standard distribution since the number of overidentifying restrictions $N_n - k$ is not fixed. Furthermore, the fact that we can express T_n in terms of projections is what allows us to propose a simple multiplier bootstrap procedure to estimate the critical value in the next section.

4. Bootstrap Test

From previous sections, it is clear that asymptotic distributions of both $\bar{M}_n(t;\hat{\theta}_n)$ and T_n depend on the DGP. To overcome this nonpivotality, we propose two simple multiplier bootstrap procedures to estimate the asymptotic distribution of T_n. We first introduce a multiplier bootstrap procedure for ML estimators, followed by a procedure for MD estimators.

4.1 Multiplier Bootstrap for ML Estimators

The proposed multiplier bootstrap test procedure is based on rescaling $\sqrt{n}\bar{M}_n(t;\tilde{\theta}_n)$. The test statistic is defined as

$$T_n^* = \bar{M}_n^*(\tilde{\theta}_n)^\top \bar{M}_n^*(\tilde{\theta}_n) \tag{14}$$

where the jth element of the vector $\bar{M}_n^*(\hat{\theta}_n)$ is $\bar{M}_n^*(t_j;\hat{\theta}_n)$, and

$$\bar{M}_n^*(t;\tilde{\theta}_n) = \frac{1}{n}\sum_{j=1}^{n}\Big(N_j(t) - \Lambda_j(t;\tilde{\theta}_n)\Big)\varepsilon_j + \dot{M}_n(t;\tilde{\theta}_n)\frac{1}{n}\sum_{j=1}^{n} l(t_j;\tilde{\theta}_n)\varepsilon_j$$

and here $\{\varepsilon_j\}$ is a sequence of i.i.d random variables with zero mean and unit variance. Furthermore, $\{\varepsilon_j\}$ is independent with the underlying counting process. Common choices of distributions of ε_j include the Standard Normal, Rademacher and Mammen's two-point distribution.

Let $\Rightarrow_*$ denote the weak convergence under the bootstrap law, and by the conditional multiplier central limit theorem, we have

Proposition 15 *With proper assumptions,*

- *Under the null,* $\sqrt{n}\bar{M}_n^*(t;\tilde{\theta}_n) \Rightarrow_* B_\Phi(t)$
- *Under* H_1, $\sqrt{n}\bar{M}_n^*(t;\tilde{\theta}_n) \Rightarrow_* B_\Phi(t)$, *except that the* θ_0 *must be replaced by* θ_1 *in the definition of* Φ, Γ *and* Ω.

- *Under $H_{1,n}$, $\sqrt{n}\bar{M}_n^*(t;\tilde{\theta}_n) \Rightarrow_* B_{\Phi^*}(t)$.*

Corollary 16 *Let $\hat{C}_{1-\alpha}$ be the $(1-\alpha)$-quantile of the empirical distribution of T_n^*, and let $C_{1-\alpha}$ be its limit. We have:*

$$\Pr\{T_n > C_{1-\alpha}\} \to \begin{cases} \alpha \text{ under the null} \\ 1 \text{ under the alternative} \\ \kappa \in (\alpha, 1) \text{ under the sequence of local alternatives} \end{cases}$$

In detail, one would test the null hypothesis as follows

1. Calculate the test statistic T_n.
2. Generate $\{\varepsilon_j\}$ a sequence of i.i.d random variables with zero mean and unit variance. This sequence is also independent of the original sample. Then calculate T_n^* by equation 14.
3. Repeat step 2 B times. This produces a set of B independent values of T_n^* that share the asymptotic distribution of T_n.
4. The proposed test of nominal level α rejects the null if $T_n > \hat{C}_{1-\alpha}$.

4.2 Multiplier Bootstrap for MD Estimators

By Equation 13, we specify the test statistic T_n^{**} as

$$T_n^{**} = \bar{M}_n^{**}(\hat{\theta}_n)^\top \mathbb{W}_n(\hat{\theta}_n)\bar{M}_n^{**}(\hat{\theta}_n) \tag{15}$$

where the jth element of the vector $\bar{M}_n^*(\hat{\theta}_n)$ is $\bar{M}_n^*(s_j;\hat{\theta}_n)$, and

$$\bar{M}_n^*(s;\hat{\theta}_n) = \frac{1}{n}\sum_{j=1}^n \Big(N_j(s) - \Lambda_j(s;\hat{\theta}_n)\Big)\,\varepsilon_j$$

$\{\varepsilon_i\}$ is a sequence of i.i.d random variables with zero mean and unit variance. Furthermore, $\{\varepsilon_i\}$ is independent with the underlying counting process.

Proposition 17 *Under Proper Assumptions and either H_0,H_1 or under $H_{1,n}$, we have*

$$\sqrt{n}\bar{M}_n^{**}(s;\hat{\theta}_n) \Rightarrow B_\Psi(s) \quad a.s$$

where under $H_{1,n}$, the θ_0 must be replaced by θ_1 in the definition of Ψ, $\Sigma_{\dot{M}\dot{M}^\top}$, $\Sigma_{\dot{M}^\top\Gamma}$ and Σ.

Since any projection is a linear continuous operator, projections of $\sqrt{n}\bar{M}_n(s;\theta_0)$ and $\sqrt{n}\bar{M}_n^{**}(s;\hat{\theta}_n)$ also enjoy the same limit distribution.

Corollary 18 *Let $\hat{C}_{1-\alpha}$ be the $(1-\alpha)$-quantile of the empirical distribution of T_n^{**}, and let $C_{1-\alpha}$ be its limit. We have:*

$$\Pr\{T_n > C_{1-\alpha}\} \to \begin{cases} \alpha \text{ under the null} \\ 1 \text{ under the alternative} \\ \kappa \in (\alpha, 1) \text{ under the sequence of local alternatives} \end{cases}$$

Hence, the proposed bootstrap test has an α asymptotic level, it is consistent, and it is able to detect alternatives tending to the null at the $\sqrt{n}$ rate.

The above corollary justifies the estimation of the asymptotical critical value of T_n by those of T_n^{**}. The following procedure is employed to estimate the critical value.

1. Minimize the objective function, obtain consistent estimators and calculate the test statistic T_n.
2. Obtain the estimated martingale process $\bar{M}_n(s;\hat{\theta}_n)$ as well as its first derivatives $\dot{M}_n(s;\hat{\theta}_n)$ and construct the $n \times k$ matrix $\dot{M}_n(\hat{\theta}_n)$.
3. Generate $\{\varepsilon_j\}$ a sequence of i.i.d random variables with zero mean and unit variance. This sequence is also independent of the original sample. Then compute the multiplier bootstrap estimated martingale vector $\bar{M}_n^*(\hat{\theta}_n)$. Calculate T_n^{**} by equation 15.
4. Repeat step 3 B times. This produces a set of B independent values of T_n^{**} that share the asymptotic distribution of T_n.
5. The proposed test of nominal level α rejects the null if $T_n > \hat{C}_{1-\alpha}$.

5. Simulations

This section illustrates the performance of our proposed test. First, we consider a one-observation counting process case where the intensity is the form of the Hawkes process:

$$DGP1.0 \quad \lambda(s|\mathcal{F}(s-)) = \mu + \sum_{i:s_i<t} \exp(\alpha x_i) \left(1 + \frac{s - s_i}{c}\right)^{-1} \tag{16}$$

This model is widely applied in the seismology literature. It aims to describe how past earthquakes and their magnitudes would affect the occurrence times of future earthquakes. In the model, $\{s_i\}$ denotes a sequence of occurrence times and $\{x_i\}$ is a sequence of corresponding magnitudes.

We set $\theta_0 = (\mu, \alpha, c)^\top = (0.02, 0.98, 0.018)^\top$. Instead of fixing the number of events (equivalent to the number of individuals in n-observation case), we fix the time interval for the convenient of simulation. We consider three time intervals: $T = (0, 3000)$,$T = (0, 4000)$

	ML Estimator			MD Estimator		
	10	5	1	10	5	1
$T = 3000$	9.2	4.4	0.7	8.7	3.5	1.0
$T = 4000$	9.7	5.2	1.1	9.1	3.8	0.6
$T = 5000$	9.7	5.2	1.3	8.9	4.3	0.6

Table 1: Size results (percentage). One-Observation Hawkes Process

	DGP1.1			DGP1.2			DGP1.3		
	10	5	1	10	5	1	10	5	1
$T = 3000$	74.1	60.2	32.9	89.8	81.2	59.2	99.8	99.5	98.5
$T = 4000$	84.3	75.9	51.6	96.8	92.1	77.0	99.9	99.9	99.9
$T = 5000$	92.7	86.5	64.7	98.5	97.2	88.0	100.0	100.0	100.0

Table 2: Power results (percentage), ML Estimator, One-Observation Hawkes process

and $T = (0, 5000)$. The number of bootstrap replications mk $B = 1000$, and we conduct 1000 experiments.

Table 1 reports the empirical rejection rate under the null hypothesis for three nominal levels: $0.1, 0.05, 0.01$. Overall, the size distortion is moderate.

Next, we consider a power comparison for the Hawkes process model. We consider two three alternative processes that deviate from the null in different directions. All three alternatives are also belonging to one class of Hawkes processes:

$$\lambda(s|\mathcal{F}(s-)) = \mu + A \sum_{i:s_i<t} \exp(\alpha x_i) \left(1 + \frac{s - s_i}{c}\right)^{-p} \tag{17}$$

with the same specification of $\theta_0 = (\mu, \alpha, c)^\top = (0.02, 0.98, 0.018)^\top$, the three alternatives are different in other two parameters:

- DPG1.1 $A = 5.0, p = 1.5$
- DGP1.2 $A = 9.0, p = 1.9$
- DGP1.3 $A = 9.0, p = 1.5$

Note that the number of occurred events increases with the value of A, but decreases with the value of p. Table 3 reports the percentage power for nominal levels of $0.1, 0.05, 0.01$.

Next, we consider n-observation models. Instead of directly specifying the intensity function, here we utilize Theorem 4 and write down the hazard rates for each duration

	DGP1.1			DGP1.2			DGP1.3		
	10	5	1	10	5	1	10	5	1
$T = 3000$	32.9	21.7	8.9	58.1	44.2	22.6	75.2	62.7	36.9
$T = 4000$	40.0	26.5	10.1	66.8	56.0	31.2	82.5	71.8	45.4
$T = 5000$	44.0	31.4	12.2	76.6	64.4	38.7	88.3	79.8	57.8

Table 3: Power results (percentage), MD Estimator, One-Observation Hawkes process

	Size (Percentage)		
	10	5	1
$n = 100$	11.7	6.9	1.7
$n = 200$	10.5	6.0	1.8
$n = 300$	10.2	4.9	1.8

Table 4: Size results (percentage). n-Observation Process

within one counting process. The null model has the following specification:

$$DGP2.0 : h_{i,d_k}(t) = \alpha t^{\alpha-1} \exp(\beta \sum_{j=1}^{k-1} t_{i,j}) \tag{18}$$

where subscript i denotes an individual (an observation process). This specification extends the static Generalized Accelerated Failure Time model of Ridder (1990) to a dynamic framework. It corresponds to an expression that specifies a duration in terms of a multiplicative structure between covariates and an error term:

$$t_{i,k}^{\alpha} = \exp(-\beta \sum_{j=1}^{k-1} t_{i,j}) u_{i,k}$$

where $\{u_{i,k}\}$ is a sequence of i.i.d $EXP(1)$ random variables.

We set the true parameters as $\theta_0 = (\alpha, \beta)^\top = (0.5, -0.05)^\top$. The sample sizes are $100, 200$ and 300. Like before, the number of bootstrap replications mk $B = 1000$, and we conduct 1000 experiments. Table 4 presents the empirical rejection rates under the null hypothesis for three nominal levels:$0.1, 0.05$ and 0.001. The results suggest that the size distortion is small even at the small sample case, and as the sample size increases, the size distortion disappears.

We consider two alternatives:

$$DGP2.1 : t_k^{\alpha} = (1 + \beta \sum_{j=1}^{k-1} t_{i,j})^{0.75} u_{i,k}$$

	DGP2.1			DGP2.2		
	10	5	1	10	5	1
$n = 100$	83.2	74.7	50.0	100.0	100.0	100.0
$n = 200$	95.0	92.8	88.3	100.0	100.0	100.0
$n = 300$	97.6	96.8	95.5	100.0	100.0	100.0

Table 5: Power results (percentage), n-Observation process

where $\{u_{i,k}\}$ is a sequence of i.i.d $EXP(1)$ random variables.. The parameters are set as $\alpha = 0.5, \beta = 0.75$.

$$DGP2.2 : t_k^{\alpha} = \exp(-\beta \sum_{j=1}^{k-1} t_{i,j}) u_{i,k}$$

where $\{u_{i,k}\}$ is a sequence of i.i.d $U(1.0, 1.5)$ random variables. The parameters are set as $\alpha = 0.5, \beta = -0.05$.

Table 5 presents results. To summarize, all four tables have illustrated the effectiveness of the proposed test. Under the null, our test controls the type I error, and under the alternative, the power of test seems depends on specific DGP.

6. Empirical Application

To illustrate how the proposed testing scheme works in practice, we consider one typical example. This example studies 'extreme occurrences' in US stock market, as measured by empirical quantiles of the Dow Jones Index (DJI), see Cavaliere et al. (2022); Embrechts et al. (2011). We consider the DJI daily returns observed over the period January 1, 1994 to December 31, 2010. The event times corresponding to (negative) extreme returns are given by the trading days where the corresponding daily return is below the 10% empirical quantile. In the end, we have $N(T) = 429$ events during the period $T = 6144$ days. Cavaliere et al. (2022) consider the following intensity specification:

$$\lambda(s; \theta) = \mu + \alpha \sum_{i:s_i<t} \beta \exp(-\beta(s - s_i))$$

Using the minimum distance estimation scheme described in section 3.2, and imposing parameter restrictions $\mu, \beta > 0$ and $0 < \alpha < 1$. With ML estimation, we obtain the following estimators: $\hat{\mu} = 0.013(0.004), \hat{\alpha} = 0.807(0.066), \hat{\beta} = 0.019(0.004)$, where the standard errors are within the parenthese. Using MD estimation method, we obtain the following estimators: $\hat{\mu} = 0.011(0.007), \hat{\alpha} = 0.839(0.102), \hat{\beta} = 0.025(0.008)$. The value of the Cramér-Von Mises statistics in ML and MD are $T_n^{ML} = 0.016$ and $T_n^{MD} = 0.005$, respectively. After bootstraping 1000 times, the bootstrap critical values at 5% level are $c_{95\%}^{ML} = 0.024$ and $c_{95\%}^{MD} = 0.007$, respectively. Hence, we conclude that the model is corrected specified.

7. Conclusion

In this paper, we propose an omnibus test for general counting process specifications. We contribute to the literature in the following three areas. First, our test statistic is based on an empirical process argument. Conventionally, goodness-of-fit test for counting processes are based on the random time change theorem, where the focus is to test whether the transformed duration follows standard exponential distribution. As demonstrated by the Doob-Meyer decomposition result, the proposed test is more intuitive, as the difference between a counting process and its compensator constitutes an empirical process. Second, we explicitly take the estimation effect into consideration. The implementation of the random time change theorem heavily relies on a correct specification. When model parameters are estimated, the estimation effect will affect the transformation in a way that the transformed duration is no longer standard exponential, and its exact distribution is difficult, if not impossible, to obtain. On the other hand, although the estimation effect still exist under the empirical process framework, with mild conditions (e.g., linear expansion of the ML estimator), the limit distribution of the test statistic is trackable. Third, our framework is valid for both the single-observational process and the n-observational process. The former is widely used in high-frequency trading data, while the later can be used to analysis individual-decision making processes.

We analyze limit distributions of the proposed test statistic under both the null and the local alternative hypotheses. A multiplier bootstrap procedure is introduced to obtain critical values. This bootstrap procedure is easy to implement and does not require estimation of the model at each round. Monte Carlo exercises are conducted, and the simulation results show that the proposed test has good size and power properties. Finally, a simple real-life example is studied to demonstrate our test procedure.

References

Bacry, E., I. Mastromatteo, and J.-F. Muzy (2015): "Hawkes processes in finance," *Market Microstructure and Liquidity*, 1, 1550005.

Baltagi, B. H. and S. H. Song (2006): "Unbalanced panel data: A survey," *Statistical Papers*, 47, 493–523.

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

Cavaliere, G., Y. Lu, A. Rahbek, and J. Stærk-Østergaard (2022): "Bootstrap inference for Hawkes and general point processes," *Journal of Econometrics*.

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

Cheysson, F. and G. Lang (2020): "Strong mixing condition for Hawkes processes and application to Whittle estimation from count data," *arXiv preprint arXiv:2003.04314*.

Daley, D. J. and D. Vere-Jones (2007): *An introduction to the theory of point processes: volume II: general theory and structure*, vol. 1,2, Springer Science & Business Media.

Domínguez, M. A. and I. N. Lobato (2015): "A simple omnibus overidentification specification test for time series econometric models," *Econometric Theory*, 31, 891–910.

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

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

Hsiao, C. (2014): *Analysis of panel data*, 54, Cambridge university press.

Ivanoff, G. (1982): "Central limit theorems for point processes," *Stochastic Processes and Their Applications*, 12, 171–186.

Kopperschmidt, K. and W. Stute (2013): "The statistical analysis of self-exciting point processes," *Stast. Sinica*, 23, 1273–1298.

Li, Y. (2022): "Analyzing Dynamic Multiple Spell Durations Using Counting Processes," *Available at SSRN 4117237*.

Mohler, G. O., M. B. Short, P. J. Brantingham, F. P. Schoenberg, and G. E. Tita (2012): "Self-exciting point process modeling of crime," *Journal of the American Statistical Association*.

ÖZTÜRK, Ö. and T. P. Hettmansperger (1997): "Generalised weighted Cramér-von Mises distance estimators," *Biometrika*, 84, 283–294.

Ridder, G. (1990): "The non-parametric identification of generalized accelerated failure-time models," *The Review of Economic Studies*, 57, 167–181.

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

Schoenberg, F. P. (2002): "On rescaled Poisson processes and the Brownian bridge," *Annals of the Institute of Statistical Mathematics*, 54, 445–457.

Stabile, G. and G. L. Torrisi (2010): "Risk processes with non-stationary Hawkes claims arrivals," *Methodology and Computing in Applied Probability*, 12, 415–429.

Swishchuk, A., R. Zagst, and G. Zeller (2021): "Hawkes processes in insurance: Risk model, application to empirical data and optimal investment," *Insurance: Mathematics and Economics*, 101, 107–124.

Zhuang, J., Y. Ogata, and D. Vere-Jones (2002): "Stochastic declustering of space-time earthquake occurrences," *Journal of the American Statistical Association*, 97, 369–380.

A. Proofs

A.1 Theorem 1

The proof of theorem 1 is based on the *Time Scaling Theorem*, also known as the *Random Time Change Theorem*. Here we briefly introduce this building block.

A.1.1 Time Rescaling Theorem

Theorem 19 *Suppose we have a counting process $N_i(t)$ with conditional intensity function $\lambda_i(t \mid \mathcal{F}_i(t-))$ on $(0,T]$ and with occurrence times $0 < S_{i1} < S_{i2}, \ldots, S_{N_i(T)} < T$. Suppose further that the duration distributions are continuous with $f_{D_{ij}|\mathcal{F}_i(S_{i(j-1)})}(d)$ on $(s_{i(j-1)}, T]$, for all $j \geq 1$. If we define*

$$Z_{i1} = \int_0^{S_{i1}} \lambda_i(t \mid \mathcal{F}_i(t-))dt$$

and

$$Z_{ij} = \int_{S_{i(j-1)}}^{S_{ij}} \lambda_i(t \mid \mathcal{F}_i(t-))dt, \quad j = 2, \ldots, N_i(T)$$

then, $Z_{i1}, \ldots, Z_{i(N_i(T))}$ are i.i.d EXP(1) random variables.

The proof of the time rescaling theorem has been extensively studied in point process literature. Here, we only provide a simplified version. A lemma is needed.

Lemma. Suppose D is a continuous random variable having p.d.f $f_D(d)$ and c.d.f $F_D(d)$, and suppose further that $f_D(d) > 0$ on an inteval (A, B) and $f_D(d) = 0$ otherwise. Let $\lambda(d)$ be the associated hazard function of D. If we define a random variable Y by $Y = G(D)$ where

$$G(d) = \int_A^d \lambda(u)du$$

then $Y \sim EXP(1)$.

Proof of the Lemma. Denote the c.d.f of EXP(1) distribution as F_{Exp}. We know that if we define Y by

$$Y = F_{Exp}^{-1}(F_D(D))$$

then $Y \sim EXP(1)$. It remains to show that for $G(d)$, we get

$$G(d) = F_{Exp}^{-1}(F_D(d))$$

.

Notice that the inverse function $F_{Exp}^{-1}$is

$$F_{Exp}^{-1}(w) = -\log(1-w)$$

Also recall that the survival function $1 - F_D(d)$ is expressed as

$$1 - F_D(d) = \exp\left(-\int_{-\infty}^{d} \lambda(u)du\right)$$

hence,

$$F_D(d) = 1 - \exp\left(-\int_{-\infty}^{d} \lambda(u)du\right)$$

Thus,

$$\begin{aligned} F_{Exp}^{-1}(F_D(d)) &= -\log(1 - F_D(d)) \\ &= \int_{A}^{d} \lambda(u)du = G(d) \end{aligned}$$

Proof of the theorem. Note that the transformed duration are

$$Z_{ij} = \int_{s_{i(j-1)}}^{s_{ij}} \lambda_i(t \mid \mathcal{F}_i(t-))dt$$

with $s_{i0} = 0$. Applying the lemma to D_{i1} with $Z_{i1} = G_1(D_{i1})$ where

$$G_1(d) = \int_{0}^{d} \lambda_i(t \mid \mathcal{F}_i(t-))dt$$

we get $Z_{i1} \sim EXP(1)$. Continuing to the next event time and defining $D_{i2} = S_{i2} - S_{i1}$ with $Z_{i2} = G_2(D_{i2})$ where

$$G_2(d) = \int_{s_{i1}}^{d} \lambda_i(t \mid \mathcal{F}_i(t-))dt$$

we get $Z_{i2} \sim EXP(1)$ and, furthermore, this is the same distribution results regardless of the value $Z_{i1} = z_{i1}$. Thus, the conditional density function $f_{Z_{i2}|Z_{i1}}(z_{i2} \mid Z_{i1} = z_{i1})$ does not depend on z_{i1}. Therefore, Z_{i1} is independent of Z_{i1}. Continuing on, we get $Z_{ij} \sim EXP(1)$ independently of all Z_{ik} for $k < j$, for all $j = 1, \ldots, N_i$ and for all possible values $N_i = N_i(T)$.

A.1.2 Proof of Theorem 1

Proof We use Figure 1 to help illustrating the proof. Fix a time t, the value of $\Lambda_i(t)$ is:

$$\Lambda_i(t) = \sum_{j=1}^{5} z_{ij} + \bar{z}_{i6}$$

where

$$z_{ij} = \Lambda_i(s_{ij}) - \Lambda_i(s_{i(j-1)}) = G_{ij}(d_{ij}), \quad j = 1, \ldots, 5, s_{i0} = 0$$

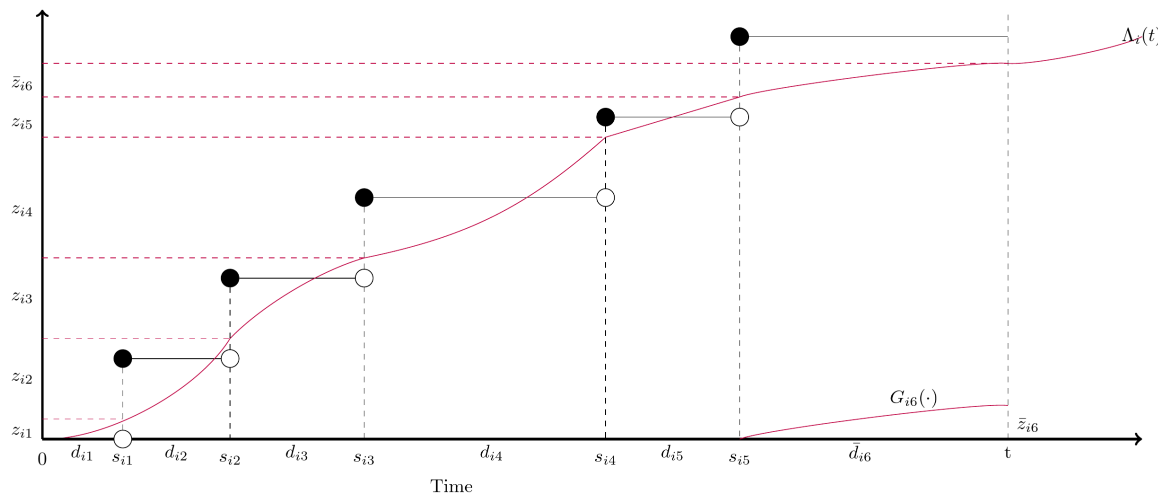

Figure 1: A Possible Realization of a Counting Process and its Cumulative Intensity

and

$$\bar{z}_{i6} = \Lambda_i(t) - \Lambda_i(s_{i5})$$

Here $G_{ij}(\cdot) : \mathbb{R}_+ \to \mathbb{R}_+$ are non-decreasing functions. In the figure, d_{ij} are complete duration, $\bar{d}_{i6}$ is an incomplete duration. Our proof consists of two steps, first, we show that $z_{ij} = H_{ij}(d_{ij})$, where $H_{ij}(\cdot)$ are the corresponding integrated hazard rates of D_{ij}. Second, we show that $\bar{z}_{i6} = H_{i6}(\bar{d}_{i6})$.

From the time rescaling theorem, we known that $z_{ij}, j = 1, \ldots, 5$ are the realizations of an EXP(1) random variable, thus

$$P(Z_{ij} > x) = P(G_{ij}(D_{ij}) > x) = P(D_{ij} > G_{ij}^{-1}(x))$$

and,

$$1 - F_{D_{ij}}(G_{ij}^{-1}(x)) = \exp\left(-H_{D_{ij}}(G_{ij}^{-1}(x))\right) = \exp(-x)$$

In the above equation, the first equality comes from the relationship between the survival function and the integrated hazard rate. Thus, we conclude

$$G_{ij}(\cdot) = H_{D_{ij}}(\cdot)$$

$$z_{ij} = G_{ij}(d_{ij}) = H_{D_{ij}}(d_{ij})$$

Next, notice that

$$P(Z_{i6} > \bar{z}_{i6}) = P(G_{i6}(D_{i6}) > \bar{z}_{i6}) = P(D_{i6} > G_{i6}^{-1}(\bar{z}_{i6}))$$

thus,

$$1 - F_{D_{i6}}(G_{i6}^{-1}(\bar{z}_{i6})) = \exp(-\bar{z}_{i6})$$

Further notice that (as illustrated by the figure)

$$G_{i6}^{-1}(\bar{z}_{i6}) = \bar{d}_{i6}$$

Hence,

$$1 - F_{D_{i6}}(\bar{d}_{i6}) = \exp(-\bar{z}_{i6})$$

and

$$\bar{z}_{i6} = -\log\left(1 - F_{ij}(\bar{d}_{i6})\right) = \int_0^{\bar{d}_{i6}} h_{D_{i6}}(x)dx = H_{D_{i6}}(\bar{d}_{i6})$$

■

B. Minimum Distance Estimation of Counting Process

The continuum of moment restrictions

$$M_0(t;\theta_0) = \mathbb{E}M(t;\theta_0) = \mathbb{E}(N_1(t) - \Lambda_1(t;\theta_0|\mathcal{F}_1(t-))) = 0$$

provides an estimation channel. To derive the estimator, we impose the following assumptions:

- A1. For each $\varepsilon > 0$.

$$\inf_{||\theta-\theta_0||\geq\varepsilon} ||\mathbb{E}\Lambda_1(\cdot;\theta_0) - \mathbb{E}\Lambda_1(\cdot;\theta)||_{\mathbb{E}\Lambda_1(\cdot;\theta_0)} > 0$$

- A2. The process $(t,\theta) \to \Lambda_j(t,\theta), j = 1, 2, \ldots$ is continuous with probability one.
- A3. $\Lambda_j(t;\theta), j = 1, 2, \ldots$ is bounded in t and θ.
- A4. $\Theta \in \mathbb{R}^k$ is compact.

Assumption A1 is a weak identification condition. Assumption A2 suggests that $\Lambda_j(\cdot;\theta), j = 1, 2, \ldots$ has a (random) Lebesgue intensity $\lambda_j(\cdot;\theta)$ with values in an appropriate Skorokhod Space. This guarantees continuity (but not differentiablity) of $\Lambda_j(t;\theta)$ in t and allows for unexpected jumps in the intensity function. Assumption A3 is used in ÖZTÜRK and Hettmansperger (1997), while Assumption A4 is standard in the literature.

By Assumption A1, we have

$$P\left(M_0(t;\theta) = 0\right) < 1, \quad \theta \neq \theta_0$$

thus, $M(t;\theta) \neq 0$ in a non-null space of $\mathcal{T}$, and we have

$$\int_{\mathcal{T}} M_0(t;\theta_0)^2 \mathbb{E}\Lambda_1(dt;\theta_0) = 0$$

but

$$\int_{\mathcal{T}} M_0(t;\theta)^2 \mathbb{E}\Lambda_1(dt;\theta_0) \neq 0 \quad \forall\theta \neq \theta_0$$

Hence,

$$\theta_0 = \arg\min_{\theta\in\Theta} \int_{\mathcal{T}} M_0(t;\theta)^2 \mathbb{E}\Lambda_1(dt;\theta_0) = \arg\min_{\theta\in\Theta} \mathbb{E}\{\langle N_1 - \Lambda_1(\cdot,\theta), N_2 - \Lambda_2(\cdot,\theta)\rangle_{N_3}\}$$

By Lemma 3 of Kopperschmidt and Stute (2013), the above equation can be re-written as:

$$\theta_0 = \arg\min_{\theta\in\Theta} ||\mathbb{E}\Lambda_1(\cdot;\theta_0) - \mathbb{E}\Lambda_1(\cdot;\theta)||^2_{\mathbb{E}\Lambda_1(\cdot;\theta_0)}$$

where

$$||f||_\mu = \Big[\int_{\mathcal{T}} f^2(t)\mu(dt)\Big]^{1/2}$$

is a semi-norm. By Lemma 5 of the same paper, we have

$$||\bar{N}_n - \bar{\Lambda}_n(\cdot,\theta)||^2_{\bar{N}_n} \xrightarrow{p} ||\mathbb{E}\Lambda_1(\cdot;\theta_0) - \mathbb{E}\Lambda_1(\cdot;\theta)||_{\mathbb{E}\Lambda_1(\cdot;\theta_0)}$$

where

$$\bar{N}_n = \frac{1}{n}\sum_{j=1}^n N_j, \quad \bar{\Lambda}_n(\cdot;\theta) = \frac{1}{n}\sum_{j=1}^n \Lambda_j(\cdot;\theta)$$

We can write the minimum distance estimator as

$$\begin{aligned}\hat{\theta}_n = \arg\min_{\theta\in\Theta} ||\bar{N}_n - \bar{\Lambda}_n(\cdot,\theta)||^2_{\bar{N}_n} &= \arg\min_{\theta\in\Theta} \int_{\mathcal{T}} \bar{M}_n(t;\theta)^2 \bar{N}_n(dt) \\ &= \arg\min_{\theta\in\Theta} \frac{1}{n}\sum_{j=1}^n \bar{M}_n(t_j;\theta)^2\end{aligned}$$

where $\bar{M}_n(t;\theta) = \bar{N}_n(t) - \bar{\Lambda}_n(t,\theta)$, and n is the number of jumps (events) in the counting process $N(s), s \in \mathcal{T}$. The quantity $||\bar{N}_n - \bar{\Lambda}_n(\cdot,\theta)||^2_{\bar{N}_n}$ represents an overall measure of fit of $\bar{\Lambda}_n(\cdot,\theta)$ to $\bar{N}_n$. This objective function is a weighted Cramér-von Mises statistic, which can be interpreted as a minimum distance estimator.

Theorem 20 *Under Assumptions A1-A4, we have*

$$\hat{\theta}_n \xrightarrow{a.s} \theta_0$$

Proof See Li (2022) ■

In order to obtain asymptotic normality, some additional assumptions are required.

- A5. $\Lambda_j(t;\cdot), j = 1, 2, \ldots$ is once differentiable in a neighborhood of θ_0 and satisfies $\dot{\Lambda}_j(t;\theta)$ is square integrable w.r.t $\mathbb{E}\Lambda_j(\cdot;\theta_0)$ where $\mathcal{N}_0$ is a neighborhood of θ_0 and $\dot{\Lambda}_j(t;\theta) = \partial\Lambda_j(t;\theta)/\partial\theta$.

- A6. $\theta_0 \in int(\Theta)$.

Assumption A5 is a standard smoothness condition. Assumption A6 is also standard.

Theorem 21 *Under Assumptions A1-A6, we have*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \left(\int_{\mathcal{T}} \dot{M}_0(t;\theta_0)\dot{M}_0(t;\theta_0)^\top \mathbb{E}\Lambda_1(dt;\theta_0)\right)^{-1} \int_{\mathcal{T}} \dot{M}_0(t;\theta_0) B_\Gamma \mathbb{E}\Lambda_1(dt;\theta_0)$$

where $\dot{M}_0(t;\theta_0) = \partial M_0(t;\theta)/\partial\theta|_{\theta=\theta_0}$ *and* B_Γ *denotes a centered Gaussian process with covariance structure given by* $\Gamma(t_1,t_2) = \mathbb{E}(M_1(t_1;\theta_0)M_1(t_2;\theta_0))$.

Proof See Li (2022). ■

This theorem naturally leads to the following corollary.

Corollary 22 *Under Assumptions A1-A8, we have*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0,\Omega)$$

where

$$\begin{aligned}\Omega = &\left(\int_{\mathcal{T}} \dot{M}_0(t;\theta_0)\dot{M}_0(t;\theta_0)^\top \mathbb{E}\Lambda_1(dt;\theta_0)\right)^{-1} \times \\ &\int_{\mathcal{T}}\int_{\mathcal{T}} \dot{M}_0(t_1;\theta_0)\dot{M}_0(t_2;\theta_0)^\top \Gamma(t_1,t_2)\mathbb{E}\Lambda_1(dt_1;\theta_0)\mathbb{E}\Lambda_1(dt_2;\theta_0)\times \\ &\left(\int_{\mathcal{T}} \dot{M}_0(t;\theta_0)\dot{M}_0(t;\theta_0)^\top \mathbb{E}\Lambda_1(dt;\theta_0)\right)^{-1}\end{aligned}$$

Proof This result follows immediately from Theorem 5 and the fact that the integrated weighted Gaussian process follows a normal distribution. ■