Working Paper Series
Department of Economics
University of Verona

Monte Carlo likelihood inference for marked doubly stochastic Poisson processes with intensity driven by marked point processes

Marco Minozzo, Silvia Centanni

WP Number: 11 March 2012

ISSN: 2036-2919 (paper), 2036-4679 (online)

Journal of Statistical Computation & Simulation
Vol. 00, No. 00, Month 200x, 1–16

RESEARCH ARTICLE

Monte Carlo likelihood inference for marked doubly stochastic Poisson processes with intensity driven by marked point processes

M. Minozzo * and S. Centanni

Department of Economics, University of Verona,
Via dell'Artigliere 19, 37129 Verona, Italy.

(*March 2012*)

In recent years, marked point processes have found a natural application in the modeling of ultra-high-frequency financial data since they do not require the integration of the data which is usually needed by other modeling approaches. Among these processes, two main classes have so far been proposed to model financial data: the class of autoregressive conditional duration models of Engle and Russel and the broad class of doubly stochastic Poisson processes with marks. In this paper we consider a class of marked doubly stochastic Poisson processes in which the intensity is driven by another marked point process. In particular, we focus on an intensity with a shot noise form that can be interpreted in terms of the effect caused by news arriving on the market. For models in this class we study likelihood inferential procedures such as Monte Carlo likelihood and importance sampling Monte Carlo expectation maximization by making use of reversible jump Markov chain Monte Carlo algorithms.

Keywords: Cox process; Marked point process; Reversible jump Markov chain Monte Carlo; Shot noise process; Ultra-high-frequency data.

AMS Subject Classification: 62F99; 62M99; 65C05; 91B28

1. Introduction

In ultra-high-frequency (UHF) databases, for each market event, such as a trade or a quote update by a market maker, the time T_i at which it took place, together with some relevant information Z_i, which can be, for instance, the volume, the price or the logreturn of the transaction, are recorded [1]. Such UHF financial data can naturally be modeled as a marked point process (MPP) which can be seen as a sequence of random times $T_1, \ldots, T_N$ where each time is complemented with a random variable, the mark, $Z_1, \ldots, Z_N$ (in general we might have a random vector associated to each time) taking value in some measurable space. Two main classes of models based on MPPs have so far been proposed for these data: the class of autoregressive conditional duration (ACD) models [2] and the class based on doubly stochastic Poisson processes (DSPPs) with marks [3]. As far as these latter processes are concerned, they are characterized by having the number of events in any given time interval as being Poisson distributed, conditionally to another positive stochastic process λ called intensity [4–6]. Moreover, given λ, the number of events in disjoint time intervals are conditionally independent.

*Corresponding author. Email: marco.minozzo@univr.it

ISSN: 0094-9655 print/ISSN 1563-5163 online

DOI: 10.1080/0094965YYxxxxxxxx
http://www.informaworld.com

In this paper we consider a class of marked DSPPs in which the intensity process is a known function of another non-explosive MPP with positive jumps that we characterize through the distributions of jump times and sizes [7, 8]. In particular, we will study an intensity jump process with a drift given by an exponential decay, whose times and sizes of jumps have a probability distribution which can be expressed analytically. Such an intensity process can be viewed as a generalization of the classical shot noise process [5]. Marked DSPPs with shot noise intensity have found applications in many fields such as quantum electronics [9], insurance [10] and finance [3, 11]. In a financial context, the intensity process can be specified to account for intraday seasonalities without the need to resort to ad hoc methods, and a shot noise form, in particular, can be used to catch the effect of economic news on the market.

A central problem faced when modeling with these processes, which are usually only partially observed since the latent intensity is typically not observable, is the estimation of the parameters. For the model currently present in the finantial literature, such as, among others, the model by [12] in which the intensity depends on the level of an unobserved Markovian state process, or the model by [3] where the unknown intensity is a non-Gaussian Ornstein-Uhlenbeck process, the estimation of the parameters is carried out by appealing to standard Bayesian approaches or simple ad hoc least squares techniques.

On the other hand, in this article, we propose some likelihood based inferential procedures which make use of simulation methods based on the Monte Carlo smoothing of the unobservable intensity by means of the reversible jump Markov chain Monte Carlo (RJMCMC) algorithm [13]. Since the (marginal) likelihood function based solely on the observation of a trajectory of the process $(T_i, Z_i)_{i=1,2,\ldots,N}$ is not available in closed form and since the integration of the complete data likelihood is not analytically feasible, neither direct maximization nor the expectation maximization (EM) algorithm [14] are practically possible. Notwithstanding, we can resort to Monte Carlo inferential procedures, such as Monte Carlo likelihood [15] or to Monte Carlo expectation maximization (MCEM) algorithms [16–18]. These procedures are stochastic versions of the respective (non-stochastic) procedures in which the analytic integration required in the computation of the marginal likelihood and in the expectation step of the EM algorithm, respectively, is replaced by a Monte Carlo conditional simulation of the unobserved intensity, given the observed trajectory.

Recently, a similar computational approach has been considered by [19] for their stochastic conditional intensity process (basically, a multivariate dynamic extension of a marked DSPP), which perform statistical inference through Monte Carlo likelihood by using an efficient importance sampling technique. Also, parametric estimation and goodness of fit testing based on Monte Carlo likelihood have been considered by [15] for some spatial parametric point processes. On the other hand, in reviewing inferential methods for DSPPs, [20], although discussing Monte Carlo likelihood for some specific parametric DSPPs, prefer, for computational reasons, to focus on minimum contrast estimation methods. Besides, a well known account of nonparametric inference for general point processes, and in particular for DSPPs, is provided by [21].

The paper is organized as follows. In Section 2 we introduce our modeling framework based on DSPPs with intensity driven by MPPs. Then we consider Monte Carlo likelihood computation in Section 3 and a computational improvement of the MCEM based on importance sampling in Section 4. Lastly, in Section 5 we present some simulation studies and in Section 6 we give some conclusions.

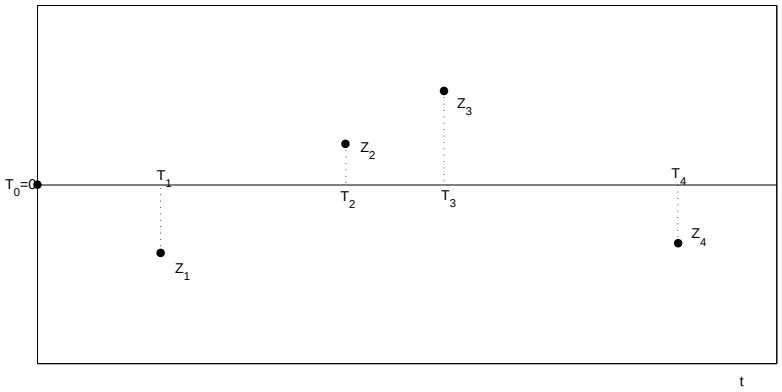

Figure 1. The beginning of a typical trajectory of a marked point process $\phi = (T_i, Z_i)_{i=0,1,2\ldots}$, with $T_0 = 0$ and $Z_0 = 0$. In our context, T_i and Z_i could represent the time and size of the ith logreturn change of some asset of interest.

2. DSPPs with intensity driven by MPPs

In a financial context in which we are interested in the modeling of UHF data, let us denote with T_i and Z_i, $i = 1, 2, \ldots$, the time and the mark, respectively, of the ith event. In our context, T_i and Z_i could be, for instance, the time and size of the ith logreturn change, that is, $Z_i = \ln(S_{T_i}/S_{T_{i-1}})$, where S_t is the price in t of some asset of interest. Then, for a filtered probability space $(\Omega, \mathcal{F}, \mathrm{P}, \{\mathcal{F}_t\}_{t\in[0,T]})$, we assume $\phi = (T_i, Z_i)_{i=0,1,2,\ldots}$, with $T_0 = 0$ and $Z_0 = 0$, to be an MPP, and in particular a marked DSPP with intensity process λ (Figure 1 shows a simulated partial realization of an MPP $(T_i, Z_i)_{i=0,1,2,\ldots}$). For our purposes, we just need to recall that DSPPs are point processes in which, given a realization of the (positive) stochastic process λ, the number of jumps $N_t - N_s$ in any time interval $(s, t]$ is Poisson distributed, that is, for $k = 0, 1, 2, \ldots$,

$$\mathrm{P}\big(N_t - N_s = k\big|\lambda\big) = \frac{\Big(\int_s^t \lambda_u \mathrm{d}u\Big)^k}{k!} \exp\Big\{ - \int_s^t \lambda_u \mathrm{d}u\Big\} ,$$

and that, given λ, the number of jumps in disjoint time intervals are independent.

In what follows, we will also assume that the intensity λ is driven by an MPP, that is, that λ depends itself on some other MPP. In other words, we assume that λ is of the form $\lambda_t = h(t, \psi_0^t)$, where h is a given function and ψ_0^t is the restriction to $[0, t]$ of an MPP $\psi = (\tau_j, X_j)_{j=0,1,2,\ldots}$, with $\tau_0 = 0$ and $X_j > 0$. For this latter MPP, we assume that the conditional distributions $p(\tau_j|\tau_1, \ldots, \tau_{j-1})$ and $p(X_j|\tau_1, \ldots, \tau_j, X_1, \ldots, X_{j-1})$ admit density. Of considerable interest is the case in which the intensity λ of the marked DSPP ϕ has a shot noise form, that is, when

$$\lambda_t = \sum_{j=0}^{M_t} X_j \; e^{-k(t-\tau_j)} \; , \tag{1}$$

where $M_t = \#\{j \in \mathbb{N} : \tau_j \leq t\}$.

Example (basic model). In the modeling framework just outlined, a particularly simple specification is given by the basic model described by Centanni and Minozzo [7, 8] in which the marks Z_i are assumed independently distributed and the parameters of the process λ, that is assumed to have the form in (1), are the mean interarrival time ν and the mean size of 'market perturbations' $1/\gamma$. In detail, for

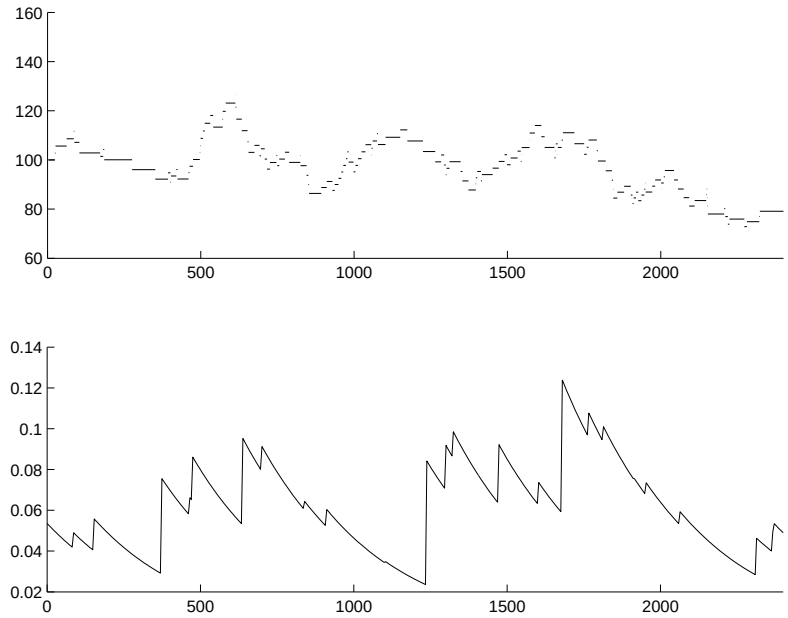

Figure 2. Simulated trajectory (top) of the price process $S_t = S_0 \exp\{\sum_{i=1}^{N_t} Z_i\}$, with starting value $S_0 = 100$ and time horizon $T = 2400$, where $N_t = \#\{i \in \mathbb{N} : T_i \leq t\}$, obtained by conditioning to a simulated trajectory of the intensity process λ (bottom), assuming the basic model with $\nu = 1/150$, $k = 0.0030$ and $\gamma = 45.89$ (here the price is assumed constant until the next event).

the basic model hold the following assumptions:

(A1) M is a Poisson process with constant intensity ν;
(A2) X_j, $j = 1, 2, \ldots$, are independently and identically distributed Exponential random variables with mean $1/\gamma$ (in particular, independent from τ_j, $j = 1, 2, \ldots$);
(A3) X_0 has a Gamma distribution with parameters ν/k and γ (that is, with mean $\mathrm{E}[X_0] = \nu/(k\gamma)$);
(A4) Z_i, $i = 1, 2, \ldots$, are independently and identically distributed random variables (which are, in particular, independent from the processes N and λ).

Due to the simplicity of these assumptions, this model has some nice probabilistic properties [7, 8]. Figure 2 shows a simulated realization of the intensity and of the 'price process' under the basic model.

Example (variation of a model by Barndorff-Nielsen and Shephard). A more complicated model, always belonging to our modeling framework, is the following variation of a model by [22] in which the intensity process is given by λ in (1), and where

(B1) τ_1 is Exponentially distributed with mean δ_1; $(\tau_j - \tau_{j-1})$, $j = 2, 3, \ldots$, are conditionally distributed, given X_{j-1}, as Exponentials with means δ/X_{j-1};
(B2) $(X_0 - 1)$ is Poisson distributed with mean η_0; $(X_j - 1)$, $j = 1, 2, \ldots$, are Poisson distributed with mean η;

and

(B3) $(T_i)_{i\in\mathbb{N}}$ are the event times of a marked DSPP with intensity λ;
(B4) $(Z_i)_{i\in\mathbb{N}}$ are conditionally independently distributed as Gaussians with mean $\mu(T_i - T_{i-1}) + \alpha \int_{T_{i-1}}^{T_i} \lambda_t \, \mathrm{d}t$, and variance σ^2.

With respect to other modeling approaches, our modeling framework has some

advantages as well as some nice features. First of all, data do not need to be sampled at fixed frequency, indeed the interarrival time between price changes gives us important information that is accounted for in the model. Secondly, our modeling framework can be useful in measuring the impact of economic news on asset prices and volatility. This problem has been widely studied in the literature. For example, [23] using a daily data set of the S&P 500 index, find that stock prices react to unexpected macroeconomic news. On the other hand, [24] use data obtained by sampling every 5 minutes to investigate the effect of economic news on foreign exchange rates. These last authors clearly show that macroeconomic news can be linked to periods of high volatility, although they find that these news have little lasting effects. Very recently, [25] models the 2 minute frequency prices of the future on the S&P 500 index with a diffusion with piecewise constant volatility and gives a procedure for estimating the times at which the volatility changes. Even in this work, a link between economic news and prices is exploited. These and other works, however, avoid the problem of modeling UHF data, so loosing the information incorporated in the times of market events. In our modeling framework, the form assumed by the intensity process, in particular when λ has a shot noise form, allows a natural interpretation for the stochastic changes of the intensity in terms of market perturbations caused by the arrival of relevant news [26]. When the jth item of news reaches the market, a sudden increase X_j in trading activity occurs, depending on the importance of the item, followed by a progressive normalization. The random variable τ_j represents the time of the arrival of the jth item of news whereas the parameter k expresses the speed of absorption of the effect of the news by the market. Another nice feature of our modeling approach is the possibility to account for intraday seasonal patterns through the specification of the model, without the need to resort to ad hoc methods. Indeed, we can incorporate in the intensity λ given by (1) some deterministic addendum accounting for the activity that the market would have in absence of random perturbations. By adequately choosing this addendum, it is possible to model many of the stylized features characterizing intraday price data [1].

3. Monte Carlo Likelihood

In a partial information setting only the pairs (T_i, Z_i) of times and marks of the DSPP ϕ (the 'logreturn' process) observed over a time period $[0, T]$ are available, whereas the intensity λ and the process ψ are not observable. With this information, assuming that the distribution of Z_i, X_j and τ_j depends on a parameter vector $\boldsymbol{\theta}$ taking values in some parameter space $\boldsymbol{\Theta}$, maximum likelihood inference on $\boldsymbol{\theta}$ would require the maximization of the likelihood based on the marginal distribution of the pairs (T_i, Z_i). In our modeling framework, this likelihood does not have, in general, a closed form, and direct analytical or numerical maximization is impossible. In similar situations, it is often possible to perform maximum likelihood inference by using the EM algorithm [14], casting the problem in a missing data setting. In our case, however, the integration required in the expectation step of the EM algorithm is analytically unfeasible, and the implementation of this algorithm is not possible. Nevertheless, likelihood inference on the parameter $\boldsymbol{\theta}$ can be carried out by Monte Carlo likelihood methods [15, 27–29] or by appealing to some stochastic versions of the EM algorithm like the stochastic EM (StEM) algorithm [16, 30, 31] or the MCEM algorithm [18], in which the analytic integration required in the expectation step of the EM algorithm is replaced by a Monte Carlo approximation based on a (not necessarily independent) random sample from the conditional distribution of the unobserved processes given the data.

Before discussing the implementation of the above Monte Carlo likelihood inferential procedures for our DSPPs with intensity driven by MPPs, let us exploit the structure of our modeling framework by identifying process trajectories with vectors (of possibly different length). In particular, with respect to an observation period with a fixed time horizon T, let us identify trajectories in $[0, T]$ of the DSPP ϕ with vectors of variable dimension $1 + 2(N_T + 1)$,

$$\boldsymbol{\phi}_0^T = (N_T, T_0, \mathbf{T}', Z_0, \mathbf{Z}')' \,,$$

where $\mathbf{T} = (T_1, \ldots, T_{N_T})'$, $\mathbf{Z} = (Z_1, \ldots, Z_{N_T})'$, $N_T = \#\{i \in \mathbb{N} : T_i \leq T\}$ and $(\,\cdot\,)'$ denotes transpose. Proceeding similarly for the trajectories of the MPP ψ, we consider the vectors of variable dimension $1 + 2(M_T + 1)$,

$$\boldsymbol{\psi}_0^T = (M_T, \tau_0, \boldsymbol{\tau}', \mathbf{X}')' \,,$$

where $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_{M_T})'$ and $\mathbf{X} = (X_0, X_1, \ldots, X_{M_T})'$ are the times and marks, from 0 to T, of the MPP ψ driving the intensity λ, and where $M_T = \#\{j \in \mathbb{N} : \tau_j \leq T\}$. Each of the vectors $\boldsymbol{\phi}_0^T$ and $\boldsymbol{\psi}_0^T$ belongs to a subset of the space

$$\mathcal{C} = \bigcup_{k=1,2,\ldots} \mathcal{C}_k = \bigcup_{k=1,2,\ldots} \left[\{k\} \times \mathbb{R}^{2(k+1)} \right] ,$$

whereas the vector $((\boldsymbol{\phi}_0^T)', (\boldsymbol{\psi}_0^T)')'$ belongs to the space $\mathcal{C} \times \mathcal{C}$. On this space we assume that there exists a dominating probability measure with respect to which we will consider all densities of interest.

With this identification of process trajectories, the complete likelihood of the observed data $\boldsymbol{\phi}_0^T$ and of the unobserved $\boldsymbol{\psi}_0^T$ can be written as

$$\begin{aligned} f(\boldsymbol{\phi}_0^T, \boldsymbol{\psi}_0^T; \boldsymbol{\theta}) &= f(\mathbf{Z}, \mathbf{T}, N_T, \mathbf{X}, \boldsymbol{\tau}, M_T; \boldsymbol{\theta}) \\ &= f(\mathbf{Z}|\mathbf{T}, N_T, \mathbf{X}, \boldsymbol{\tau}, M_T; \boldsymbol{\theta})\ f(\mathbf{T}, N_T|\mathbf{X}, \boldsymbol{\tau}, M_T; \boldsymbol{\theta})\ f(\mathbf{X}, \boldsymbol{\tau}, M_T; \boldsymbol{\theta}) \,, \end{aligned} \tag{2}$$

where $f(\mathbf{T}, N_T|\mathbf{X}, \boldsymbol{\tau}, M_T; \boldsymbol{\theta}) = f(\mathbf{T}|\mathbf{X}, \boldsymbol{\tau}, M_T; \boldsymbol{\theta})\ \mathrm{P}(T_{N_T+1} > T|\mathbf{T}, \mathbf{X}, \boldsymbol{\tau}, M_T; \boldsymbol{\theta})$.

Example [basic model (continued)]. Let us consider the basic model with parameter vector $\boldsymbol{\theta} = (\gamma, \nu)'$ (k is assumed to be known). In this case the complete data likelihood (2) is given by $L(\gamma, \nu) = L_A \cdot L_B \cdot L_C(\gamma, \nu)$, where L_A and L_B do not depend on $\boldsymbol{\theta}$ and are given by

$$L_A = f(\mathbf{Z}|\mathbf{T}, N_T, \mathbf{X}, \boldsymbol{\tau}, M_T) = f(\mathbf{Z}|N_T),$$

and

$$\begin{aligned}
L_B &= f(\mathbf{T}, N_T|\mathbf{X}, \boldsymbol{\tau}, M_T) \\
&= f(\mathbf{T}|\mathbf{X}, \boldsymbol{\tau}, M_T)\ \mathrm{P}(T_{N_T+1}>T|\mathbf{T}, \mathbf{X}, \boldsymbol{\tau}, M_T) \\
&= \Big[\prod_{i=1}^{N_T} f(T_i|T_{i-1}, \mathbf{X}, \boldsymbol{\tau}, M_T)\Big] \cdot \mathrm{P}(T_{N_T+1} > T|T_{N_T}, \mathbf{X}, \boldsymbol{\tau}, M_T) \\
&= \Big[\prod_{i=1}^{N_T} \lambda_{T_i}\ \exp\Big\{-\int_{T_{i-1}}^{T_i} \lambda_t\ \mathrm{d}t\Big\}\Big] \exp\Big\{-\int_{T_{N_T}}^{T} \lambda_t\ \mathrm{d}t\Big\} \\
&= \Big[\prod_{i=1}^{N_T} \lambda_{T_i}\Big] \cdot \exp\Big\{-\int_0^T \lambda_t\ \mathrm{d}t\Big\} \\
&= \Big[\prod_{i=1}^{N_T} \sum_{j=0}^{M_{T_i}} X_j\ \exp\{-k(T_i-\tau_j)\}\Big] \\
&\quad \cdot \exp\Big\{-\frac{1}{k}\sum_{j=0}^{M_T} X_j(1-\exp(-k(T-\tau_j)))\Big\},
\end{aligned}$$

considering that λ_t is given by (1) and

$$\int_0^t \lambda_s\ \mathrm{d}s = \frac{1}{k}\sum_{j=0}^{M_t} X_j(1-\exp\{-k(t-\tau_j)\}).$$

The only part of the complete likelihood which depends on $\boldsymbol{\theta}$ is

$$\begin{aligned}
L_C(\gamma,\nu) &= f(\mathbf{X}, \boldsymbol{\tau}, M_T; \gamma, \nu) = f(\mathbf{X}|\boldsymbol{\tau}, M_T; \gamma, \nu)\ f(\boldsymbol{\tau}, M_T; \nu) \\
&= f(\mathbf{X}|M_T; \gamma, \nu)\ f(\boldsymbol{\tau}, M_T; \nu) \\
&= \Big[f(x_0)\cdot\prod_{j=1}^{M_T} f(x_j)\Big] \cdot \Big[\prod_{j=1}^{M_T} f(\tau_j|\tau_{j-1})\Big] \cdot \mathrm{P}(\tau_{M_T+1} > T|\tau_{M_T}) \\
&= \Big[\gamma^{\nu/k} x_0^{\nu/k-1} \exp\{-\gamma x_0\}\frac{1}{\Gamma(\nu/k)}\Big] \\
&\quad \cdot \Big[\prod_{j=1}^{M_T} \gamma\ \exp\{-\gamma x_j\}\ \nu\ \exp\{-\nu(\tau_j-\tau_{j-1})\}\Big] \cdot \exp\{-\nu(T-\tau_{M_T})\}.
\end{aligned}$$

Example [variation of a model by Barndorff-Nielsen and Shephard (continued)]. Under assumptions B1, B2, B3 and B4 of the variation of the model by Barndorff-Nielsen and Shephard, and assuming as parameter vector $\boldsymbol{\theta} = (\delta_1, \delta, \eta_0, \eta, \mu, \alpha, \sigma^2)'$ (as in the basic model we assume that the function $h(t, \psi_0^t)$ is known, that is, that k is known), the complete data likelihood (2) is given by

$$L(\delta_1, \delta, \eta_0, \eta, \mu, \alpha, \sigma^2) = L_A(\mu, \alpha, \sigma^2) \cdot L_B \cdot L_C(\delta_1, \delta, \eta_0, \eta)\ ,$$

where $L_B = f(\mathbf{T}, N_T|\mathbf{X}, \boldsymbol{\tau}, M_T)$ does not depend on $\boldsymbol{\theta}$ and is as in the basic model,

and the other two terms on the right hand side are given by:

$$L_A(\mu,\alpha,\sigma^2) = f(\mathbf{Z}|\mathbf{T},N_T,\mathbf{X},\boldsymbol{\tau},M_T;\mu,\alpha,\sigma^2) = \prod_{i=1}^{N_T} f(Z_i|T_i,T_{i-1})$$
$$= \prod_{i=1}^{N_T}\left[\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}\Big(Z_i-\mu(T_i-T_{i-1})-\alpha\int_{T_{i-1}}^{T_i}\lambda_t\,\mathrm{d}t\Big)^2\right\}\right];$$

and

$$L_C(\delta_1,\delta,\eta_0,\eta) = f(\mathbf{X},\boldsymbol{\tau},M_T;\delta_1,\delta,\eta_0,\eta)$$
$$= f(x_0)\cdot\left(\prod_{j=1}^{M_T} f(x_j)\right)\cdot f(\tau_1)\cdot\left(\prod_{j=2}^{M_T} f(\tau_j|\tau_{j-1},x_{j-1})\right)\cdot \mathrm{P}(\tau_{M_T+1}>T|\tau_{M_T})$$
$$= \frac{e^{-\eta_0}\,\eta_0^{(x_0-1)}}{(x_0-1)!}\cdot\left[\prod_{j=1}^{M_T}\frac{e^{-\eta}\,\eta^{(x_j-1)}}{(x_j-1)!}\right]\cdot\frac{1}{\delta_1}\exp\Big\{-\frac{1}{\delta_1}\tau_1\Big\}$$
$$\cdot\left[\prod_{j=2}^{M_T}\frac{x_{j-1}}{\delta}\exp\Big\{-\frac{x_{j-1}}{\delta}(\tau_j-\tau_{j-1})\Big\}\right]\cdot\exp\Big\{-\frac{x_{M_T}}{\delta}(T-\tau_{M_T})\Big\}.$$

Now, to carry out likelihood inference on the parameter $\boldsymbol{\theta}$ we would have to obtain the marginal likelihood based only on the data $\boldsymbol{\phi}_0^T$ by integrating the complete data likelihood (2) with respect to the unobserved $\boldsymbol{\psi}_0^T$ which is, in general, an impossible task. On the other hand, this marginal likelihood can effectively be simulated. For models in our class, and for a fixed parameter point $\tilde{\boldsymbol{\theta}}$, the log-likelihood (in the parameter $\boldsymbol{\theta}$) based on the marginal distribution of the observed trajectory, from 0 to T, of the DSPP ϕ can be written as

$$l(\boldsymbol{\theta}) = \ln\big\{f(\boldsymbol{\phi}_0^T;\boldsymbol{\theta})\big\} = \ln\left\{\int f(\boldsymbol{\phi}_0^T,\boldsymbol{\psi}_0^T;\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\psi}_0^T\right\} \tag{3}$$
$$= \ln\left\{\int \frac{f(\boldsymbol{\phi}_0^T,\boldsymbol{\psi}_0^T;\boldsymbol{\theta})}{f(\boldsymbol{\phi}_0^T,\boldsymbol{\psi}_0^T;\tilde{\boldsymbol{\theta}})}\,f(\boldsymbol{\phi}_0^T,\boldsymbol{\psi}_0^T;\tilde{\boldsymbol{\theta}})\,\mathrm{d}\boldsymbol{\psi}_0^T\right\}$$
$$= \ln\left\{\int \frac{f(\boldsymbol{\phi}_0^T|\boldsymbol{\psi}_0^T;\boldsymbol{\theta})\,f(\boldsymbol{\psi}_0^T;\boldsymbol{\theta})}{f(\boldsymbol{\phi}_0^T|\boldsymbol{\psi}_0^T;\tilde{\boldsymbol{\theta}})\,f(\boldsymbol{\psi}_0^T;\tilde{\boldsymbol{\theta}})}\,f(\boldsymbol{\psi}_0^T|\boldsymbol{\phi}_0^T;\tilde{\boldsymbol{\theta}})\,f(\boldsymbol{\phi}_0^T;\tilde{\boldsymbol{\theta}})\,\mathrm{d}\boldsymbol{\psi}_0^T\right\}$$
$$= \ln\big\{f(\boldsymbol{\phi}_0^T;\tilde{\boldsymbol{\theta}})\big\} + \ln\left\{\int \frac{f(\boldsymbol{\phi}_0^T|\boldsymbol{\psi}_0^T;\boldsymbol{\theta})\,f(\boldsymbol{\psi}_0^T;\boldsymbol{\theta})}{f(\boldsymbol{\phi}_0^T|\boldsymbol{\psi}_0^T;\tilde{\boldsymbol{\theta}})\,f(\boldsymbol{\psi}_0^T;\tilde{\boldsymbol{\theta}})}\,f(\boldsymbol{\psi}_0^T|\boldsymbol{\phi}_0^T;\tilde{\boldsymbol{\theta}})\,\mathrm{d}\boldsymbol{\psi}_0^T\right\},$$

and the log-likelihood ratio against the fixed parameter point $\tilde{\boldsymbol{\theta}}$ is given by

$$\ln\left\{\frac{f(\boldsymbol{\phi}_0^T;\boldsymbol{\theta})}{f(\boldsymbol{\phi}_0^T;\tilde{\boldsymbol{\theta}})}\right\} = \ln\left\{\mathrm{E}_{\tilde{\boldsymbol{\theta}}}\left[\frac{f(\boldsymbol{\phi}_0^T|\boldsymbol{\psi}_0^T;\boldsymbol{\theta})\,f(\boldsymbol{\psi}_0^T;\boldsymbol{\theta})}{f(\boldsymbol{\phi}_0^T|\boldsymbol{\psi}_0^T;\tilde{\boldsymbol{\theta}})\,f(\boldsymbol{\psi}_0^T;\tilde{\boldsymbol{\theta}})}\middle|\boldsymbol{\phi}_0^T\right]\right\}, \tag{4}$$

where $\mathrm{E}_{\tilde{\boldsymbol{\theta}}}[\,\cdot\,|\boldsymbol{\phi}_0^T]$ denotes expectation with respect to $f(\boldsymbol{\psi}_0^T|\boldsymbol{\phi}_0^T;\tilde{\boldsymbol{\theta}})$. Thus, assuming that we can simulate a sample $\boldsymbol{\psi}_0^{T(s)}$, $s=1,\ldots,S$, from the conditional distribution

$f(\boldsymbol{\psi}_0^T|\boldsymbol{\phi}_0^T;\tilde{\boldsymbol{\theta}})$, the log-likelihood ratio (4) can naturally be approximated by

$$\ln\left\{\frac{1}{S}\sum_{s=1}^{S}\frac{f(\boldsymbol{\phi}_0^T|\boldsymbol{\psi}_0^{T(s)};\boldsymbol{\theta})\ f(\boldsymbol{\psi}_0^{T(s)};\boldsymbol{\theta})}{f(\boldsymbol{\phi}_0^T|\boldsymbol{\psi}_0^{T(s)};\tilde{\boldsymbol{\theta}})\ f(\boldsymbol{\psi}_0^{T(s)};\tilde{\boldsymbol{\theta}})}\right\}. \tag{5}$$

Of course, the use of this Monte Carlo approximation (as well as of the algorithm in the following section) crucially relies on the existence of an efficient simulation algorithm to sample from $f(\boldsymbol{\psi}_0^T|\boldsymbol{\phi}_0^T;\tilde{\boldsymbol{\theta}})$. In our modeling framework, a (dependent) sampler for this conditional distribution can be constructed by using trans-dimensional Markov chain Monte Carlo (MCMC) algorithms. For a fixed time interval $[0,T]$, we are interested in the conditional distribution (smoothing distribution) of the MPP ψ in $[0,T]$, that is, of $\psi_0^T=(\tau_j,X_j)_{j:\tau_j\le T}$, or of the intensity λ_0^T, given the observation of the DSPP ϕ in $[0,T]$, that is, of $\phi_0^T=(T_i,Z_i)_{i:T_i\le T}$. Since we do not know in advance the number of τ_j arrived before T, to approximate the smoothing distribution, in a given time interval $[0,T]$, of ψ_0^T, given ϕ_0^T, we can not use standard MCMC algorithms, but we can resort to the RJMCMC algorithm of [13]. Details on how to build a sampler giving a Monte Carlo solution to this smoothing problem are given in [8]. This conditional sampler can also be used for predicting future continuations of the observed trajectory of the 'logreturn' process. Let us note that a somehow similar sampler has been studied by [32] to carry out fully Bayesian inference for a class of stochastic volatility models.

As far as the goodness of the above Monte Carlo approximation is concerned, for every fixed $\boldsymbol{\theta}$, the almost sure convergence, as $S\to\infty$, of (5) to (4) is guaranteed by standard ergodic theorems for Markov chains. Let us remark that 'almost sure' here means for almost all sample paths of the Monte Carlo sampler; the data $\boldsymbol{\phi}_0^T$ being considered fixed. Once a simulated log-likelihood has been generated, this can be used to carry out inference on the unknown parameters following standard practice. We could find the maximum likelihood estimate, study and draw profiles, and use it to carry out model selection. In this respect, [27] gives also conditions for the convergence (as the number of conditionally simulated samples grows to infinity) of the maximum likelihood estimate based on the simulated likelihood to the true maximum likelihood estimate.

4. Monte Carlo EM with importance sampling

In addition to the simulation method just described in the previous section, for models belonging to our class, pointwise inference on the unknown parameters, in particular, maximum likelihood inference, can also be performed through the use of MCEM algorithms [33]. Here, we recall for simplicity the MCEM algorithm for our modelling framework. For furhter details see [7, 8]. Considering that the complete data likelihood can be written as in (2), if the guess for $\boldsymbol{\theta}$ at the rth iteration ($r=1,2,\ldots,R$) of the MCEM is $\boldsymbol{\theta}_r$, the algorithm proceeds as follows:

(S step) draw S_r samples $(\mathbf{X},\boldsymbol{\tau},M_T)^{(s)}$, $s=1,\ldots,S_r$, from the conditional distribution $f(\mathbf{X},\boldsymbol{\tau},M_T|\mathbf{Z},\mathbf{T},N_T;\boldsymbol{\theta}_r)$;

(E step) approximate the expectation

$$Q(\boldsymbol{\theta},\boldsymbol{\theta}_r)=\int \ln\{f(\mathbf{Z},\mathbf{T},N_T,\mathbf{X},\boldsymbol{\tau},M_T;\boldsymbol{\theta})\} \cdot f(\mathbf{X},\boldsymbol{\tau},M_T|\mathbf{Z},\mathbf{T},N_T;\boldsymbol{\theta}_r)\ \mathrm{d}\mathbf{X}\ \mathrm{d}\boldsymbol{\tau}\ \mathrm{d}M_T\ ,$$

with the Monte Carlo estimate

$$Q_r(\boldsymbol{\theta}, \boldsymbol{\theta}_r) = \frac{1}{S_r} \sum_{s=1}^{S_r} \ln\{f(\mathbf{Z}, \mathbf{T}, N_T, \mathbf{X}^{(s)}, \boldsymbol{\tau}^{(s)}, M_T^{(s)}; \boldsymbol{\theta})\} \; ;$$

(M step) take as the new guess $\boldsymbol{\theta}_{r+1}$ the value of $\boldsymbol{\theta}$ which maximizes $Q_r(\boldsymbol{\theta}, \boldsymbol{\theta}_r)$.

By choosing specific values for S_r, $r = 1, 2, \ldots, R$, we get particular versions of the algorithm. For $S_r = 1$ we have the StEM algorithm [34]. This algorithm does not converge pointwise and estimates are given by some summary statistics. For S_r fixed and greater than one we have the MCEM algorithm of [18]. If S_r is very large, the algorithm behaves similarly to the (deterministic) EM algorithm, whereas if S_r is increasing with r, we have a 'simulated annealing' MCEM (with $1/S_r$ playing the role of the temperature).

Note that in the above MCEM algorithm, at the S step of each iteration we need to sample from the conditional distribution of the underlying MPP ψ, given the observation of the 'logreturn' process ϕ. The computational burden of these simulations can considerably be reduced by considering a variation of the algorithm that exploits ideas from importance sampling. In this variation, simulations from the above conditional distribution need to be carried out only once. At the subsequent iterations of the MCEM algorithm the simulation in the S step is replaced by a reweighting of (a portion of) the initial sample. A similar importance sampling improvement of the MCEM algorithm has also been studied, in another context, by [35].

Considering that at the rth iteration ($r = 1, 2, \ldots, R$) of the MCEM we start with the guess $\boldsymbol{\theta}_r$, and being $\tilde{\boldsymbol{\theta}}$ a fixed parameter point, the quantity $Q(\boldsymbol{\theta}, \boldsymbol{\theta}_r)$ which is approximated by a Monte Carlo estimate in the E step of the MCEM algorithm can be rewritten as

$$\begin{aligned}
Q(\boldsymbol{\theta}, \boldsymbol{\theta}_r) &= \int \ln\{f(\boldsymbol{\phi}_0^T, \boldsymbol{\psi}_0^T; \boldsymbol{\theta})\} \; f(\boldsymbol{\psi}_0^T | \boldsymbol{\phi}_0^T; \boldsymbol{\theta}_r) \; \mathrm{d}\boldsymbol{\psi}_0^T \\
&= \int \ln\{f(\boldsymbol{\phi}_0^T, \boldsymbol{\psi}_0^T; \boldsymbol{\theta})\} \; f(\boldsymbol{\psi}_0^T | \boldsymbol{\phi}_0^T; \boldsymbol{\theta}_r) \frac{f(\boldsymbol{\phi}_0^T, \boldsymbol{\psi}_0^T; \tilde{\boldsymbol{\theta}})}{f(\boldsymbol{\phi}_0^T, \boldsymbol{\psi}_0^T; \tilde{\boldsymbol{\theta}})} \; \mathrm{d}\boldsymbol{\psi}_0^T \\
&= \int \ln\{f(\boldsymbol{\phi}_0^T, \boldsymbol{\psi}_0^T; \boldsymbol{\theta})\} \; \frac{f(\boldsymbol{\phi}_0^T | \boldsymbol{\psi}_0^T; \boldsymbol{\theta}_r) \; f(\boldsymbol{\psi}_0^T; \boldsymbol{\theta}_r)}{f(\boldsymbol{\phi}_0^T; \boldsymbol{\theta}_r)} \; \frac{f(\boldsymbol{\psi}_0^T | \boldsymbol{\phi}_0^T; \tilde{\boldsymbol{\theta}}) \; f(\boldsymbol{\phi}_0^T; \tilde{\boldsymbol{\theta}})}{f(\boldsymbol{\phi}_0^T | \boldsymbol{\psi}_0^T; \tilde{\boldsymbol{\theta}}) \; f(\boldsymbol{\psi}_0^T; \tilde{\boldsymbol{\theta}})} \; \mathrm{d}\boldsymbol{\psi}_0^T \\
&= \frac{f(\boldsymbol{\phi}_0^T; \tilde{\boldsymbol{\theta}})}{f(\boldsymbol{\phi}_0^T; \boldsymbol{\theta}_r)} \int \ln\{f(\boldsymbol{\phi}_0^T, \boldsymbol{\psi}_0^T; \boldsymbol{\theta})\} \; \frac{f(\boldsymbol{\phi}_0^T | \boldsymbol{\psi}_0^T; \boldsymbol{\theta}_r) \; f(\boldsymbol{\psi}_0^T; \boldsymbol{\theta}_r)}{f(\boldsymbol{\phi}_0^T | \boldsymbol{\psi}_0^T; \tilde{\boldsymbol{\theta}}) \; f(\boldsymbol{\psi}_0^T; \tilde{\boldsymbol{\theta}})} \; f(\boldsymbol{\psi}_0^T | \boldsymbol{\phi}_0^T; \tilde{\boldsymbol{\theta}}) \; \mathrm{d}\boldsymbol{\psi}_0^T,
\end{aligned}$$

where the term out of the integral is a constant that does not depend on $\boldsymbol{\theta}$ and $\boldsymbol{\psi}_0^T$. Then, assuming a sample $\boldsymbol{\psi}_0^{T(s)}$, $s = 1, \ldots, S$, from the conditional distribution $f(\boldsymbol{\psi}_0^T | \boldsymbol{\phi}_0^T; \tilde{\boldsymbol{\theta}})$, the quantity $Q(\boldsymbol{\theta}, \boldsymbol{\theta}_r)$ can be approximated, forgetting constants, by

$$\tilde{Q}_r(\boldsymbol{\theta}, \boldsymbol{\theta}_r) = \frac{1}{S} \sum_{s=1}^{S} \ln\{f(\boldsymbol{\phi}_0^T, \boldsymbol{\psi}_0^{T(s)}; \boldsymbol{\theta})\} \cdot w_s(\boldsymbol{\psi}_0^{T(s)}, \boldsymbol{\theta}_r, \tilde{\boldsymbol{\theta}}) \; ,$$

where the weights $w_s(\boldsymbol{\psi}_0^{T(s)}, \boldsymbol{\theta}_r, \tilde{\boldsymbol{\theta}})$ are given by

$$w_s(\boldsymbol{\psi}_0^{T(s)}, \boldsymbol{\theta}_r, \tilde{\boldsymbol{\theta}}) = \frac{f(\boldsymbol{\phi}_0^T | \boldsymbol{\psi}_0^{T(s)}; \boldsymbol{\theta}_r) \; f(\boldsymbol{\psi}_0^{T(s)}; \boldsymbol{\theta}_r)}{f(\boldsymbol{\phi}_0^T | \boldsymbol{\psi}_0^{T(s)}; \tilde{\boldsymbol{\theta}}) \; f(\boldsymbol{\psi}_0^{T(s)}; \tilde{\boldsymbol{\theta}})} \; . \tag{6}$$

Thus, a computational improvement over the MCEM algorithm of the previous section can be obtained with the following algorithm in which we simulate S_R samples $(\mathbf{X}, \boldsymbol{\tau}, M_T)^{(s)}$, $s = 1, \ldots, S_R$, (typically with S_R large) from the conditional distribution $f(\boldsymbol{\psi}_0^T | \boldsymbol{\phi}_0^T; \tilde{\boldsymbol{\theta}})$, that is, from $f(\mathbf{X}, \boldsymbol{\tau}, M_T | \mathbf{Z}, \mathbf{T}, N_T; \tilde{\boldsymbol{\theta}})$, at the S step of the first iteration, and replace the sampling in the S step of subsequent iterations by a simple reweighting of (an increasing portion of) the initial sample. In particular, if the guess for $\boldsymbol{\theta}$ at the rth iteration ($r = 2, 3, \ldots, R$) is $\boldsymbol{\theta}_r$, the algorithm proceeds as follows:

(S step) calculate the weights $w_s(\boldsymbol{\psi}_0^{T(s)}, \boldsymbol{\theta}_r, \tilde{\boldsymbol{\theta}})$, $s = 1, \ldots, S_r$, given in (6);
(E step) calculate

$$\tilde{Q}_r(\boldsymbol{\theta}, \boldsymbol{\theta}_r) = \frac{1}{S_r} \sum_{s=1}^{S_r} \ln\{f(\boldsymbol{\phi}_0^T, \boldsymbol{\psi}_0^{T(s)}; \boldsymbol{\theta})\} \cdot w_s(\boldsymbol{\psi}_0^{T(s)}, \boldsymbol{\theta}_r, \tilde{\boldsymbol{\theta}}) \; ;$$

(M step) take as the new guess $\boldsymbol{\theta}_{r+1}$ the value of $\boldsymbol{\theta}$ which maximizes $\tilde{Q}_r(\boldsymbol{\theta}, \boldsymbol{\theta}_r)$.

Let us note that here we have assumed that S_r is increasing with r, so that $S_r \leq S_R$, for any $r < R$. With respect to the algorithms of the previous section, the replacement of the sampling with the reweighting task in the S step above allows to save considerable computing time. In particular, since the samples are obtained with an RJMCMC sampler, this avoids to waste computing time with a burn-in period at each iteration of the MCEM algorithm.

5. Simulation Studies

To give an illustration of the performances of the above simulation algorithms when inferring on the parameters of our DSPPs, we run some simulation experiments. We report here the results of some of these experiments performed with MATLAB 6.0.0 under the basic model (running the MATLAB procedures on an Intel(R) Pentium(R) 4 CPU 3.00 GHz 2.99 GHz, 512 MB RAM). As far as Monte Carlo likelihood is concerned, we considered the basic model with parameter vector $\boldsymbol{\theta} = (\gamma, \nu)'$. Figure 3 shows the simulated likelihood surface for γ and ν assuming the basic model with $k = 0.01$ known, relative to a data set simulated under the basic model with $\gamma = 1.5$, $\nu = 0.033$, $k = 0.01$ and time horizon $T = 2400$. For this set of parameters, the expected number of times T_i is $\mathrm{E}[N_T] = \nu T/(k\gamma) = 5333$ and the simulated data set contained roughly 6000 event times T_i. In a way, this gives an indication of the information contained in the data about the unknown parameters, which we will use for comparing different data sets. For the generation of the pseudorandom conditional samples, we have assumed $\tilde{\gamma} = 0.5$ and $\tilde{\nu} = 0.01$. After a burn-in of 5000, we considered other 10000 RJMCMC samples, recording a trajectory every 10, so that $S = 1000$. In the figure, the maximum corresponds to a likelihood of 0.368.

Similarly, Figure 4 shows the simulated likelihood surface for γ and ν assuming the basic model with $k = 0.005$ known, relative to a data set simulated under the basic model with $\gamma = 0.15$, $\nu = 0.0166$, $k = 0.005$ and time horizon $T = 2400$. For this set of parameters, the expected number of times T_i is $\mathrm{E}[N_T] = \nu T/(k\gamma) = 53333$ and the simulated data set contained roughly 60000 event times T_i. Here again we have assumed $\tilde{\gamma} = 0.5$ and $\tilde{\nu} = 0.01$. As before, after a burn-in of 5000, we considered other 10000 RJMCMC samples, recording a trajectory every 10, so that $S = 1000$. In this figure, the maximum of the likelihood corresponds to a value of 89.032.

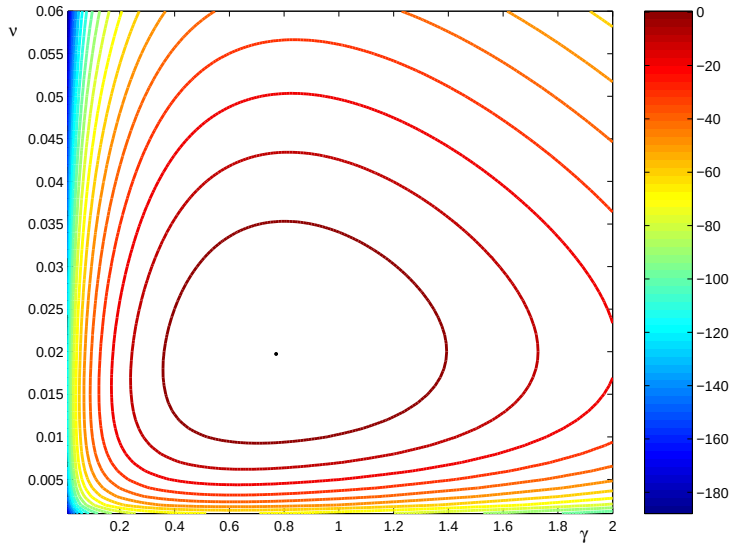

Figure 3. Simulated likelihood surface for γ and ν assuming the basic model with $k = 0.01$ known, relative to a data set simulated under the basic model with $\gamma = 1.5$, $\nu = 0.033$, $k = 0.01$ and time horizon $T = 2400$. The conditional samples have been generated assuming $\tilde{\gamma} = 0.5$ and $\tilde{\nu} = 0.01$.

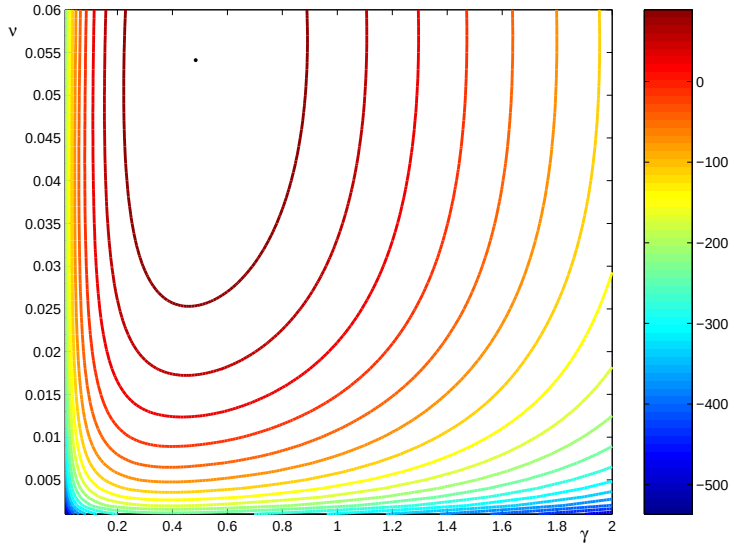

Figure 4. Simulated likelihood surface for γ and ν assuming the basic model with $k = 0.005$ known, relative to a data set simulated under the basic model with $\gamma = 0.15$, $\nu = 0.0166$, $k = 0.005$ and time horizon $T = 2400$. The conditional samples have been generated assuming $\tilde{\gamma} = 0.5$ and $\tilde{\nu} = 0.01$.

To study the behavior of the likelihood when the information in the data is increasing, we considered a time horizon T of 4800 and 9600, in addition to $T = 2400$ considered in the previous simulations. Figure 5 shows the likelihood surface for γ and ν assuming the basic model with $k = 0.01$ known, relative to a data set simulated under the basic model with $\gamma = 1.5$, $\nu = 0.033$, $k = 0.01$ (as in Figure 3) and time horizon $T = 4800$. For this set of parameters, the expected number of times T_i is $\mathrm{E}[N_T] = \nu T/(k\gamma) = 10666$. For the generation of the conditional samples, we have assumed as before $\tilde{\gamma} = 0.5$ and $\tilde{\nu} = 0.01$. After a burn-in of 5000, we considered 10000 RJMCMC samples, recording a trajectory every 10, so that $S = 1000$. In the figure, the maximum corresponds to a value of -0.4252. As we can see, the surface of the likelihood in this case is much steeper than the surface of the likelihood in Figure 3.

With the same settings, but simulating the data with a time horizon $T = 9600$,

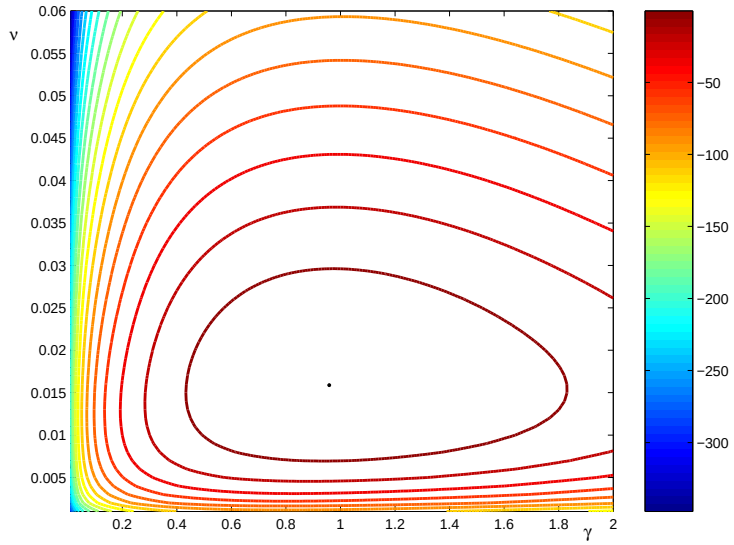

Figure 5. Simulated likelihood surface for γ and ν assuming the basic model with $k = 0.01$ known, relative to a data set simulated under the basic model with $\gamma = 1.5$, $\nu = 0.033$, $k = 0.01$ and time horizon $T = 4800$. The conditional samples have been generated assuming $\tilde{\gamma} = 0.5$ and $\tilde{\nu} = 0.01$.

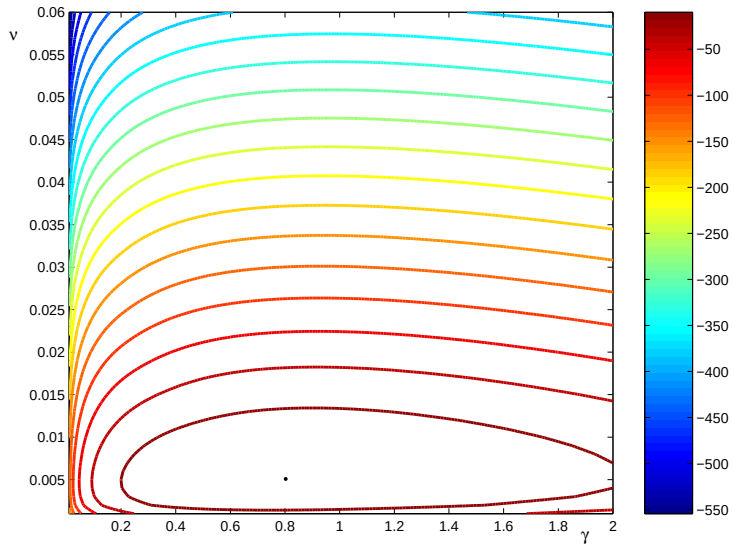

Figure 6. Simulated likelihood surface for γ and ν assuming the basic model with $k = 0.01$ known, relative to a data set simulated under the basic model with $\gamma = 1.5$, $\nu = 0.033$, $k = 0.01$ and time horizon $T = 9600$. The conditional samples have been generated assuming $\tilde{\gamma} = 0.5$ and $\tilde{\nu} = 0.01$.

Figure 6 shows, as we could expect, that in this case the likelihood is even steeper than the likelihood in Figure 5. Let us remark that, as far as computing time is concerned, each of the above four figures took approximately two minutes to be generated.

Once a simulated likelihood surface has been obtained, it can be maximized to obtain approximate maximum likelihood estimates of the parameters. To study the distribution of this approximate maximum likelihood estimator, simulation experiments similar to the following have been carried out. Considering the same parameter values used to generate Figure 3, that is, $k = 0.01$ (treated as a known constant), $\gamma = 1.5$ and $\nu = 0.33$, and a time horizon of $T = 2400$, we simulated 100 samples $\lambda^{(m)}$ and 100 corresponding trajectories of time points $(T_i)^{(m)}_{\{i=1,2,\ldots,N_{2400}\}}$, $m = 1, \ldots, 100$. For each trajectory $(T_i)^{(m)}_{\{i=1,2,\ldots,N_{2400}\}}$, for $m = 1, \ldots, 100$, we computed the approximate maximum likelihood estimates $\hat{\gamma}^{(m)}$ and $\hat{\nu}^{(m)}$ of the param-

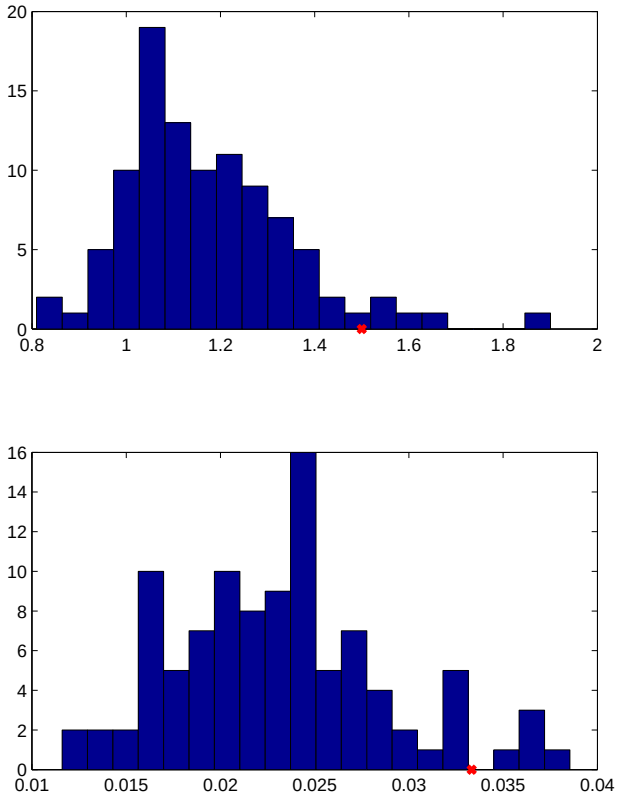

Figure 7. Histograms of the approximated maximum likelihood estimates for γ (top) and ν (bottom) with $k = 0.01$ known, relative to 100 data sets simulated under the basic model with $\gamma = 1.5$, $\nu = 0.033$, and time horizon $T = 2400$. The crosses on the horizontal axis indicate the true parameter values. The conditional samples have been generated assuming $\tilde{\gamma} = 1$ and $\tilde{\nu} = 0.04$.

eters γ and ν, by maximizing the corresponding simulated likelihood surface. The conditional samples used to compute the simulated likelihood surface have been generated assuming $\tilde{\gamma} = 1$ and $\tilde{\nu} = 0.04$. Figure 7 shows the histograms of the approximate maximum likelihood estimates for γ (top) and ν (bottom). The crosses on the horizontal axis indicate the true parameter values. Note that this approximate maximum likelihood estimator is somehow biased. Simulation experiments seem to show that the bias is lower if $\tilde{\gamma}$ and $\tilde{\nu}$ are closer to the true parameter values.

6. Conclusions

In this work we have considered likelihood inference for a class of marked DSPPs with intensity driven by MPPs, that can be used in the modeling of UHF financial data, appealing to simulation methods such as Monte Carlo likelihood and Monte Carlo expectation maximization. Apart from theoretical considerations about the convergence of these methods, we have shown, through simulation experiments, that these methods can actually be implemented for our models. Indeed, with today computing power, the algorithms considered can be implemented to tackle considerably large data sets.

Let us finally remember that in this paper we have not been concerned with

any consideration about asymptotic properties of the estimators considered, like consistency or asymptotic normality, as the 'size' of the data increases to infinity, that is, as the time horizon T goes to infinity. The simulation procedures above are guaranteed just to approximate, as the size of the Monte Carlo sample goes to infinity, the exact (and unfeasible) likelihood procedures relatively to a given set of observations gathered over a finite and fixed time horizon T. Nevertheless, for some of these Monte Carlo estimation procedures we have investigated their sampling properties for T fixed (and finite) through some simulation experiments and found that they resulted reasonably satisfactory.

References

[1] D.M. Guillaume, M.M. Dacorogna, R.R. Davé, U.A. Müller, R.B. Olsen and O.V. Pictet, *From the bird's eye to the microscope: a survey of new stylized facts of the intra-daily foreign exchange markets*, Financ. Stoch. 1 (1999), pp 95–129.

[2] R.F. Engle and J.R. Russel, *Autoregressive conditional duration: a new model for irregularly spaced transaction data*, Econometrica 66 (1998), pp 1127–1162.

[3] T. H. Rydberg and N. Shephard, *A modelling framework for the prices and times of trades made on the New York stock exchange*, in *Nonlinear and Nonstationary Signal Processing*, W.J. Fitzgerald, R.L. Smith, A.T. Walden, P.C. Young, eds., 2000 pp 217–246.

[4] P. Brémaud, *Point Processes and Queues: Martingale Dynamics*, Springer, New York, 1981.

[5] D.R. Cox and V. Isham, *Point Processes*, Chapman and Hall, London, 1980.

[6] G. Last and A. Brandt, *Marked Point Process on the Real Line: The Dynamic Approach*, Springer, New York, 1995.

[7] S. Centanni and M. Minozzo, *Estimation and filtering by reversible jump MCMC for a doubly stochastic Poisson model for ultra-high-frequency financial data*, Stat. Model. 6 (2006), pp 97–118.

[8] S. Centanni and M. Minozzo, *A Monte Carlo approach to filtering for a class of marked doubly stochastic Poisson processes*, J. Am. Stat. Assoc. 101 (2006), pp 1582–1597.

[9] M.C. Teich and B.E.A. Saleh, *Branching processes in quantum electronics* IEEE J. Sel. Top Quant 6 (2000), pp 1450–1457.

[10] A. Dassios and J. Jang, *Pricing of catastrophe reinsurance and derivatives using the Cox process with shot noise intensity*, Financ. Stoch. 7 (2003), pp 73–95.

[11] D. Duffie and N. Garleanu, *Risk and valuation of collateralized debt obligations*, Financ. Anal. J. 57 (2001), pp 41–59.

[12] R. Frey and W.J. Runggaldier, *A nonlinear filtering approach to volatility estimation with a view towards high frequency data*, Int. J. Theor. Appl. Financ. 4 (2001), pp 271–300.

[13] P.J.Green, *Reversible jump Markov chain Monte Carlo computation and Bayesian model determination*, Biometrika 82 (1995), pp 711–732.

[14] A.P. Dempster, N.M. Laird and D.B. Rubin, *Maximum likelihood from incomplete data via the EM algorithm*, J. Roy. Stat. Soc. B 39 1977, pp 1–38.

[15] C.J. Geyer, *Likelihood inference for spatial point processes*, in *Stochastic geometry: likelihood and computation*, O.E. Barndorff-Nielsen, W.S. Kendall and M.N.M. van Lieshout, eds., Chapman and Hall, London, 1999, pp 79–140.

[16] G. Celeux and J. Diebolt, *The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem*, Computation. Stat. Quaterly 2 (1985), pp 73–82.

[17] M. Lavielle and E. Moulines,. *A simulated annealing version of the EM algorithm for non-Gaussian deconvolutions*, Stat. Comput. 7 (1997), pp 229–236.

[18] G.C.G. Wei and M.A. Tanner, *A Monte Carlo implementation of the EM algorithm and poor man's data augmentation algorithm*, J. Am. Stat. Assoc. 85 (1990), pp 699–704.

[19] L. Bauwens and N. Hautsch, *Stochastic conditional intensity processes*, J. Financ. Econom. 4 (2006), pp 450–493.

[20] J. Møller and R.P. Waagepetersen, *Statistical inference for Cox processes*, in *Spatial Cluster Modelling*, D. Denison and A.B. Lawson eds., Chapman and Hall, London, 2002, pp 37–60.

[21] A.F. Karr, *Point Processes and Their Statistical Inference* 2nd ed, Marcel Dekker Inc., New York, 1991.

[22] O.E. Barndorff-Nielsen and N. Shephard, *Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics* (with discussion), J. Roy. Stat. Soc. B 63 (2001), pp 167–241.

[23] D.K. Pierce and V.V. Roley, *Stock prices and economics news*, J. Bus. 58 (1985), pp 49–67.

[24] T. Andersen and T. Bollerslev, *Deutsche Mark-Dollar volatility: intraday activity patterns, macroeconomic announcements, and longer run dependencies*, J. Financ. 53 (1998), pp 219–265.

[25] R.T. Peters, *Shocks in the S&P 500*. Mimeo, University of Amsterdam, 2006.

[26] P.S. Kalev, W.M. Liu, P.K. Pham and E. Jarnecic, *Public information arrival and volatility of intraday stock returns*. J. Bank. Financ. 28 (2004), pp 1441–1467.

[27] C.J. Geyer, *On the convergence of Monte Carlo maximum likelihood calculations*, J. Roy. Stat. Soc. B 56 (1994), pp 261–274.

[28] C.J. Geyer, *Estimation and optimization of functions*, in *Markov Chain Monte Carlo in Practice*, W.R. Gilks, S. Richardson and D.J. Spiegelhalter eds., Chapman and Hall, London, 1996, pp 241–258.

[29] C.J. Geyer and E.A. Thompson, *Constrained Monte Carlo maximum likelihood for dependent data*, J. Roy. Stat. Soc. B 54 (1992), pp 657–699.

[30] G. Celeux, D. Chauveau and J.Diebolt, *On stochastic versions of the EM algorithm*, Mimeo ,Report N. 2514, Institut National de Recherche en Informatique et en Automatique, Rhône-Alpes, Faculté de Médicine de Grenoble, Dept. de Statistique, Latronche, France, 1995.

[31] G. Celeux, D. Chauveau and J.Diebolt, *Stochastic versions of the EM algorithm: an experimental study in the mixture case*, J. Stat. Comput.Sim. 55 (1996), pp 287–314.

[32] G.O. Roberts, O. Papaspiliopoulos and P. Dellaportas, *Bayesian inference for non-Gaussian Ornstein-Uhlenbeck stochastic volatility processes*, J. Roy. Stat. Soc. B 66 (2004), pp 369–393.

[33] G. Fort and E. Moulines, *Convergence of the Monte Carlo expectation maximization for curved exponential families*, Ann. Stat. 31 (2003), pp 1220–1259.

[34] S.F. Nielsen, *The stochastic EM algorithm: estimation and asymptotic results*, Bernoulli 6 (2000), pp 457–489.

[35] R.A. Levine and G. Casella, *Implementations of the Monte Carlo EM algorithm*, J. Comput. Graph. Stat. 10 (2001), pp 422–439.