J. Appl. Prob. **50,** 1006–1024 (2013)
Printed in England

INFERENCE FOR A NONSTATIONARY SELF-EXCITING POINT PROCESS WITH AN APPLICATION IN ULTRA-HIGH FREQUENCY FINANCIAL DATA MODELING

FENG CHEN,* *The University of New South Wales*
PETER HALL,** *The University of Melbourne*

Abstract

Self-exciting point processes (SEPPs), or Hawkes processes, have found applications in a wide range of fields, such as epidemiology, seismology, neuroscience, engineering, and more recently financial econometrics and social interactions. In the traditional SEPP models, the baseline intensity is assumed to be a constant. This has restricted the application of SEPPs to situations where there is clearly a self-exciting phenomenon, but a constant baseline intensity is inappropriate. In this paper, to model point processes with varying baseline intensity, we introduce SEPP models with time-varying background intensities (SEPPVB, for short). We show that SEPPVB models are competitive with autoregressive conditional SEPP models (Engle and Russell 1998) for modeling ultra-high frequency data. We also develop asymptotic theory for maximum likelihood estimation based inference of parametric SEPP models, including SEPPVB. We illustrate applications to ultra-high frequency financial data analysis, and we compare performance with the autoregressive conditional duration models.

Keywords: Asymptotic normality; consistency; Hawkes process; intensity process; martingale central limit theorem; maximum likelihood estimator; nonstationary; point process; self-exciting; ultra-high frequency

2010 Mathematics Subject Classification: Primary 60G55
Secondary 60F05; 62F12; 62P20

1. Introduction

The self-exciting point process (SEPP) is a useful statistical model for point patterns that have a temporal clustering feature. Since being proposed by Hawkes (1971), the SEPP and its extensions have found applications in a wide ranges of areas, such as earthquake occurrence modeling (Ogata 1988) and prediction (Vere-Jones 1995), neuron firing process modeling (Chornoboy *et al.* 1988), triggered optical emission modeling (Teich and Saleh 2000), financial securities trading and quote arrival time modeling (Engle and Lunde 2003), credit rating transition modeling (Koopman *et al.* 2008), general ultra-high frequency financial data modeling (Monteiro 2009), and social network interaction modeling (Crane and Sornette 2008).

An important problem in applications of point processes is estimation of model parameters. The literature contains results concerning the estimation of SEPPs and associated asymptotic theory. Ogata (1978) established consistency and asymptotic normality of maximum likelihood estimators under stationarity and ergodicity conditions. Chornoboy *et al.* (1988) derived

Received 22 November 2011; revision received 13 December 2012.
* Postal address: School of Mathematics and Statistics, The University of New South Wales, Sydney, NSW 2052, Australia. Email address: feng.chen@unsw.edu.au
** Postal address: Department of Mathematics and Statistics, The University of Melbourne, Melbourne, VIC 3010, Australia.

consistency and asymptotic normality of maximum likelihood estimators for the multivariate SEPP under regularity conditions on the excitation components of the intensity processes. Chornoboy *et al.* (1988) also proposed an expectation-maximization procedure to calculate maximum likelihood estimators, and established the convergence of the procedure. Rathbun (1996) studied asymptotic properties of maximum likelihood estimators (MLEs) for spatial-temporal SEPPs under stationarity conditions.

All these contributions assume that the baseline intensity of SEPP models is constant in time. While there are many situations where a SEPP seems an appropriate model for the event times of interest, a constant baseline event intensity seems unrealistic. For instance, in the modeling of earthquakes the sequence of aftershocks following a major quake may be modeled by a SEPP, but it is well known that the background/baseline aftershock rate decays over time (Utsu 1961). In the modeling of intra-day stock trading, a SEPP appears to be a natural candidate model, but a constant baseline trading intensity would not be able to account for the phenomenon that the trading intensity is almost always much higher during market open and market close (Engle and Russell 1998). In situations such as these, SEPPs with constant background intensities typically fit poorly to the data, and SEPP models with time-varying background intensities (SEPPVB) seem to be more appropriate candidates.

In this paper we show that the SEPPVB models are viable alternatives to the autoregressive conditional duration (ACD) type SEPP models, proposed by Engle and Russell (1998) for modeling ultra-high frequency (UHF) financial data. Due to the stationarity requirement of ACD type models, the data has to be detrended or otherwise 'stationarized' before ACD models can be fitted. The focus of ACD model-based inference is on the residual process, rather than on the original SEPP, and statistical properties of the transformations remain largely unexplored. This makes inference concerning the systematic or signal part of the SEPP difficult. In contrast, SEPPVB models do not require stationarity and therefore can be fitted directly to the original point process data. This, together with the asymptotic theory that we develop, allows simultaneous inference about the signal and noise components of the point process. Moreover, in the real-life, ultra-high frequency stock trading example that we treat, the SEPPVB models fit to the data better than an ACD model with comparable complexity.

Although inference for parametric SEPPVB models, using maximum likelihood, is straightforward to implement, results concerning the properties of MLEs in this setting seem not to be available. Therefore, we provide an asymptotic theory for MLEs of the SEPPVB models, to facilitate statistical inference based on this approach. Due to the time-varying baseline intensity, the SEPP models we shall consider are typically nonstationary, and so the more familiar asymptotic scenario of letting the observation time of the process tend to infinity (Chornoboy *et al.* 1988; Ogata 1978) does not seem appropriate. The asymptotics we shall consider are those where the baseline intensity itself tends to infinity while the observation time interval remains fixed.

The remainder of the paper is as follows. Section 2 presents the model and the MLE. Section 3 studies asymptotic properties of the MLE, and we prove that the MLE is consistent and asymptotically normal. In Section 4 we explain how to assess the goodness of fit of SEPPVB models. In Section 5 we report the results of numerical studies based on simulated and real-life data. All proofs are deferred to Section 6. Section 7 concludes with a discussion.

2. The self-exciting point process with varying baseline intensity

Let $\{t_i, i = 1, 2, \ldots, \}$ be a simple point process on the fixed interval $[0, 1]$. That is, $0 < t_1 < t_2 < \cdots \leq 1$. Let the associated counting process be $N(t) = \max\{i : t_i \leq t\}$, $t \in [0, 1]$. Let

https://doi.org/10.1239/jap/1389370096 Published online by Cambridge University Press

$(\Omega, \mathcal{A}, \mathbb{P})$ be the underlying probability space upon which all the relevant random elements are defined. Denote the natural history of $N(t)$ by $\mathcal{F} = \{\mathcal{F}_t\}_{t\in[0,1]}$, so that $\mathcal{F}_t = \sigma\{N(s); s \le t\}$. The intensity process of $N(t)$, here denoted by $\lambda(t)$, is an $\mathcal{F}$ predictable process such that $M(t) = N(t) - \int_0^t \lambda(u)\,\mathrm{d}u$ is a zero mean locally square integrable martingale with respect to $\mathcal{F}$ under $\mathbb{P}$. The self-exciting intensity point process, as in Hawkes (1971), is a counting process with an intensity process of the following self-exciting form:

$$\lambda(t) = \nu + \int_{[0,t)} g(t-u)\,\mathrm{d}N(u), \tag{1}$$

where $\nu > 0$ is a baseline intensity and the function g measures the influence of an event on the intensity process. It is assumed that $g(x) \ge 0$ for $x \in [0, 1]$, $g(x) = 0$ for $x < 0$, and $\int_0^1 g(t)\,\mathrm{d}t < 1$. The time-varying background intensity SEPP $N(t)$ we shall consider has a time dependent function $\nu(t)$ in place of the constant ν in (1). That is, the intensity process of $N(t)$ is

$$\lambda(t) = \nu(t) + \int_{[0,t)} g(t-u)\,\mathrm{d}N(u).$$

The statistical problem in which we are interested concerns determining the two functions ν and g which determine the distribution of the process N. This paper considers a parametric problem where ν and g are known up to a finite dimensional parameter θ. That is, the intensity of N satisfies

$$\lambda(t;\theta) = \nu(t;\theta) + \int_{[0,t)} g(t-u;\theta)\,\mathrm{d}N(u),$$

where $\theta \in \Theta \subset \mathbb{R}^d$ and d is the dimension of the parameter.

The likelihood function for θ is simply the Radon–Nikodým density of the distribution of the SEPP relative to a suitably chosen measure on the Skorokhod space $\mathcal{D}([0, 1])$ of right continuous functions with left limits, defined on $[0, 1]$, such as the distribution of Poisson process with unit intensity, regarded as a function of θ. Up to a multiplicative constant, with large π denoting product-integration (Andersen *et al.* 1993, Definition II.6.1), the likelihood is

$$\begin{aligned} L(\theta) &= \mathop{\pi}_{t\in[0,1]} \Big\{ \lambda(t;\theta)^{\mathrm{d}N(t)} (1 - \lambda(t;\theta)\,\mathrm{d}t)^{1-\mathrm{d}N(t)} \Big\} \\ &= \prod_{i=1}^{N(1)} \lambda(t_i;\theta) \exp\Big\{ -\int_0^1 \lambda(t;\theta)\,\mathrm{d}t \Big\}. \end{aligned}$$

The loglikelihood, score function, and observed information for the parameter are respectively,

$$\begin{aligned} \ell(\theta) &= \log L(\theta) \\ &= \sum_{i=1}^{N(1)} \log \lambda(t_i;\theta) - \int_0^1 \lambda(t;\theta)\,\mathrm{d}t \\ &= \int_0^1 \log \lambda(t;\theta)\,\mathrm{d}N(t) - \int_0^1 \lambda(t;\theta)\,\mathrm{d}t, \end{aligned}$$

$$S(\theta) = \partial_\theta \ell(\theta) = \int_0^1 \frac{\partial_\theta \lambda(t;\theta)}{\lambda(t;\theta)}\,\mathrm{d}N(t) - \int_0^1 \partial_\theta \lambda(t;\theta)\,\mathrm{d}t = \int_0^1 \frac{\partial_\theta \lambda(t;\theta)}{\lambda(t;\theta)}\,\mathrm{d}M_\theta(t), \tag{2}$$

$$I(\theta) = -\partial_\theta S(\theta)^\top = -\int_0^1 \frac{\partial_\theta^{\otimes 2} \lambda(t;\theta)}{\lambda(t;\theta)}\,\mathrm{d}M_\theta(t) + \int_0^1 \frac{(\partial_\theta \lambda(t;\theta))^{\otimes 2}}{\lambda(t;\theta)^2}\,\mathrm{d}N(t),$$

https://doi.org/10.1239/jap/1389370096 Published online by Cambridge University Press

where

$$M_\theta(t) = N(t) - \int_0^t \lambda(u;\theta)\,\mathrm{d}u.$$

Throughout we use the notation

$$\partial_\theta = \frac{\partial}{\partial\theta} \quad \text{and} \quad \partial_\theta^{\otimes 2} = \frac{\partial^2}{\partial\theta\,\partial\theta^\top}.$$

The MLE $\hat\theta$ can be obtained by maximizing $\ell(\theta)$ over Θ, or by solving the score equation $S(\theta) = 0$.

3. Consistency and asymptotic normality of the MLE

In this section we consider asymptotic properties of the MLE for θ. For this purpose, we assume that there is a sequence of self-exciting intensity SEPP models N_n with their baseline intensity functions proportional to a sequence of positive constants $a_n \nearrow \infty$, that is, the intensity of N_n is

$$\lambda_n(t;\theta) = a_n\mu(t;\theta) + \int_0^t g(t-u;\theta)\,\mathrm{d}N_n(u),$$

where $\mu(t;\theta)$ is a normalized baseline intensity function.

We shall show that under suitable regularity conditions, the MLE $\hat\theta$ behaves like the MLE in a parametric model based on independent and identically distributed (i.i.d.) data, that is, $\hat\theta$ is consistent and asymptotically normally distributed, and the MLE based tests such as the Wald, score, and likelihood ratio tests are all asymptotically χ^2 under the null hypothesis. For ease of reference we list all the regularity conditions here.

(C1) Both the normalized baseline intensity $\mu(\cdot;\theta)$ and the excitation function $g(\cdot;\theta)$ are positive and continuous on [0, 1] for all $\theta \in \Theta$.

(C2) The parameter space Θ is compact and its interior is connected and contains a d-dimensional nonempty open ball which, in turn, contains the true parameter.

(C3) For each $t \in [0, 1]$, the functions $\mu(t;\theta)$ and $g(t;\theta)$ are twice continuously differentiable in θ and their partial derivatives, up to order 2 with respect to θ, are all uniformly equicontinuous when regarded as families of functions of θ.

(C4) For each θ, $\partial_\theta g(t,\theta)$, and $\partial_\theta^{\otimes 2} g(t,\theta)$ are continuously differentiable in t.

(C5) The matrix-valued function

$$\mathcal{I}(\theta) = \int_0^1 \frac{\{\partial_\theta\mu(t;\theta) + \int_0^t \partial_\theta g(t-u;\theta)h(u;\theta)\,\mathrm{d}u\}^{\otimes 2}}{\mu(t;\theta) + \int_0^t g(t-u;\theta)h(u;\theta)\,\mathrm{d}u}\,\mathrm{d}t \tag{3}$$

is nonsingular at the true parameter value, where the function $h(t;\theta)$ is determined by the functions μ and g and is given in (5) below.

Under C1, by the cluster Poisson process representation of SEPP (Hawkes and Oakes 1974) we have $\mathbb{E}\{N_n(1)^2\} < \infty$, and therefore the process

$$M_\theta^n(t) = N_n(t) - \int_0^t \lambda_n(u,\theta)\,\mathrm{d}u, \qquad t \in [0, 1],$$

is a square integrable martingale. Moreover, the mean intensity process of $N_n(t)$, $\eta_n(t;\theta) = \mathbb{E}\{\mathrm{d}N_n(t)\}/\mathrm{d}t = \mathbb{E}\{\lambda_n(t;\theta)\}$, satisfies the linear Volterra integral equation

$$\eta(t) = a_n \mu(t;\theta) + \int_0^t g(t-u;\theta)\eta(u)\,\mathrm{d}u. \tag{4}$$

By the theory of integral equations (e.g. Theorem 3.5 of Linz 1985), a unique solution of (4) on [0, 1] exists and is given by

$$\eta_n(t;\theta) = a_n\left\{\mu(t;\theta) + \int_0^t R(t-u;\theta)\mu(u;\theta)\,\mathrm{d}u\right\} \equiv a_n h(t;\theta), \tag{5}$$

where the *resolvent kernel* $R(\cdot;\theta)$ is the unique solution of the integral equation

$$R(t) = g(t;\theta) + \int_0^t g(t-u;\theta)R(u)\,\mathrm{d}u, \qquad t \in [0, 1]. \tag{6}$$

While general numerical procedures, such as Laplace transformations and inverse transformations, might be required to solve the integral equations (4) and (6), for special excitation functions explicit solutions are available. For instance, if $g(t) = \gamma_1 \mathrm{e}^{-\gamma_2 t}$ then the solution to (6) is easily obtained by differentiation and solving the resulting differential equation; it is $R(t) = \gamma_1 \mathrm{e}^{-(\gamma_2-\gamma_1)t}$.

We now state our main results concerning the properties of the MLE. Hereafter, we use θ_0 to denote the true parameter.

Theorem 1. *Under C1-C5, with probability tending to 1, $\hat{\theta}$ exists as a solution to the score equation $S(\theta) = 0$, and $\hat{\theta} \xrightarrow{\mathbb{P}} \theta_0$.*

The following result states that the MLE is asymptotically normally distributed with a variance-covariance matrix that can be estimated by inverting the observed information matrix.

Theorem 2. *Under C1-C5, we have*

$$a_n^{1/2}(\hat{\theta} - \theta_0) \xrightarrow{\mathrm{D}} N(0, \mathcal{I}(\theta_0)^{-1}), \tag{7}$$

where $\mathcal{I}(\theta)$ is as in (3). Moreover, we have $a_n^{-1} I(\hat{\theta}) \xrightarrow{\mathbb{P}} \mathcal{I}(\theta_0)$.

Proofs of Theorems 1 and 2 are given in Section 6.

4. Assessing adequacy of the model

In practice, after fitting a model to a data set we normally would like to assess whether the model provides an adequate fit. If a point pattern $\{t_i\}$ on the interval [0, 1] does follow a SEPP with a specified conditional intensity $\lambda(t)$, then the integral transformed point pattern $\Lambda(t_i)$ should follow a unit intensity Poisson process on $[0, \Lambda(1)]$, where $\Lambda(t) = \int_0^t \lambda(u)\,\mathrm{d}u$ is the cumulative intensity process, or the compensator, of the SEPP. Given that a Poisson process is monitored from time 0 until a specific time T and since the joint distribution of the ordered event times of the process is equal to that of the order statistics of an equal number of uniformly distributed times on the interval $[0, T]$, we can assess the adequacy of a fitted SEPP by checking the uniformity of the transformed event times $\hat{\Lambda}(t_i)$ on the interval $[0, \hat{\Lambda}(1)]$, where $\hat{\Lambda}$ is obtained by substituting in λ the unknown parameters by their estimates. The uniformity can

https://doi.org/10.1239/jap/1389370096 Published online by Cambridge University Press

be visually checked through a Quantile-Quantile plot (QQ-plot) or through formal tests, such as the χ^2 test or the Kolmogorov-Smirnov (KS) test. In the following real-life examples we shall use the KS test since it avoids a somewhat arbitrary choice of bin boundaries when calculating the χ^2 statistic. Large p-values of the test indicate acceptable model fits. Of course, in interpreting the magnitude of p-values we should bear in mind that the transformation function $\hat{\Lambda}$ carries the randomness of the data, and the distribution of the test statistic would be more spread-out than that calculated from a prespecified Λ, and, therefore, we should be more tolerant of smaller p-values. In any case, the p-values should be assessed together with other diagnostic checks, such as the QQ-plot.

5. Numerical results

5.1. Simulation study

In this section we report the results of a small simulation study. The SEPP model we simulated has a quadratic baseline intensity function,

$$\nu(t) = \mathrm{e}^{\beta_1} + \{\mathrm{e}^{\beta_2} + \mathrm{e}^{\beta_3}\}^2 (t - \mathrm{e}^{\beta_2}/(\mathrm{e}^{\beta_2} + \mathrm{e}^{\beta_3}))^2, \qquad t \in [0, 1],$$

and an exponential excitation function, $g(t) = \exp(\gamma_1 - \mathrm{e}^{\gamma_2} t)$, $t \geq 0$. Here the shape of the baseline intensity was motivated by the stock trading example considered below in Section 5.2, where the trading intensities seem high at market open and close, and low around middle of the day. The exponential forms were used to ensure positivity of the intensity. We simulated 100 sample paths of the SEPP, and estimated the γ_i and β_i using MLEs based on each of the simulated paths.

The results of the simulation are shown in Table 1. The first row gives the true values of each parameter. The second row gives the average of the 100 estimates. The third row gives the empirical standard error of each estimator, i.e. the standard deviation of the 100 estimates. The fourth row gives the average of the 100 standard error estimates obtained by inverting the observed information. The fifth row gives the normal QQ-plot of the 100 estimates of each parameter. The sixth and last row gives the p-values of the two-sided Kolmogorov-Smirnov tests of normality of the 100 estimates for each parameter, where the mean and variance of the hypothesized normal distribution are chosen to be the emprical mean and variance, respectively, of the 100 estimates.

From Table 1 we can see that the estimates are all close to their respective true parameter values, relative to the standard errors. The averages of the estimated standard errors are also close to the empirical standard errors. From the QQ-plots we can see that the distributions of the estimators are normal-looking, although the distributions of the estimates of the parameters involved in the excitation function have heavier than normal tails. The p-values of the KS tests of normality are all above the usual significance level of 0.05. We therefore conclude that the MLE works well, as predicted by our asymptotic theory.

5.2. Trade arrival process of an Australian stock

As pointed out by Engle and Russell (1998), financial durations, or equivalently the intensity of arrival of financial events, measure the speed of a financial market. Therefore, modeling of financial durations can cast light on the microeconomic structure of the financial market. In this section we apply the SEPPVB model to an ultra-high frequency financial dataset for the intraday trading times of a common stock traded on the Australian Stock Exchange, the ANZ stock. The data contain the times of all trades of the ANZ stock that occurred in December 2008,

TABLE 1: Results of the simulation study.

	γ_1	γ_2	β_1	β_2	β_3
True	4.3	6.0	7.0	4.0	3.8
Mean Est.	4.190	6.087	7.008	4.015	3.809
Emp. SE	0.802	0.950	0.092	0.056	0.058
Mean $\widehat{\text{SE}}$	0.626	0.721	0.109	0.056	0.063
QQ-plot	Figure 1(a)	Figure 1(b)	Figure 1(c)	Figure 1(d)	Figure 1(e)
p-val.	0.15	0.13	0.09	0.87	0.93

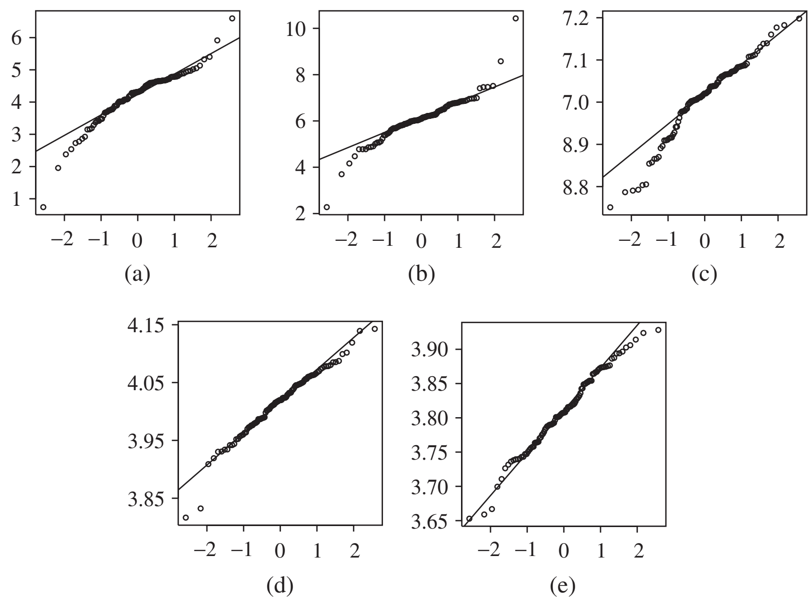

FIGURE 1.

Greenwich Mean Time (GMT). In this analysis we consider only the times of the trades that fall inside normal trading hours, 10:00-16:00 Australian Eastern Standard Time (AEST), except on 24 and 31 December when the normal trading period is 10:00-14:10. There are a total of 21 trading days in the month. The daily number of trades varies between 2620 and 11900 with a mean of 6633. Due to the finite precision in time recording, there are multiple trades associated with some trade times. Following convention we consider multiple trades at the same point in time as a single trade. With this interpretation, the number of trades reduces to the range [1493, 5821] with a mean of 3399. We fitted the SEPPVB to each of the 21 daily datasets, where the varying background intensity was assumed to be a B-spline with interior knots at each of the trading hours (11:00, 12:00, ...), and for the excitation function we used both the exponential decay function $g(t;\theta) = \gamma_1 \exp(-\gamma_2 t)$ and the polynomial decay function $g(t;\theta) = \gamma_1/(1+t)^{1+\gamma_2}$. We also assumed that the parameters involved in the baseline intensity function and in the excitation functions were separated.

The goodness-of-fit (GOF) results are shown in Figures 2–4 and Table 2. For the purpose of comparison we have also included the GOF results of fitting the Generalized Gamma ACD

https://doi.org/10.1239/jap/1389370096 Published online by Cambridge University Press

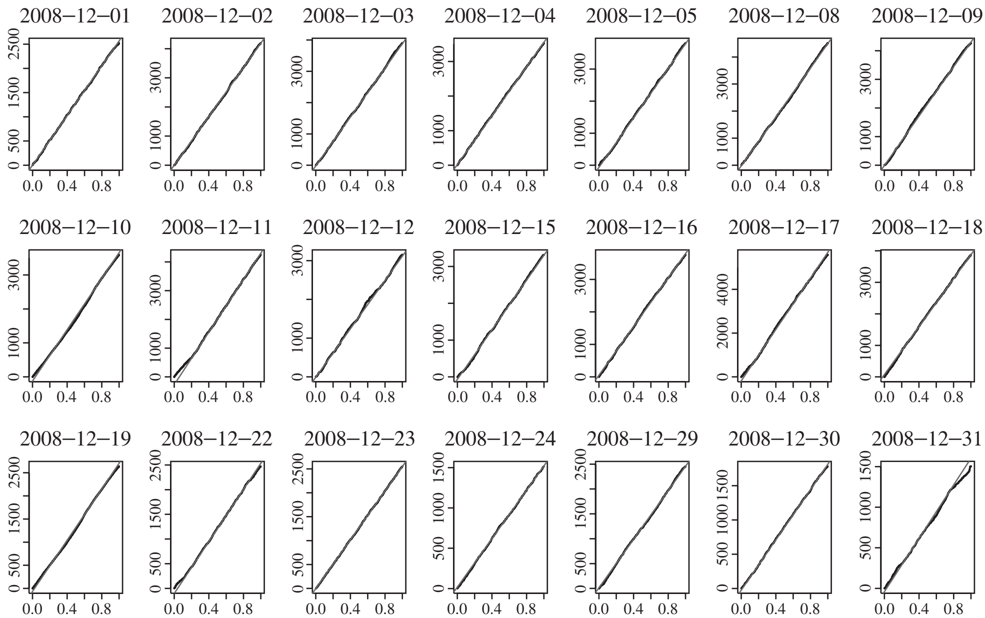

FIGURE 2: QQ-plots of the transformed event times in the SEPPVB-ED model.

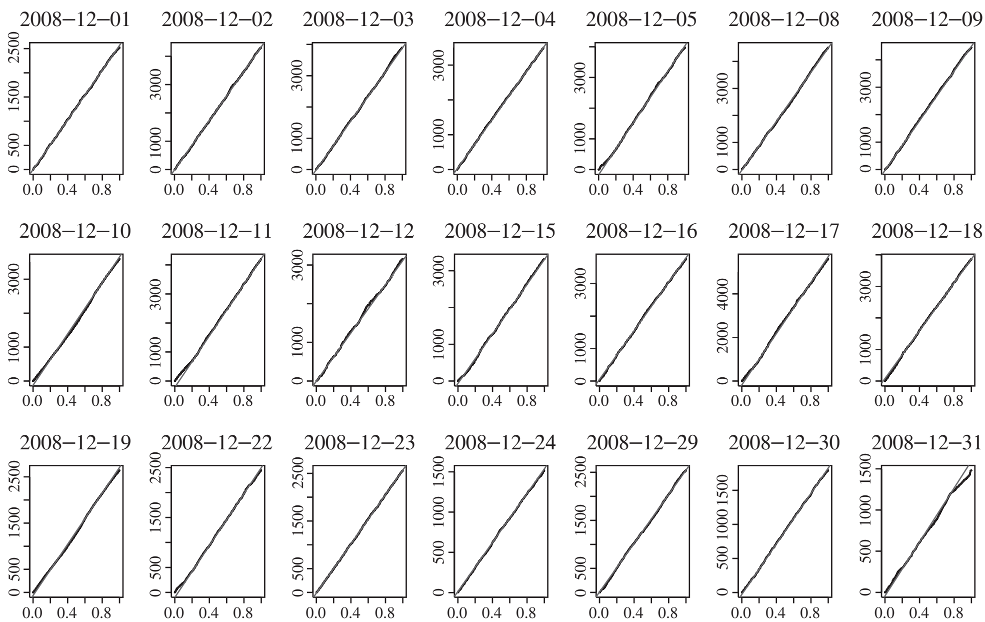

FIGURE 3: QQ-plots of the transformed event times in the SEPPVB-PD model.

https://doi.org/10.1239/jap/1389370096 Published online by Cambridge University Press

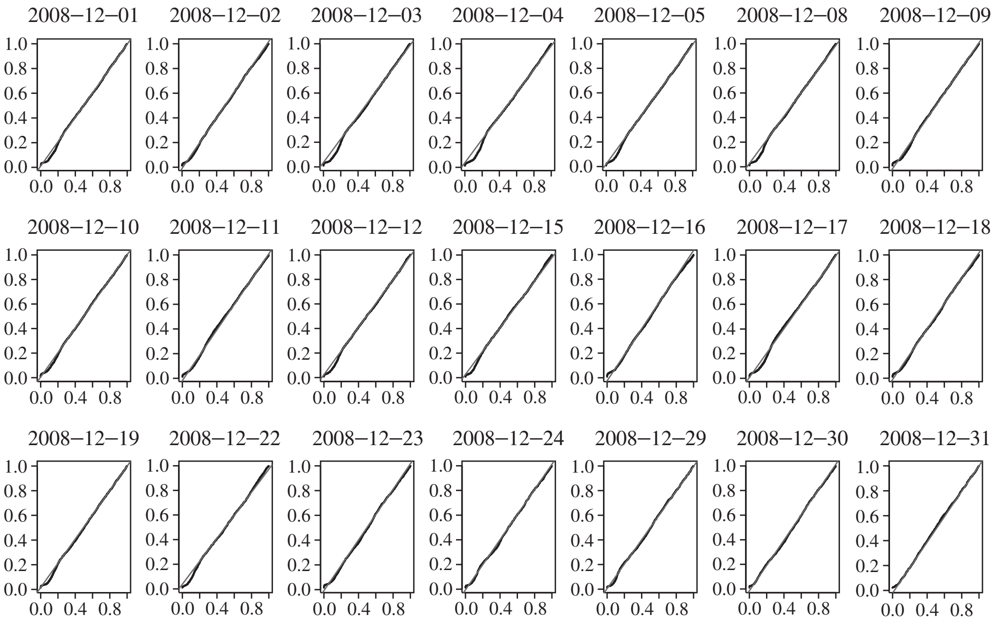

FIGURE 4: QQ-plots of the transformed event times in the GGACD(1,1) model.

TABLE 2: Goodness-of-fit results in the case of fitting the SEPPVB and GGACD models to the intraday trade times data of ANZ stock in December 2008.

Date	*p*-values of the KS tests of uniformity of the Rosenblatt residuals		
	SEPPVB-ED	SEPPVB-PD	GGACD(1,1)
12-01	0.14	0.14	0.01
12-02	0.43	0.13	0.0011
12-03	0.22	0.23	0.0005
12-04	0.12	0.16	1.5e-05
12-05	0.013	9.1e-05	0.00022
12-08	0.063	0.062	9.1e-05
12-09	0.014	0.058	1.3e-05
12-10	0.00016	0.00016	0.0039
12-11	5.3e-09	1.1e-06	0.00011
12-12	0.014	0.013	0.012
12-15	0.00028	0.00027	0.013
12-16	0.013	0.013	0.0021
12-17	2.6e-05	4.1e-05	8.1e-09
12-18	0.092	0.074	0.00071
12-19	0.075	0.075	0.0096
12-22	0.016	0.015	0.1
12-23	0.96	0.96	0.012
12-24	0.42	0.37	0.19
12-29	0.42	0.4	0.054
12-30	0.42	0.38	0.25
12-31	0.0041	0.00048	0.47

https://doi.org/10.1239/jap/1389370096 Published online by Cambridge University Press

(GGACD, cf. Zhang *et al.* 2001) model with autoregressive order (1,1) to the diurnally adjusted data in Figures 2–4 and Table 2. The GGACD model is a generalization of the Exponential and Weibull ACD models of Engle and Russell (1998). It assumes the durations $d_i = t_i - t_{i-1}$ between successive events of the SEPP have a generalized gamma distribution, given the past information of the process up to the last event time, and provided that the conditional mean of the duration, $\psi_i = \mathbb{E}[d_i \mid \mathcal{F}_{t_{i-1}}]$, satisfies an autoregressive relationship of order (m, q). That is,

$$d_i = \psi_i \epsilon_i,$$
$$\psi_i = \omega + \sum_{j=1}^{m} \alpha_j d_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j},$$

where the innovation ϵ_i is independent of $\mathcal{F}_{t_{i-1}}$ and has a generalized gamma distribution with density given by

$$f_\epsilon(x; a, k) = \frac{a x^{ka-1} \mathrm{e}^{-(x/b)^a}}{b^{ka}\Gamma(k)}, \quad \text{with} \quad b = \frac{\Gamma(k)}{\Gamma(k + 1/a)}.$$

In obtaining the diurnally adjusted data we followed Engle and Russell (1998) and estimated the expected raw durations as functions of the time of the day using regression splines, and then divided the raw residuals by the corresponding estimates of expected durations. The knots and order of the splines were the same as those used in the SEPPVB model. To assess the GOF of the GGACD model, we used the Rosenblatt (1952) transform method (see also Bauwens *et al.* 2004). We calculated the residuals $\hat{u}_i = \hat{F}_i(d_i)$, with $\hat{F}_i$ being the estimator of the conditional distribution function of d_i given $\mathcal{F}_{t_{i-1}}$, and obtained from the MLEs $\hat{a}$, $\hat{k}$, $\hat{\omega}$, $\hat{\alpha}_j$, $\hat{\beta}_j$, and the fitted conditional durations

$$\hat{\psi}_i = \hat{\omega} + \sum_{j=1}^{m} \hat{\alpha}_j d_{i-j} + \sum_{j=1}^{q} \hat{\beta}_j \psi_{i-j}.$$

We then calculated p-values of the two-sided KS tests of uniformity of the distribution of the $\hat{u}_i$. Small p-values indicate a poor fit.

From Figures 2–4 and Table 2 we note that, overall, the SEPPVB models provide a considerably better fit than the GGACD(1,1) model. Of the two SEPPVB models, the model with the exponential decay excitation function (SEPPVB-ED), seems to fit slightly better than the model with the polynomial decay function (SEPP-PD), although the results are for most days nearly identical.

The estimated parameters of the excitation function are shown in Table 3, from which we can see the estimated integral of the excitation function varies in the range 0.10–0.66. This suggests that between 10% and 66% of the transactions of the ANZ shares are likely to be the result of noisy trading excited by informative trading.

The estimated time-varying baseline intensity functions are shown in Figures 5 and 6, from which we note that the background intensities have a number of somewhat consistent features, such as high intensities near the open and close of the local market. We also note that 11:00 tends to be a time when the background intensity peaks locally. This phenomenon could be due to the fact that there are other major Asian stock markets opening at 11:00 or in the hour following, such as the Malaysian and Singaporean stock exchanges (at 11:00 AEST), the Hong Kong Stock Exchange (11:20 AEST), and the Shanghai and Shenzhen Stock Exchanges (at 11:30 AEST).

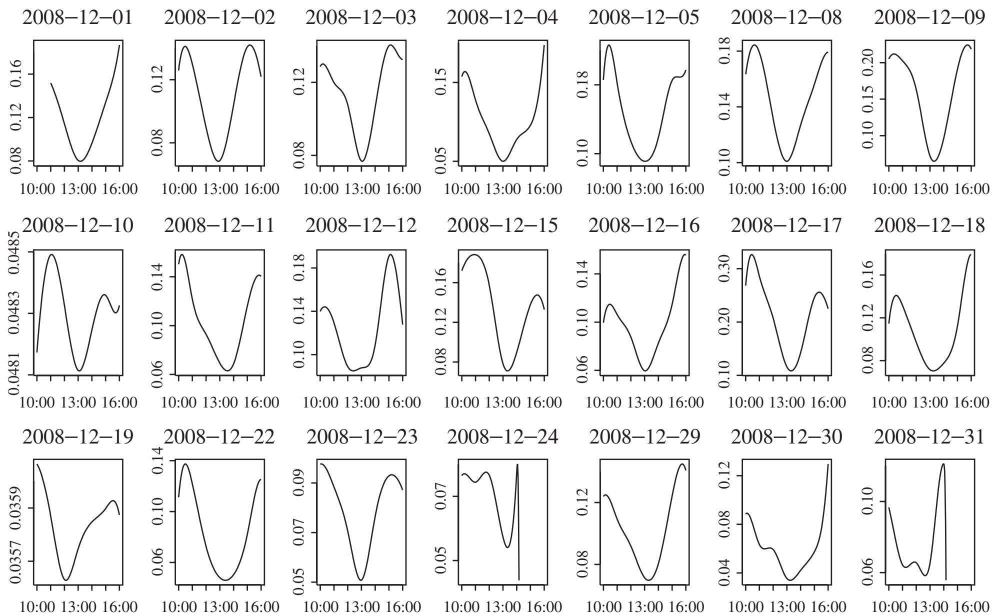

FIGURE 5: Estimated background intensities (in trades per second) in the SEPPVB model with exponentially decaying excitation.

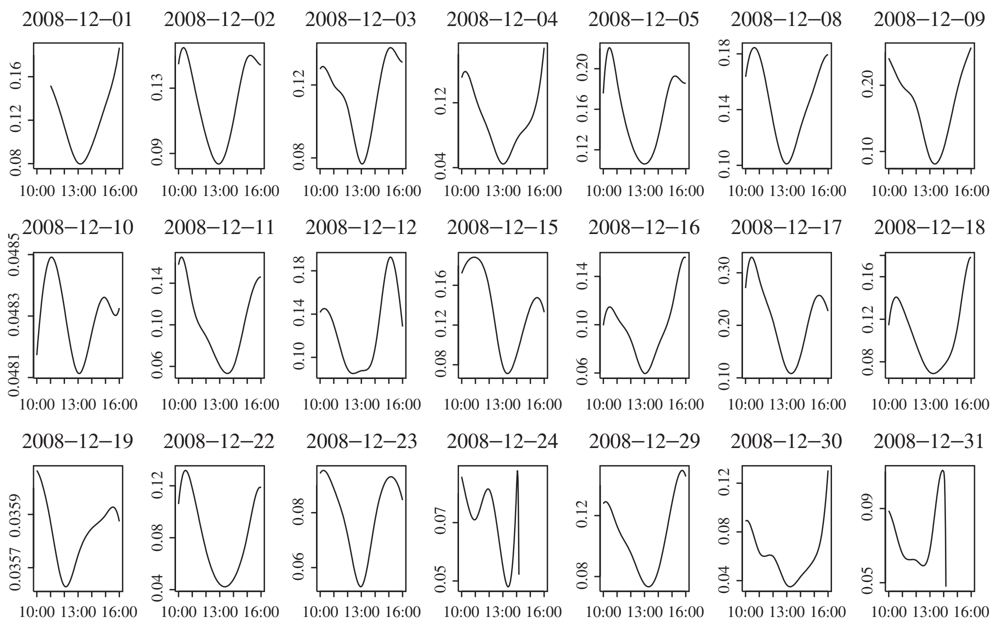

FIGURE 6: Estimated background intensities (in trades per second) in the SEPPVB model with polynomially decaying excitation.

https://doi.org/10.1239/jap/1389370096 Published online by Cambridge University Press

TABLE 3: Estimates of the parameters of the excitation functions in the SEPPVB models. In brackets are the standard errors (SE). An NaN SE value indicates the corresponding inverted Hessian matrix failed to be positive definite.

Date	SEPPVB-ED			SEPPVB-PD		
	$\log \gamma_1$	$\log \gamma_2$	$\int g \approx \gamma_1/\gamma_2$	$\log(\gamma_1)$	$\log(\gamma_2)$	$\int g \approx \gamma_1/\gamma_2$
12-01	11.05(0.087)	12.88(0.087)	0.16(0.0088)	11.05(0.087)	12.88(0.087)	0.16(0.0088)
12-02	9.34(0.059)	10.22(0.072)	0.41(0.015)	9.42(0.062)	10.36(0.078)	0.39(0.015)
12-03	9.23(0.073)	10.25(0.095)	0.36(0.016)	9.23(0.081)	10.25(0.11)	0.36(0.019)
12-04	9.24(0.066)	10.19(0.072)	0.38(0.013)	9.21(0.076)	10.15(0.099)	0.39(0.018)
12-05	11.31(0.073)	13.23(0.072)	0.15(0.0069)	11.32(0.072)	13.25(0.07)	0.15(0.0068)
12-08	9.37(NaN)	10.57(NaN)	0.30(NaN)	9.37(0.08)	10.57(0.11)	0.30(0.015)
12-09	11.43(0.062)	13.09(0.064)	0.19(0.0074)	11.45(0.062)	13.13(0.062)	0.19(0.0073)
12-10	8.25(0.16)	8.67(0.22)	0.66(0.044)	8.25(0.21)	8.67(0.29)	0.66(0.057)
12-11	9.49(0.058)	10.24(0.073)	0.47(0.016)	9.47(0.057)	10.21(0.07)	0.48(0.016)
12-12	11.16(0.081)	13.18(0.078)	0.13(0.0071)	11.16(0.081)	13.18(0.078)	0.13(0.0071)
12-15	11.08(0.09)	13.23(0.092)	0.12(0.0067)	11.08(0.09)	13.23(0.092)	0.12(0.0067)
12-16	9.68(0.056)	10.51(0.065)	0.44(0.014)	9.68(0.056)	10.51(0.064)	0.44(0.014)
12-17	11.47(0.053)	13.25(0.049)	0.17(0.006)	11.47(0.053)	13.25(0.049)	0.17(0.006)
12-18	9.56(0.058)	10.50(0.069)	0.39(0.014)	9.56(0.058)	10.49(0.069)	0.39(0.014)
12-19	7.86(NaN)	8.29(NaN)	0.65(NaN)	7.86(NaN)	8.29(NaN)	0.65(NaN)
12-22	8.81(0.12)	10.15(0.16)	0.26(0.021)	8.75(0.12)	10.04(0.16)	0.28(0.021)
12-23	9.54(0.079)	10.65(0.097)	0.33(0.016)	9.54(0.08)	10.66(0.098)	0.33(0.016)
12-24	8.77(0.098)	9.95(0.11)	0.31(0.019)	8.77(0.097)	9.96(0.11)	0.31(0.019)
12-29	11.04(0.11)	13.37(0.1)	0.10(0.0068)	11.05(0.11)	13.39(0.1)	0.10(0.0068)
12-30	8.90(0.099)	10.17(0.12)	0.28(0.018)	8.91(0.099)	10.18(0.12)	0.28(0.018)
12-31	9.26(0.11)	10.69(0.12)	0.24(0.017)	9.24(0.11)	10.65(0.12)	0.24(0.017)

6. Technical arguments

Proof of Theorem 1. If we can show that θ_0 is a suitably separated zero of the uniform in-probability limit, $S(\theta)$, of the normalized score function $a_n^{-1}S_n(\theta)$, taking the limit in a neighborhood of θ_0, then, in view of the Z-estimator master theorem (e.g Theorem 5.9 of van der Vaart 1998), the proof is complete.

We have seen from (5) that the mean of $N_n(t)$ is $a_n \int_0^t h(u;\theta_0)\,\mathrm{d}u \equiv a_n H(t;\theta_0)$. So we expect that

$$\sup_{t\in[0,1]} |a_n^{-1}N_n(t) - H(t;\theta_0)| \xrightarrow{\mathbb{P}} 0. \tag{8}$$

Since the normalized score function can be written as

$$\begin{aligned} a_n^{-1}S_n(\theta) = &\int_0^1 \frac{\partial_\theta \mu(t;\theta) + \int_0^t \partial_\theta g(t-u;\theta)\,\mathrm{d}a_n^{-1}N_n(u)}{\mu(t;\theta) + \int_0^t g(t-u;\theta)\,\mathrm{d}a_n^{-1}N_n(u)}\,\mathrm{d}a_n^{-1}N_n(t) \\ &- \int_0^1 \left\{\partial_\theta \mu(t;\theta) + \int_0^t \partial_\theta g(t-u;\theta)\,\mathrm{d}a_n^{-1}N_n(u)\right\}\mathrm{d}t, \end{aligned}$$

https://doi.org/10.1239/jap/1389370096 Published online by Cambridge University Press

it also seems natural to expect its limit to be

$$s(\theta) = \int_0^1 \frac{\partial_\theta \mu(t;\theta) + \int_0^t \partial_\theta g(t-u;\theta) h(u;\theta_0)\,\mathrm{d}u}{\mu(t;\theta) + \int_0^t g(t-u;\theta) h(u;\theta_0)\,\mathrm{d}u} h(t;\theta_0)\,\mathrm{d}t \\ - \int_0^1 \left\{ \partial_\theta \mu(t;\theta) + \int_0^t \partial_\theta g(t-u;\theta) h(u;\theta_0)\,\mathrm{d}u \right\} \mathrm{d}t.$$

Since $h(t;\theta_0)$ satisfies the Volterra equation (4) (with a_n there dropped), it is clear that $s(\theta_0) = 0$. The derivative matrix of $s(\theta)$ is given by

$$\partial_\theta s(\theta)^\top = \int_0^1 \frac{\partial_\theta^{\otimes 2} \mu(t;\theta) + \int_0^t \partial_\theta^{\otimes 2} g(t-u;\theta) h(u;\theta_0)\,\mathrm{d}u}{\mu(t;\theta) + \int_0^t g(t-u;\theta) h(u;\theta_0)\,\mathrm{d}u} h(t;\theta_0)\,\mathrm{d}t \\ - \int_0^1 \frac{\{\partial_\theta \mu(t;\theta) + \int_0^t \partial_\theta g(t-u;\theta) h(u;\theta_0)\,\mathrm{d}u\}^{\otimes 2}}{\{\mu(t;\theta) + \int_0^t g(t-u;\theta) h(u;\theta_0)\,\mathrm{d}u\}^2} h(t;\theta_0)\,\mathrm{d}t \\ - \int_0^1 \left\{ \partial_\theta^{\otimes 2} \mu(t;\theta) + \int_0^t \partial_\theta^{\otimes 2} g(t-u;\theta) h(u;\theta_0)\,\mathrm{d}u \right\} \mathrm{d}t. \tag{9}$$

In view of condition C5, the matrix $\partial_\theta s(\theta)^\top$ is negative semi-definite at θ_0 since the first and last terms on the right-hand side of (9) cancel when $\theta = \theta_0$. Noting the continuity of $\mu(\cdot;\theta)$, $g(\cdot;\theta)$, and $h(\cdot;\theta)$ in θ, we see that $\partial_\theta s(\theta)^\top$ is negative definite in a neighborhood U of θ_0. So, θ_0 is the unique and well separated zero of $s(\theta)$ in U, in the sense that

$$\sup_{\theta \in U:\ \|\theta - \theta_0\| \geq \epsilon} \|s(\theta)\| > \|s(\theta_0)\| = 0 \quad \text{for all } \epsilon > 0.$$

Now we only need to establish the uniform in-probability convergence,

$$\sup_{\theta \in \Theta} \|a_n^{-1} S_n(\theta) - s(\theta)\| \xrightarrow{\mathbb{P}} 0. \tag{10}$$

To this end, we first show that $a_n^{-1} S_n(\theta) - s(\theta) \xrightarrow{\mathbb{P}} 0$ for each $\theta \in \Theta$. With $g'(t;\theta)$ denoting $\mathrm{d}g(t;\theta)/\mathrm{d}t$, by integration by parts and the assumed conditions on the excitation function g, we have the following estimates for $j = 0, 1$ and all $t \in [0, 1]$:

$$\left| \int_0^t \partial_\theta^j g(t-u;\theta) \{ \mathrm{d}a_n^{-1} N(u) - h(u;\theta_0)\,\mathrm{d}u \} \right| \\ = \left| \{a_n^{-1} N(t) - H(t;\theta_0)\} \partial_\theta^j g(0;\theta) + \int_0^t \{a_n^{-1} N(u) - H(u;\theta_0)\} \partial_\theta^j g'(t-u;\theta)\,\mathrm{d}u \right| \\ \leq \left\{ \left| \partial_\theta^j g(0;\theta) \right| + \int_0^1 \left| \partial_\theta^j g'(t;\theta) \right| \mathrm{d}t \right\} \times \sup_{t \in [0,1]} |a_n^{-1} N(t) - H(t;\theta_0)|,$$

where $|\cdot|$ and $\leq$ are understood component-wise when applied to a vector. Therefore, using integration by parts and the continuous mapping theorem (e.g. Kallenberg 2002, Lemma 4.3), we deduce that $a_n^{-1} S_n(\theta) - s(\theta) \xrightarrow{\mathbb{P}} 0$ for each θ, provided we can establish (8).

We now derive (8) using a Poisson representation method (e.g. Kurtz 1983). Although the law of large numbers result of Kurtz (1983) for SEPPs is not directly applicable, the proof below is modeled on that of Kurtz' Theorem 2.1. Let Y be a unit intensity Poisson process, and X_n a SEPP defined by

$$X_n(t) = Y\Big(a_n \int_0^t \{\mu(u;\theta_0) + \int_{[0,u)} g(u-s;\theta_0)\,\mathrm{d}X_n(s)\}\,\mathrm{d}u\Big), \qquad t \in [0,1].$$

Since the intensity of X_n is given by $a_n\mu(t;\theta_0) + \int_{[0,t)} g(t-u;\theta_0)\,\mathrm{d}X_n(u)$ and $X_n(0) = N_n(0) = 0$, we see that N_n and X_n are equal in distribution. Therefore, we will satisfy (8) if we can prove it with N_n replaced by X_n, that is,

$$\sup_{t\in[0,1]} |a_n^{-1}X_n(t) - H(t;\theta_0)| \xrightarrow{\mathbb{P}} 0.$$

Recall that $h(t;\theta_0)$ satisfies the integral equation (4) with the a_n dropped. So, if we let $\tilde{Y}(t) = Y(t) - t$ be the compensated Poisson process, then, by Fubini's theorem and integration by parts, we can write

$$\begin{aligned}
a_n^{-1}X_n(t) - H(t;\theta_0) &= a_n^{-1}\tilde{Y}\bigg(a_n \int_0^t \bigg\{\mu(u;\theta_0) + \int_{[0,u)} g(u-s;\theta_0)\,\mathrm{d}a_n^{-1}X_n(s)\bigg\}\,\mathrm{d}u\bigg) \\
&\quad + \int_0^t \int_{[0,u)} g(u-s;\theta_0)\,\mathrm{d}\{a_n^{-1}X_n(s) - H(s;\theta_0)\}\,\mathrm{d}u \\
&= a_n^{-1}\tilde{Y}\bigg(a_n \int_0^t \mu(u;\theta_0) + a_n^{-1}X_n(u)g(t-u;\theta_0)\,\mathrm{d}u\bigg) \\
&\quad + \int_0^t \{a_n^{-1}X_n(u) - H(u;\theta_0)\}g(t-u;\theta_0)\,\mathrm{d}u \\
&\equiv \epsilon_n(t) + \int_0^t \{a_n^{-1}X_n(u) - H(u;\theta_0)\}g(t-u;\theta_0)\,\mathrm{d}u. \qquad (11)
\end{aligned}$$

If we write $\bar{\mu} = \sup_{t\in[0,1]} \mu(t;\theta_0)$ and $\bar{g} = \sup_{t\in[0,1]} g(t;\theta_0)$, and define another self-exciting process Z_n by

$$Z_n(t) = Y\bigg[a_n \int_0^t \{\bar{\mu} + \bar{g}a_n^{-1}Z_n(u)\}\,\mathrm{d}u\bigg],$$

then it is clear that $\sup_{u\in[0,t]} X_n(u) = X_n(t) \le Z_n(t)$ and by a strong law of large numbers of Kurtz for self-exciting counting processes (see e.g. Kurtz 1981, Theorem 8.1), with probability 1, $a_n^{-1}Z_n(t) \to Z(t)$ uniformly in t on compact intervals, for $Z(t)$ satisfying

$$Z(t) = \int_0^t \bar{\mu} + \bar{g}Z(u)\,\mathrm{d}u,$$

or explicitly, $Z(t) = (\bar{\mu}/\bar{g})\{\exp(\bar{g}t) - 1\}$. So, for sufficiently large n,

$$\begin{aligned}
\sup_{t\in[0,1]} |\epsilon_n(t)| &= \sup_{t\in[0,1]} \bigg| a_n^{-1}\tilde{Y}\bigg\{a_n \int_0^t \mu(u;\theta_0) + a_n^{-1}X_n(u)g(t-u;\theta_0)\,\mathrm{d}u\bigg\}\bigg| \\
&\le \sup_{s\in[0,a_n(\bar{\mu}+(Z(1)+1)\bar{g})]} |a_n^{-1}\tilde{Y}(s)| \\
&\xrightarrow{\mathbb{P}} 0,
\end{aligned}$$

https://doi.org/10.1239/jap/1389370096 Published online by Cambridge University Press

where the in-probability convergence is implied by Doob's inequality (e.g. Revuz and Yor 1999, Theorem II.1.7): since $a_n^{-1}\tilde{Y}$ is a martingale, it follows for all $\epsilon, T > 0$, that

$$\mathbb{P}\left\{\sup_{s\in[0,a_nT]} a_n^{-1}|\tilde{Y}(s)| > \epsilon\right\} \le \mathbb{E}\{a_n^{-2}\tilde{Y}(a_nT)^2\}/\epsilon^2 = a_n^{-2}a_nT/\epsilon^2 \to 0.$$

Noticing (11) implies $|a_n^{-1}X_n(t) - H(t;\theta_0)| \le |\epsilon_n(t)| + \int_0^t |a_n^{-1}X_n(u) - H(u;\theta_0)|\bar{g}\,\mathrm{d}u$, and, therefore, also

$$\begin{aligned}\sup_{u\in[0,t]} &|a_n^{-1}X_n(u) - H(u;\theta_0)| \\ &\le \sup_{u\in[0,1]} |\epsilon_n(u)| + \bar{g}\int_0^t \sup_{s\in[0,u]} |a_n^{-1}X_n(s) - H(s;\theta_0)|\,\mathrm{d}u, \qquad t\in[0,1],\end{aligned}$$

we have, by Gronwall's inequality (e.g. Kallenberg 2002, Lemma 21.4),

$$\sup_{u\in[0,t]} |a_n^{-1}X_n(u) - H(u;\theta_0)| \le \sup_{u\in[0,1]} |\epsilon_n(u)|e^{\bar{g}t} \xrightarrow{\mathbb{P}} 0, \qquad t\in[0,1].$$

This proves (8) and, therefore, $a_n^{-1}S_n(\theta) - s(\theta) \xrightarrow{\mathbb{P}} 0$ for each $\theta\in\Theta$.

To complete the proof, it remains to show this convergence is uniform in θ. Now note for each $\epsilon > 0$, by the assumed regularity conditions on μ and g, we can find a $\delta > 0$ such that, for all $\theta_1,\theta_2\in\Theta$, $\|\theta_1-\theta_2\| < \delta$ implies $\|a_n^{-1}S_n(\theta_1) - a_n^{-1}S_n(\theta_2)\| \le \epsilon/3$ with probability tending to 1, and, moreover, $\|s(\theta_1) - s(\theta_2)\| \le \epsilon/3$. For such a δ, since Θ is a compact set in $\mathbb{R}^d$, there exist a finite number of open balls with radius δ whose union covers Θ. Let $\vartheta_1,\dots,\vartheta_K$ denote the centers of these balls, and, for all θ, let $\vartheta_{i(\theta)}$ be the center of a ball that contains θ. Note that

$$\begin{aligned}\mathbb{P}\Big(\sup_\theta &\|a_n^{-1}S_n(\theta) - s(\theta)\| > \epsilon\Big) \\ &\le \mathbb{P}\Big(\sup_\theta \|a_n^{-1}S_n(\theta) - a_n^{-1}S_n(\vartheta_{i(\theta)})\| > \epsilon/3\Big) \\ &\quad + \mathbb{P}\Big(\sup_\theta \|a_n^{-1}S_n(\vartheta_{i(\theta)}) - s(\vartheta_{i(\theta)})\| > \epsilon/3\Big) + \mathbb{P}\Big(\sup_\theta \|s(\vartheta_{i(\theta)}) - s(\theta)\| > \epsilon/3\Big).\end{aligned}$$

By definition, $\|\vartheta_{i(\theta)} - \theta\| < \delta$, so the third term on the right equals 0 and the first term converges to 0 as $n\to\infty$. The second term above also converges to 0 since it is bounded by

$$\sum_{i=1}^K \mathbb{P}(\|a_n^{-1}S_n(\vartheta_i) - s(\vartheta_i)\| > \epsilon/3) \to 0.$$

This completes the proof of (10), and therefore of the theorem.

Proof of Theorem 2. The proof is based on the mean value theorem and the martingale central limit theorem (Andersen *et al.* 1993; Fleming and Harrington 1991). First note that, by the mean value theorem,

$$0 = a_n^{-1/2}S_n(\hat{\theta}) = a_n^{-1/2}S_n(\theta_0) + a_n^{-1}\partial_\theta S_n(\theta)|_{\theta=\theta^*}a_n^{1/2}(\hat{\theta} - \theta_0),$$

where θ^* lies on the line segment joining θ_0 and $\hat{\theta}$. So we have, at least formally,

$$a_n^{1/2}(\hat{\theta}-\theta_0) = -\{a_n^{-1}\partial_\theta S_n(\theta)|_{\theta=\theta^*}\}^{-1} a_n^{-1/2} S_n(\theta_0).$$

By Slutsky's theorem, it suffices to show

$$a_n^{-1/2} S_n(\theta_0) \xrightarrow{\mathrm{D}} N(0, \mathcal{I}(\theta_0)), \tag{12}$$

and

$$a_n^{-1}\partial_\theta S_n(\theta)|_{\theta=\theta^*} \xrightarrow{\mathbb{P}} -\mathcal{I}(\theta_0). \tag{13}$$

To show (12), we note from (2) that $a_n^{-1/2} S_n(\theta_0)$ has the martingale transform structure, $a_n^{-1/2} S_n(\theta_0) = \int_0^1 U_n(t)\,\mathrm{d}M_n(t)$, with

$$U_n(t) = a_n^{-1/2}\frac{\partial_\theta \mu(t;\theta_0) + \int_0^t \partial_\theta g(t-u;\theta_0)\,\mathrm{d}a_n^{-1}N_n(u)}{\mu(t;\theta_0) + \int_0^t g(t-u;\theta_0)\,\mathrm{d}a_n^{-1}N_n(u)},$$
$$M_n(t) = N_n(t) - \int_0^t \left\{a_n\mu(u;\theta_0) + \int_0^u g(u-s;\theta_0)\,\mathrm{d}N_n(s)\right\}\mathrm{d}u.$$

Note that, since $M_n(t)$ is a square-integrable martingale and $U_n(t)$ is locally bounded and predictable, $\tilde{M}_n(t) = \int_0^t U_n(s)\,\mathrm{d}M_n(s)$, $t \in [0,1]$, is a locally square integrable martingale, and $a_n^{-1/2} S_n(\theta_0) = \tilde{M}_n(1)$. Since the predictable variation process of $N_n(t)$ is equal to its compensator, that is, $\langle N_n\rangle(t) = \int_0^t \lambda_n(u;\theta_0)\,\mathrm{d}u$, the predictable variation process of $\tilde{M}(t)$ is

$$\begin{aligned}\langle \tilde{M}_n\rangle(t) &= \int_0^t U_n(s)^{\otimes 2}\,\mathrm{d}\langle \tilde{M}_n\rangle(s)\\ &= \int_0^t \frac{\{\partial_\theta \mu(u;\theta_0) + \int_0^u \partial_\theta g(u-s;\theta_0)\,\mathrm{d}a_n^{-1}N_n(s)\}^{\otimes 2}}{\mu(u;\theta_0) + \int_0^u g(u-s;\theta_0)\,\mathrm{d}a_n^{-1}N_n(s)}\,\mathrm{d}u,\end{aligned}$$

which converges to

$$v(t) \equiv \int_0^t \frac{\{\partial_\theta \mu(u;\theta_0) + \int_0^u \partial_\theta g(u-s;\theta_0)\,\mathrm{d}H(s;\theta_0)\}^{\otimes 2}}{\mu(u;\theta_0) + \int_0^u g(u-s;\theta_0)\,\mathrm{d}H(s;\theta_0)}\,\mathrm{d}u$$

in probability for $t \in [0,1]$.

Next we look at the large jumps of $\tilde{M}_n(t)$. Write $\tilde{M}_{n,\epsilon}(t) = \int_0^t U_n(s)\mathbf{1}\{|U_n(s)| > \epsilon\}\,\mathrm{d}M_n(s)$, where, as usual, the function $\mathbf{1}\{|\cdot| > \epsilon\}$ is understood dimension-wise when applied to a vector. Then $\tilde{M}_{n,\epsilon}(t)$ accumulates all the jumps of $\tilde{M}_n$, before time t, that exceed ϵ. Noting that $U_n(t) = O_{\mathbb{P}}(a_n^{-1/2})$ uniformly in $t \in [0,1]$, and by Lenglart's inequality (Andersen *et al.* 1993; Fleming and Harrington 1991; Lenglart 1977), we deduce that $\sup_{s\in[0,t]} |\tilde{M}_{n,\epsilon}(s)| \xrightarrow{\mathbb{P}} 0$ dimension-wise for all $t \in [0,1]$. By the martingale central limit theorem, we conclude that $\tilde{M}(t)$ converges in distribution to a Gaussian process with covariance function $v(t)$. Therefore, $a_n^{-1/2} S_n(\theta_0) = \tilde{M}(1)$ converges in distribution to a zero mean normal random vector with variance matrix $v(1) = \mathcal{I}(\theta_0)$. This proves (12).

Since we have seen that $\sup_{t\in[0,1]} |a_n^{-1}N_n(t) - H(t;\theta_0)| \xrightarrow{\mathbb{P}} 0$ in the proof of Theorem 1, and using an argument essentially the same as that leading to $\sup_{\theta\in\Theta} \|a_n^{-1}S_n(\theta) - s(\theta)\| \xrightarrow{\mathbb{P}} 0$ in the proof of Theorem 1, it can be shown that $\sup_{\theta\in\Theta} \|a_n^{-1}\partial_\theta S_n(\theta)^\top - \partial_\theta s(\theta)\| \xrightarrow{\mathbb{P}} 0$. Since θ^*

https://doi.org/10.1239/jap/1389370096 Published online by Cambridge University Press

is sandwiched between $\hat\theta$ and θ_0, and $\hat\theta \xrightarrow{\mathbb{P}} \theta_0$, it follows that $\theta^* \xrightarrow{\mathbb{P}} \theta_0$ as well. By the continuity of $\partial_\theta s(\theta)^\top$ in θ and the continuous mapping theorem, we also have

$$\partial_\theta s(\theta^*)^\top \xrightarrow{\mathbb{P}} \partial_\theta s(\theta_0)^\top = -\mathcal{I}(\theta_0).$$

Furthermore, (13) follows from the observation that

$$\begin{aligned}
&\|a_n^{-1}\partial_\theta S_n(\theta^*)^\top - \partial_\theta s(\theta_0)^\top\| \\
&\quad\le \|a_n^{-1}\partial_\theta S_n(\theta^*)^\top - \partial_\theta s(\theta^*)^\top\| + \|\partial_\theta s(\theta^*)^\top - \partial_\theta s(\theta_0)^\top\| \\
&\quad\le \sup_{\theta\in\Theta} \|a_n^{-1}\partial_\theta S_n(\theta)^\top - \partial_\theta s(\theta)^\top\| + \|\partial_\theta s(\theta^*)^\top - \partial_\theta s(\theta_0)^\top\|,
\end{aligned}$$

which proves (7).

Finally we note that the property $a_n^{-1} I(\hat\theta) \xrightarrow{\mathbb{P}} \mathcal{I}(\theta_0)$ follows along almost the same lines as the above proof of $a_n^{-1}\partial_\theta S_n(\theta^*)^\top \xrightarrow{\mathbb{P}} -\mathcal{I}(\theta_0)$.

7. Discussion

In this paper we proved that maximum likelihood estimators of parametric SEPPs (or Hawkes processes) with time-varying background intensity (SEPPVBs) are consistent and asymptotically normal as the normalizing constant in the baseline intensity tends to infinity, under weak regularity conditions on the model. We also demonstrated that SEPPVBs comprise a promising and competitive alternative to the more popular ACD type models initially developed by Engle and Russell (1998) for financial duration modeling.

In establishing consistency we used a Cramér type method (Cramér 1946). It is natural to wonder whether a Wald type method (Wald 1949) would work as well. The apparent answer would seem to be 'no', since under our model there does not appear to be a function definable as the limit of the normalized log-likelihood which plays the role of the expectation of the log density in the i.i.d. data case. However, if suitably 'hacked', a Wald method still works. The trick is to note there is a sequence of deterministic functions which have a common global maximizer at the true parameter value, and are uniformly close (in probability) to the loglikelihood function. In fact, if we take

$$\begin{aligned}
m_n(\theta) = &\int_0^1 \log\{a_n\mu_\theta(t) + a_n(g_\theta * h_{\theta_0})(t)\} a_n h_{\theta_0}(t)\,\mathrm{d}t \\
&- \int_0^1 \{a_n\mu_\theta(t) + a_n(g_\theta * h_\theta)(t)\}\,\mathrm{d}t,
\end{aligned}$$

with $*$ denoting convolution, then it can be shown that $|\log L_n(\theta) - m_n(\theta)| = o_p(a_n)$ uniformly in a neighborhood of θ_0, and, moreover, $m_n(\theta_0) - m_n(\theta) \ge 0$ for all n by the elementary inequality $\log x \ge 1 - x^{-1}$ for $x > 0$. A similar hacking of the Cramér method has been used by Chen (2011) to establish consistency of local polynomial estimators of counting process intensity functions and their derivatives.

The asymptotic setting in this paper does not allow the excitation function to vary with the limit index. However, based on the shape of the estimated excitation function in the UHF financial data example, it seems justifiable to consider the asymptotics in a context where the shape of the excitation function varies while its integral is held fixed, e.g. via a scaling factor which is allowed to tend to infinity with the intensity. Asymptotics of this type would make interesting research problems.

In the UHF data example we considered above, we treated multiple trades with the same time stamp as a single trade. This was only for convenience, and is likely to have led to underestimated baseline trading intensities. To remedy this, one possibility is to jitter the tied event times, and average the resulting estimates from multiple jittered data sets. Another treatment is to consider point process models that allow simultaneous jumps, or marked point process (MPP) models. Multiplicity of an event time, or the size of a jump on the path of the MPP, can enter the model as event marks together with other features of the event, such as price and volume of a trade in the UHF financial data example. Inference for self-exciting MPPs is another interesting research problem.

As suggested by a referee, the multivariate Hawkes process, or mutually exciting point process, also has applications in many applied settings, such as finance (Embrechts *et al.* 2011). The inference for mutually exciting point processes with time-varying baseline intensities is also an important problem.

Acknowledgements

The comments of an anonymous reviewer have led to improved presentation, and we are grateful. Data used for this work was supplied by the Securities Industry Research Centre of Asia-Pacific (SIRCA) on behalf of Reuters. FC thanks Ryan Findlayson, Alexander Meister, and Aihua Xia for stimulating discussions on technicalities involved in the proof of the asymptotic theory. Most of the work was done while FC was visiting the Department of Mathematics and Statistics of the University of Melbourne on a special study leave. FC's work was partly supported by a UNSW ECR grant.

References

Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer, New York.

Bauwens, L., Giot, P., Grammig, J. and Veredas, D. (2004). A comparison of financial duration models via density forecasts. *Internat. J. Forecasting* **20,** 589–609.

Chen, F. (2011). Maximum local partial likelihood estimators for the counting process intensity function and its derivatives. *Statistica Sinica* **21,** 107–128.

Chornoboy, E., Schramm, L. and Karr, A. (1988). Maximum likelihood identification of neural point process systems. *Biol. Cybernetics* **59,** 265–275.

Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press.

Crane, R. and Sornette, D. (2008). Robust dynamic classes revealed by measuring the response function of a social system. *Proc. Nat. Acad. Sci.* **105,** 15649–15653.

Embrechts, P., Liniger, T. and Lin, L. (2011). Multivariate Hawkes processes: an application to financial data. *J. Appl. Prob.* **48,** 367–378.

Engle, R. F. and Lunde, A. (2003). Trades and quotes: A bivariate point process. *J. Financial Econometrics* **1,** 159–188.

Engle, R. F. and Russell, J. R. (1998). Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica*. **66,** 1127–1162.

Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. Wiley, New York.

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

Hawkes, A. G. and Oakes, D. (1974). A cluster process representation of a self-exciting process. *J. Appl. Prob.* **11,** 493–503.

Kallenberg, O. (2002). *Foundations of Modern Probability*, 2nd edn. Springer, New York.

Koopman, S. J., Lucas, A. and Monteiro, A. (2008). The multi-state latent factor intensity model for credit. *J. Econometrics* **142,** 399–424.

Kurtz, T. G. (1981). *Approximation of Population Processes*. Soc. Industrial Appl. Math. Philadelphia, Pennsylvania.

Kurtz, T. G. (1983). Gaussian approximations for Markov chains and counting processes. In *Proc. 44th Session Internat. Statist. Inst.* (Madrid, 1983; Bull. Inst. Internat. Statist. **50**), Vol. 1, pp. 361–376.

Lenglart, E. (1977). Relation de domination entre deux processus. *Ann. Inst. H. Poincaré (B)* **13,** 171–179.

Linz, P. (1985). *Analytical and Numerical Methods for Volterra Equations*. SIAM, Philadelphia.
Monteiro, A. A. (2009). The econometrics of randomly spaced financial data: a survey. UC3M working paper, Statistics and Econometrics 09–24, Universidad Carlos III de Madrid. Available at http://hdl.handle.net/10016/5995.
Ogata, Y. (1978). The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Ann. Inst. Statist. Math.* **30,** 243–261.
Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *J. Amer. Statist. Assoc.* **83,** 9–27.
Rathbun, S. L. (1996). Asymptotic properties of the maximum likelihood estimator for spatio-temporal point processes. *J. Statist. Planning Infer.* **51,** 55–74.
Revuz, D. and Yor, M. (1999). *Continuous Martingales and Brownian Motion*, 3rd edn. Springer, New York.
Rosenblatt, M. (1952). Remarks on a multivariate transformation. *Ann. Math. Statist.* **23,** 470–472.
Teich, M. C. and Saleh, B. E. A. (2000). Branching processes in quantum electronics. *IEEE J. Sel. Topics in Quantum Electronics* **6,** 1450–1457.
Utsu, T. (1961). A statistical study of the occurrence of aftershocks. *Geophysical Magazine* **30,** 521–605.
van der Vaart, A. (1998). *Asymptotic Statistics*. Cambridge University Press, New York.
Vere-Jones, D. (1995). Forecasting earthquakes and earthquake risk. *Internat. J. Forecasting* **11,** 503–538.
Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist.* **20,** 595–601.
Zhang, M. Y., Russell, J. R. and Tsay, R. S. (2001). A nonlinear autoregressive conditional duration model with applications to financial transaction data. *J. Econometrics* **104,** 179–207.