A Nonlinear Filtering Approach to Volatility Estimation with a View Towards High Frequency Data

Rüdiger Frey
Swiss Banking Institute
University of Zürich
Plattenstr 14
CH-8032 Zürich
freyr@isb.unizh.ch*

Wolfgang J. Runggaldier
Dipartimento di Matematica
Pura ed Applicata
Universitá di Padova, Via Belzoni 7
I-35131-Padova, Italy
runggal@math.unipd.it

May 24, 2000

Abstract

In this paper we consider a nonlinear filtering approach to the estimation of asset price volatility. We are particularly interested in models which are suitable for high frequency data. In order to describe some of the typical features of high frequency data we consider marked point process models for the asset price dynamics. Both jump-intensity and jump-size distribution of this marked point process depend on a hidden state variable which is closely related to asset price volatility. In our setup volatility estimation can therefore be viewed as nonlinear filtering problem with marked point process observations. We develop efficient recursive methods to compute approximations to the conditional distribution of this state variable using the so-called reference probability approach to nonlinear filtering.

Key Words: Stochastic volatility models, volatility estimation, nonlinear filtering theory, reference probability approach, filter approximations

1 Introduction

The considerable amount of empirical evidence against the Black-Scholes model of geometric Brownian motion for asset price fluctuations has led to the development of more flexible models which are able to cope with some of the empirical deficiencies of Black-Scholes. In particular, there is a growing literature on stochastic volatility models (SV-models), see e.g.[9] for a survey. SV-models are designed to mimic the stochastic nature of asset price volatility. In this class of models the instantaneous volatility is assumed to depend on some latent stochastic process which is not adapted to the noise driving the asset price process. This latent process is often interpreted as rate at which new economic information

*This work was done while the first author was at the Department of Mathematics, ETH Zurich. Financial support from UBS is gratefully acknowledged.

is absorbed by the market (see e.g. [1]) or – in trader's language – as temperature of the market.

In the context of SV-models challenging statistical problems arise: in real markets historical asset price volatility[1] is not directly observable. Hence both the current level of volatility and parameters that govern the evolution of volatility dynamics have to be estimated from observable quantities such as the time series of past asset prices. In the present paper we develop an approach to these estimation problems which is based on nonlinear filtering techniques. We are particulary interested in models which are suitable for high frequency data (tick data), i.e. for data which record every quote or every trade. To model these data adequately we have to depart from the diffusion models routinely used in derivative asset analysis, for the following reason:

It is well-known that on very small time-scales real world asset price processes are not very well described by diffusions. In particular, real asset prices are piecewise constant and jump only at discrete points in time, e.g. in reaction to trades or to significant new information. Diffusion processes to the contrary have continuous trajectories with nonzero quadratic variation. Moreover, the quadratic variation of a diffusion can be approximated arbitrarily well by the sum of the squared increments. Hence in a diffusion model with continuous observations of the asset price the current volatility level can be estimated with arbitrary precision from past price information. This is in sharp contrast to the case of piecewise constant processes, where quadratic variation over very small time intervals is useless for volatility estimation. This shows that sample path properties do matter a lot for volatility estimation. In particular, volatility estimation should not be studied in the context of continuously observable diffusion models.

We therefore propose an alternative model, where asset prices are given by a marked point process. Both jump-intensity and jump-size distribution of this marked point process depend on a hidden state variable which is closely related to asset price volatility. In our setup volatility estimation can therefore be viewed as a (nonlinear) filtering problem with marked point process observations. We extend an approach pioneered by Kushner (see e.g. [14]) and develop efficient recursive methods to compute approximations to the conditional distribution of this state variable. While this leads to some new results in nonlinear filtering theory we view the present paper more as a contribution to financial mathematics: along with [8] and [7] this is one of the few applications of nonlinear filtering techniques in Mathematical Finance. Marked point processes as models for high frequency data in Finance have independently been studied by a number of authors in the recent literature. Here we mention only the papers [17], [18] and [12]. The hedging of derivatives in our model is studied in the companion-paper [10].

2 The Model

In this paper we assume that the asset price S changes only at discrete, random points in time $T_1 < T_2 < \ldots \leq T$, where T represents some final time horizon for all economic

[1] In this paper the term volatility refers always to the local variance of asset prices (the *historical volatility*). Of course *implied volatility*, which is computed from observed market prices of derivatives is – if prices of derivatives are available – observable by definition.

activities. These time points represent instants at which a large trade occurs or at which a market maker updates his quotes in reaction to new information. We shall work directly with the logarithmic asset price $Y_t := \log S_t$, both for convenience and in order to be in line with most of the econometric literature on financial time series. At time T_n the log-price Y_{T_n} equals the value of some *shadow log price process* $L = (L_t)_{0 \le t \le T}$. We model the dynamics of L by a stochastic differential equation (SDE) driven by a Brownian motion $w^{(2)} = \left(w_t^{(2)}\right)_{0 \le t \le T}$. In our setup the diffusion coefficient of this SDE is influenced by an unobservable state variable process $X = (X_t)_{0 \le t \le T}$ which is independent of the filtration generated by $w^{(2)}$.

The random times T_n are modelled as jump-times of some point process $N = (N_t)_{0 \le t \le T}$ whose intensity depends on the level of the state variable process. This appears reasonable from an economic viewpoint, as the rate at which information is absorbed by the market is very likely to have an influence on trading activity.

We assume that the state variable process X is a time-homogeneous Markov process with RCLL trajectories defined on some underlying filtered probability space $(\Omega, \mathcal{F}, P, \{\mathcal{F}_t\})$. We are particularly interested in the following two cases:

A1) (X is a diffusion) Consider a Wiener process $w^{(1)} = (w_t^{(1)})_{0 \le t \le T}$ on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$, which is independent of $w^{(2)}$, and continuous functions α and β: $\mathbb{R} \to \mathbb{R}$. X is a global solution to the SDE

$$dX_t = \alpha(X_t)dt + \beta(X_t)dw_t^{(1)} . \tag{1}$$

Moreover, the coefficients α and β are such that weak uniqueness holds for (1).

A2) X is a continuous-time finite-state Markov chain with state space $E_M := \{x_1, \cdots, x_M\}$ and generator matrix $R = \{r_{ij}\}_{i,j=1,\cdots,M}$.

Remark 2.1 In the subsequent analysis we assume that the dynamics of the state-variable process are known. This is no restriction, as the important case where parameters of the dynamics of X are unknown and have to be estimated from the time series of past prices can be incorporated by modelling the state variable process as $d+1$-dimensional process for some $d > 0$. Suppose for instance that α and β in (1) depend on a d-dimensional parameter vector $\Theta = \theta^1, \ldots, \theta^d$, i.e. $\alpha = \alpha(x; \Theta)$ and $\beta = \beta(x; \Theta)$. This can be modelled by a $d+1$-dimensional state variable process X with known dynamics, if we define X by $X_t := (\tilde{X}_t, \Theta)$, where $\tilde{X}$ solves the SDE $d\tilde{X}_t = \alpha(\tilde{X}_t; \Theta)dt + \beta(\tilde{X}_t; \Theta)dw_t^{(1)}$.

We make the following assumption on the dynamics of the shadow log price process L:

A3) L solves the SDE

$$dL_t = \sqrt{v(t, X_{t-})}dw_t^{(2)} \tag{2}$$

for a Brownian motion $w^{(2)}$ on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$ which is independent of $\{\mathcal{F}_t^X\}$, the filtration generated by the state variable process. The function $v : [0, T] \times \mathbb{R} \to \mathbb{R}$ is continuous; moreover, there are constants $\underline{v}$ and $\overline{v}$ with $0 < \underline{v} \le \overline{v} < \infty$ such that $v(t, x) \in [\underline{v}, \overline{v}]$ for all $t \in [0, T]$ and all $x \in \mathbb{R}$.

Using Itô's formula it is immediately seen that under A3) $\exp(L_t)$ follows a standard stochastic volatility model as discussed for instance in [9]. We now introduce the dynamics of the actual log price process Y. For this purpose let $N = (N_t)_{0 \le t \le T}$ be a point process with $\{\mathcal{F}_t\}$- intensity $\lambda(t, X_{t-})$ on our underlying probability space. We assume throughout the paper that $\lambda : [0,T] \times \mathbb{R} \to \mathbb{R}$ is continuous and takes its values in the interval $[\underline{\lambda}, \overline{\lambda}]$ for constants $\underline{\lambda}$ and $\overline{\lambda}$ with $0 < \underline{\lambda} \le \overline{\lambda} < \infty$; typically λ will be an increasing in the second argument.We allow for explicit time-dependency in the functions v and λ in order to incorporate seasonality effects which are typical for high frequency data. For a discussion of the properties of high frequency data we refer to [13].

Let $(T_n)_{n \in \mathbb{N}}$ be the jump-times of the point process N and note that, by definition, T_{N_t} and $T_{N_{t-}}$ are the times of the last jump prior to t and *strictly prior* to t respectively. We define the log asset price by

$$Y_t := L_{T_{N_t}}. \tag{3}$$

Note that Y is a right continuous process with piecewise constant trajectories; by construction the jump-times of Y and N coincide P-a.s.

Besides $\{\mathcal{F}_t\}$ we shall consider the filtrations $\{\mathcal{F}_t^X\}$, $\{\mathcal{F}_t^Y\}$ generated by the state variable and the observation processes respectively, as well as the filtration $\{\mathcal{G}_t\}$ defined by

$$\mathcal{G}_t = \mathcal{F}_T^X \vee \mathcal{F}_t^Y . \tag{4}$$

Note that X is $\mathcal{G}_0$-measurable by definition, i.e. $\{\mathcal{G}_t\}$ contains information about all the future of the state variable process.

For our filtering results we need an additional assumption on the process N:

A4) Under P the process N is a doubly stochastic Poisson process (Cox process), i.e. it admits the $(P, \{\mathcal{G}_t\})$-intensity $\lambda(t, X_{t-})$ (see also Chapter I of [3]).

Remark 2.2 A4) implies that, given information about all the future of the state variable, N is a Poisson process with conditionally deterministic intensity $\lambda(t, X_{t-})$. This is a stronger requirement than the assumption that N is a point process with $(P, \{\mathcal{F}_t\})$-intensity $\lambda(t, X_{t-})$; in particular, A4) rules out the possibility that N and X have common jumps. Put differently A4) implies that the actual trading activity - the realization of the point process N - does not affect the law of the state-variable. In economic terms this means that in our model trading is caused by exogenous factors such as fundamental information and not by the observed past trading activity. Admittedly it is difficult to judge if this assumption is met in practice; however, we are confident that our model is a useful first approximation to reality.

We now determine the $(P, \{\mathcal{G}_t\})$-jump-size distribution of Y. Define for $t > 0$

$$\Sigma_t := \int_{T_{N_{t-}}}^{t} v(s, X_{s-})ds \text{ and, for } n = 1, 2, \dots , \quad \Sigma_n := \int_{T_{n-1}}^{T_n} v(s, X_{s-})ds . \tag{5}$$

Suppose that the process N has a jump at time t. By (3) Y has a jump of size $\Delta Y_t = L_t - L_{T_{N_{t-}}}$. Equation (2) now implies that conditionally on $\mathcal{G}_t$ we have $\Delta Y_t \sim \Phi(dz; \Sigma_t)$, where

$\Phi(dz;\Sigma_t)$ is shorthand for a normal distribution with mean zero and variance Σ_t. Moreover, the $(P,\{\mathcal{G}_t\})$-jump-intensity of Y is obviously equal to the $(P,\{\mathcal{G}_t\})$-jump-intensity of N and hence given by $\lambda(t,X_{t-})$. Following e.g. [3] we shall express this fact by saying that the marked point process Y has the $(P,\{\mathcal{G}_t\})-$ local characteristics $(\lambda(t,X_{t-}),\Phi(dz;\Sigma_t))$.

3 The filtering problem

Our filtering problem consists of determining, for each $t \leq T$, the conditional expectation $E\{F(X_t)|\mathcal{F}_t^Y\}$ where $F(\cdot)$ is a continuous and bounded function, X_t is the unobserved state variable process, and Y_t is the observed log-price process.

This is a nonlinear filtering problem, for which an exact solution is practically impossible to obtain. We thus determine an approximation to $E\{F(X_t)|\mathcal{F}_t^Y\}$, leading to a weak-sense approximation for the conditional distribution $p_t(X_t|\mathcal{F}_t^Y)$ of X_t, given the observations of Y_s up to and including t. An important practical aspect of our procedure, especially for high frequency data, will be that this approximation can be computed recursively.

Among the two major approaches to nonlinear filtering, namely the so-called innovations approach (see [11], [16]) and the reference probability approach (see [3], [14]), the latter is better suited for our approximations. In the next subsection we shall present its main features.

3.1 The reference probability approach

This approach is based on the fact that, for a probability measure Q on $(\Omega,\mathcal{F},\mathcal{G}_t)$, that is absolutely continuous with respect to P with Radon-Nikodym derivative

$$dQ_{|\mathcal{G}_t}\Big/dP_{|\mathcal{G}_t} := \Lambda_t > 0, \tag{6}$$

we have (see e.g. Lemma L.5 in Ch. VI of [3])

$$E^P\{F(X_t)|\mathcal{F}_t^Y\} = \frac{E^Q\{F(X_t)\Lambda_t^{-1}|\mathcal{F}_t^Y\}}{E^Q\{\Lambda_t^{-1}|\mathcal{F}_t^Y\}} \tag{7}$$

where, to avoid ambiguity, we use now E^P to denote expectation with respect to P. Formula (7) is usually called Kallianpur-Striebel formula and is related to Bayes' formula : Λ_t^{-1} can in fact be interpreted as a version of the likelihood of the state variable $(X_s)_{s\leq t}$, given the observations $(Y_s)_{s\leq t}$. If a version Λ_t^{-1} of the likelihood can be found such that, under the corresponding measure Q, the processes X and Y are independent, then the conditional expectations on the right in (7) reduce to ordinary expectations. If the measure transformation (6) can furthermore be chosen so that the restrictions of P and Q on $\mathcal{F}_T^X$ coincide, these expectations can be evaluated using the original law of X. Reducing the conditional expectation on the left in (7) to two ordinary expectations will be of great help in computing approximate solutions to our filtering problem.

An interesting variant of our model, in which state variable and observation process cannot be made independent under a transformation of measures, is when the state variable process is modelled as a jump process that has common jumps with the observations. Notice that, in this case, assumption A.4 does not hold (see Remark 2.2). For this particular case a specific approximation methodology has been developed in [4].

We shall now construct a new measure $Q \sim P$ on $\mathcal{G}_T$ such that the marked point process Y admits the $(Q, \{\mathcal{G}_t\})$-local characteristics $(1, \Phi(dz; \tau_t))$ where $\Phi(dz; \tau_t)$ stands for a normal distribution with mean zero and variance $\tau_t := \underline{v}(t - T_{N_{t-}})$. It will be shown in Proposition 3.1 below that, under such a measure Q, the observation and state variable processes are independent and that the marginal distributions of X under P and Q coincide. Define the function $h(t, z)$ as quotient of the Lebesgue-densities of $\Phi(dz; \tau_t)$ and $\Phi(dz; \Sigma_t)$, i.e.

$$h(t, z) := \left(\frac{\Sigma_t}{\tau_t} \right)^{\frac{1}{2}} \exp \left\{ -\frac{1}{2} z^2 \frac{\Sigma_t - \tau_t}{\Sigma_t \tau_t} \right\} , \tag{8}$$

and let the process Λ_t be a solution to the integral equation

$$\Lambda_t = 1 + \int_0^t \int_{\mathbb{R}} \Lambda_{s-} \left(\lambda^{-1}(s, X_{s-}) h(s, z) - 1 \right) q^Y(ds \times dz) , \tag{9}$$

where q^Y denotes the $(P, \{\mathcal{G}_t\})$-compensated counting measure associated to Y (see e.g. [3], Chapter VIII). We now have

Proposition 3.1 *Let $Y = (Y_t)_{t \geq 0}$ be a marked point process with $(P, \{\mathcal{G}_t\})-$ local characteristics $(\lambda(t, X_{t-}), \Phi(dz; \Sigma_t))$. Define Λ_t as solution to (9). Then Λ is a strictly positive martingale and $E^P[\Lambda_T] = 1$. Moreover, defining an equivalent measure $Q \sim P$ via $(dQ/dP)_{|\mathcal{G}_t} := \Lambda_t$, we have the following :*

(i) Y admits the $(Q, \{\mathcal{G}_t\})$-local characteristics $(1, \Phi(dz; \tau_t))$.

(ii) The restrictions of P and Q on $\mathcal{F}_T^X$ coincide.

(iii) X and Y are $Q-$independent.

Proof : Λ_t is the semimartingale exponential of the $(P, \{\mathcal{G}_t\})-$local martingale

$$Z_t := \int_0^t \int_{\mathbb{R}} \left(\lambda^{-1}(s, X_{s-}) h(s, z) - 1 \right) q^Y(ds \times dz) . \tag{10}$$

As the jumps of Z are strictly larger than minus one, Λ is a positive local martingale. Since $\Sigma_t / \tau_t < \overline{v} / \underline{v}$ and $\Sigma_t \geq \tau_t$, the argument of the exponential function in the expression (8) of the function h is always non-positive. Moreover, $\lambda(t, X_{t-})$ and $\lambda^{-1}(t, X_{t-})$ are both bounded by our assumptions. Hence all conditions of Theorem T11 in Chapter VIII of [3] are satisfied so that $E^P[\Lambda_T] = 1$ and Λ is a $(P, \{\mathcal{G}_t\})$-martingale. Theorem T10 in Chapter VIII of [3] now implies that Y admits the $(Q, \{\mathcal{G}_t\})$-local characteristics $(1, h(t, z)\Phi(dz; \Sigma_t))$ $= (1, \Phi(dz; \tau_t))$ by definition of h in (8).

To prove (ii) note that $\mathcal{G}_0 = \mathcal{F}_T^X$ and that Λ is a $(P, \{\mathcal{G}_t\})-$martingale. Hence we get for any bounded measurable function $g : \mathbb{R}^n \to \mathbb{R}$ and any $0 \leq t_1 < \cdots < t_n \leq T$

$$\begin{aligned} E^Q \{ g(X_{t_1}, \cdots, X_{t_n}) \} &= E^P \left\{ E^P \{ \Lambda_T g(X_{t_1}, \cdots, X_{t_n}) \mid \mathcal{G}_0 \} \right\} \\ &= E^P \left\{ g(X_{t_1}, \cdots, X_{t_n}) \, E^P \{ \Lambda_T | \, \mathcal{G}_0 \} \right\} \\ &= E^P \{ g(X_{t_1}, \cdots, X_{t_n}) \} , \end{aligned} \tag{11}$$

where it was used that $E^P\{\Lambda_T|\,\mathcal{G}_0\}=1$.

We now turn to (iii). Using again the $\mathcal{G}_0$-measurability of X we get

$$
\begin{aligned}
E^Q\ &\{g(X_{t_1},\cdots,X_{t_n})\,h(Y_{t_1},\cdots,Y_{t_m})\}=E^P\{g(X_{t_1},\cdots,X_{t_n})\,h(Y_{t_1},\cdots,Y_{t_m})\,\Lambda_T\}\\
&=E^P\left\{g(X_{t_1},\cdots,X_{t_n})\,E^P\{h(Y_{t_1},\cdots,Y_{t_m})\,\Lambda_T|\,\mathcal{G}_0\}\right\}. \qquad (12)
\end{aligned}
$$

By (i) the second factor in the rhs of (12)equals $E^Q\{h(Y_{t_1},\cdots,Y_{t_m})\}$. Hence (12) equals

$$
E^P\{g(X_{t_1},\cdots,X_{t_n})\}\,E^Q\{h(Y_{t_1},\cdots,Y_{t_m})\}=E^Q\{g(X_{t_1},\cdots,X_{t_n})\}\,E^Q\{h(Y_{t_1},\cdots,Y_{t_m})\},
$$

and the claim follows.

Q.E.D.

The Q-independence allows us to model X and Y on a product space; this will be convenient in the development of recursive approximations to our filtering problem in the next sections. More precisely, we consider two filtered probability spaces $(\Omega^X,\mathcal{F}^X,\{\mathcal{F}^X_t\},Q^X)$ and $(\Omega^Y,\mathcal{F}^Y,\{\mathcal{F}^Y_t\},Q^Y)$ supporting $\{\mathcal{F}^X_t\}$- and $\{\mathcal{F}^Y_t\}$-adapted processes $(X_t)_{0\le t\le T}$ and $(Y_t)_{0\le t\le T}$ respectively, such that

- X is a Markov process on $(\Omega^X,\mathcal{F}^X,\{\mathcal{F}^X_t\},Q^X)$ having the same law as our state variable process on the original probability space.
- Y is a marked point process on $(\Omega^Y,\mathcal{F}^Y,\{\mathcal{F}^Y_t\},Q^Y)$ with (conditional) deterministic local characteristics $(1,\Phi(dz;\tau_t))$.

Define the product space $\Omega=\Omega^X\times\Omega^Y,\mathcal{F}=\left(\mathcal{F}^X\otimes\mathcal{F}^Y\right),\{\mathcal{F}_t\}=\{\mathcal{F}^X_t\}\otimes\{\mathcal{F}^Y_t\},Q=Q^X\otimes Q^Y$ as well as the filtration $\{\mathcal{G}_t\}$ with $\mathcal{G}_t=\mathcal{F}^X_T\otimes\mathcal{F}^Y_t$ and extend X and Y to processes on Ω in the canonical manner. Then this new model is from a probabilistic viewpoint equivalent to our original model under the measure Q.

Define $L_n:=L_{T_n},\ \tau_n:=\underline{v}(T_n-T_{n-1})$ and recall the definition of Σ_n in (5). By the exponential formula of Lebesgue-Stieltjes calculus, Λ_t is given by

$$
\begin{aligned}
\Lambda_t\ &=\ \textstyle\prod_{T_n\le t}\lambda^{-1}(T_n,X_{T_n-})h(T_n,L_n-L_{n-1})\\
&\qquad\cdot\exp\left\{\int_0^t\int_{\mathbb{R}}\left(1-\lambda^{-1}(s,X_{s-})h(s,z)\right)\lambda(s,X_{s-})\Phi(dz;\Sigma_s)\,ds\right\}\\
&=\ \textstyle\prod_{T_n\le t}\left\{\lambda^{-1}(T_n,X_{T_n-})\dfrac{\sqrt{\Sigma_n}}{\sqrt{\tau_n}}\exp\left[-\dfrac{1}{2}(L_n-L_{n-1})^2\dfrac{\Sigma_n-\tau_n}{\Sigma_n\tau_n}\right]\right\}\\
&\qquad\cdot\exp\left[\textstyle\int_0^t(\lambda(s,X_{s-})-1)\ ds\right]
\end{aligned}
\qquad (13)
$$

Obviously, Λ_t is a function of X^t and Y^t, the trajectories of the processes X and Y up to time t; we write $\Lambda_t=\Lambda_t(X^t,Y^t)$.

By the Kallianpur-Striebel formula (7), our filtering problem is solved once we have determined (approximately) the functional

$$V_t(y;F) := E^Q\left\{F(X_t)\Lambda_t^{-1}(X^t,Y^t)\,|\,Y^t(\omega)=y^t\right\} \tag{14}$$

for any possible trajectory y_s, $0\le s\le T$ of the marked point process Y. Here, $Y^t(\omega)=y^t$ means of course that the trajectory of $Y_s(\omega)$ equals y up to time t. Using the special structure of our product space it is immediate that

$$V_t(y;F) := E^{Q^X}\left\{F(X_t)\Lambda_t^{-1}(X^t,y^t)\right\}\ . \tag{15}$$

3.2 Approximate solution of the filtering problem

For the *approximate* evaluation of the functional $V_t(y;F)$ we use an approach pioneered by H. Kushner (see e.g. [14]) : we approximate the state variable process X by a sequence of discrete-time, finite-state Markov chains $(X_k^n)_{k=0,1,\dots}$, whose piecewise constant interpolations $X_t^n := X^n_{[(\Delta^n)^{-1}t]}$, $0\le t\le T$ converge to X in distribution on the Skorokhod space as $n\to\infty$ and the time intervals $\Delta^n\to 0$; we denote this by $X^n\Rightarrow X$. In this section we show that the convergence $X^n\Rightarrow X$ essentially implies the convergence of the corresponding filters; efficient recursive numerical procedures for evaluating the functional (15), applied to the processes X^n, are developed in the next section. We have

Proposition 3.2 *Consider a sequence X^n of RCLL-processes on $(\Omega^X,\mathcal{F}^X,\{\mathcal{F}_t^X\},Q^X)$ and a bounded and continuous function $F:\mathbb{R}\to\mathbb{R}$. Define, for a typical trajectory y of the marked point process Y and for any $t\in[0,T]$, the sequence of functionals*

$$V_t^n(y;F) := E^{Q^X}\left\{F(X_t^n)\Lambda_t^{-1}((X^n)^t,y^t)\right\}\ . \tag{16}$$

Assume that X has no deterministic times of discontinuity, i.e. $Q^X\{X_t\neq X_{t-}\}=0$ for every $t\in(0,T]$, and that that $X^n\Rightarrow X$. Then

$$V_t^n(y;F)\to V_t(y;F)\qquad as\qquad n\to\infty\qquad for\ every\qquad t\in(0,T].$$

Proof : Fix $t\in(0,T]$ and some typical trajectory y of Y. Denote by $D_{[0,T]}$ the Skorokhod space of all RCLL-functions on $[0,T]$, and by μ^X the law of the limit process X. As X has no deterministic times of discontinuity, the mapping

$$\psi\,:\,D_{[0,T]}\to\mathbb{R}\,,\qquad \psi(X)=F(X_t)\cdot\Lambda_t^{-1}(X^t,y^t)$$

is μ^X-a.s. continuous for every $t\in[0,T]$; moreover, ψ is bounded by the assumptions introduced in Section 2. Hence the convergence $X^n\Rightarrow X$ immediately implies the result (see Theorem 5.5 of [2]).

Q.E.D.

Remark 3.3 The hypotheses of Proposition 3.2 hold in particular for Markov processes satisfying Assumptions A1 or A2. Obviously, X has no deterministic times of discontinuity in these cases. The construction of Markov-chain approximations to processes X satisfying A1 or A2 is nowadays standard; see e.g. the books [6], [15] for details.

4 Recursive computation of the approximating filter

We fix an approximating index n and with it the time discretization step $\Delta = \Delta^n$. For the corresponding discretized Markov chain $X = \{X_k\}_{k=1,2,\cdots}$ we denote by $\{x_1, \cdots, x_M\}$ the set of possible values of the state and by $\Pi = \{p_{ij}\}_{i,j=1,\cdots,M}$ the transition probability matrix. With the partition of the time axis into intervals of sufficiently small length, it suffices to compute the approximating filter V^n only at the discrete time points $k\Delta$.

We shall now derive a recursion for the $M-$vector q_k^y with components

$$q_k^y(x_i) := V_{k\Delta}\left(y; \delta_{\{x_i\}}\right) = E^{Q^X}\left\{1_{\{X_{k\Delta}=x_i\}}\Lambda_{k\Delta}^{-1}(X^{k\Delta}, y^{k\Delta})\right\}; \tag{17}$$

this allows us to obtain

$$V_{k\Delta}(y; F) = \sum_{i=1}^{M} F(x_i)\, q_k^y(x_i) \tag{18}$$

for any continuous and bounded F.

By (18) and the Kallianpur-Striebel formula we have

$$E^P\{F(X_{k\Delta})|\mathcal{F}_{k\Delta}^Y\} = \frac{\sum_{i=1}^M F(x_i)\, q_k^y(x_i)}{\sum_{j=1}^M q_k^y(x_j)} \tag{19}$$

showing that $p_k^y(x_i) := q_k^y(x_i) \Big/ \sum_{j=1}^M q_k^y(x_j)$ is the *filter distribution* for the approximating state variable process $\{X_k\}$. Consequently, $q_k^y(x_i)$ can be seen as a corresponding unnormalized filter distribution.

For the generic time step k and with T_m being the jump times (if there are) of Y in $(k\Delta, (k+1)\Delta]$, define the matrix

$$E^k := diag\left(E_j^k;\ j = 1, \cdots, M\right) \tag{20}$$

where

$$\begin{aligned} E_j^k \ := \ & \exp\left\{\Delta - \int_{k\Delta}^{(k+1)\Delta} \lambda(t, x_j)\, dt + \sum_{k\Delta < T_m \le (k+1)\Delta} \left[\frac{1}{2}\log\left(\frac{\underline{v}(T_m - T_{m-1})}{\int_{T_{m-1}}^{T_m} v(t, x_j)\, dt}\right)\right.\right. \\ & + \left.\left. \frac{1}{2}\frac{(L_m - L_{m-1})^2}{T_m - T_{m-1}} \cdot \frac{\int_{T_{m-1}}^{T_m} v(t, x_j) dt - \underline{v}(T_m - T_{m-1})}{\underline{v}\int_{T_{m-1}}^{T_m} v(t, x_j) dt} + \log\lambda(T_m, x_j)\right]\right\}. \end{aligned} \tag{21}$$

Proposition 4.1 *For the $M-$vector q_k^y of unnormalized conditional expectations as defined in (17) we have the following recursion*

$$q_{k+1}^y = \left(\Pi^T \cdot E^k\right) \cdot q_k^y \tag{22}$$

where Π^T is the transpose of the transition probability matrix of the approximating state variable process and E^k is the diagonal matrix defined in (20), (21).

Proof : From the definition in (17) we have

$$\begin{aligned} q_{k+1}^{y}(x_i) &= E^{Q^X}\left\{1_{\{X_{(k+1)\Delta}=x_i\}}\Lambda^{-1}_{(k+1)\Delta}\left(X^{(k+1)\Delta}, y^{(k+1)\Delta}\right)\right\} \\ &= E^{Q^X}\left\{\Lambda^{-1}_{k\Delta}\left(X^{k\Delta}, y^{k\Delta}\right) E^{Q^X}\left\{\frac{\Lambda^{-1}_{(k+1)\Delta}\left(X^{(k+1)\Delta}, y^{(k+1)\Delta}\right)}{\Lambda^{-1}_{k\Delta}\left(X^{k\Delta}, y^{k\Delta}\right)} 1_{\{X_{(k+1)\Delta}=x_i\}}\middle|\mathcal{F}^X_{k\Delta}\right\}\right\} \\ &= \sum_{j=1}^{M} E^{Q^X}\left\{\Lambda^{-1}_{k\Delta}\left(X^{k\Delta}, y^{k\Delta}\right) 1_{\{X_{k\Delta}=x_j\}}\right. \\ &\qquad \left. E^{Q^X}\left\{\frac{\Lambda^{-1}_{(k+1)\Delta}\left(X^{(k+1)\Delta}, y^{(k+1)\Delta}\right)}{\Lambda^{-1}_{k\Delta}\left(X^{k\Delta}, y^{k\Delta}\right)} 1_{\{X_{(k+1)\Delta}=x_i\}}|X_{k\Delta}=x_j\right\}\right\} \end{aligned} \tag{23}$$

Using the fact that X_k is constant on $[k\Delta, (k+1)\Delta)$, we obtain

$$E^{Q^X}\left\{\frac{\Lambda^{-1}_{(k+1)\Delta}\left(X^{(k+1)\Delta}, y^{(k+1)\Delta}\right)}{\Lambda^{-1}_{k\Delta}\left(X^{k\Delta}, y^{k\Delta}\right)} 1_{\{X_{(k+1)\Delta}=x_i\}}|X_{k\Delta}=x_j\right\} = p_{ji}\cdot E_j^k \tag{24}$$

where p_{ji} is the transition probability, from state x_j to x_i of the approximating chain $\{X_k\}$ and E_j^k is as in (21). Hence we obtain

$$q_{k+1}^{y}(x_i) = \sum_{j=1}^{M} q_k(x_j)^y \cdot p_{ji}\cdot E_j^k \tag{25}$$

which, written in matrix form, gives (22).

Q.E.D.

Remark 4.2 The expression for E_j^k in (21) simplifies considerably, if v and λ are time independent; in that case E_j^k itself becomes time independent. In the special case when one neglects the information coming from the jump-heights of Y, which corresponds to assuming $v(t,x)\equiv \underline{v}=\overline{v}$, we obtain a known formula (see e.g. [5]) observations forming a conditional Poisson process.

References

[1] T. G. Andersen. Return volatility and trading volume: an information flow interpretation of stochastic volatility. *Journal of Finance*, 51(1):169–204, 1996.

[2] P.Billingsley, *Convergence of Probability Measures* John Wiley, New York, 1968.

[3] P.Bremaud, *Point Processes and Queues : Martingale Dynamics*, Springer Verlag, New York, 1981.

[4] A. Calzolari and G. Nappo, A Filtering Problem with Counting Observations : Approximation with Error Bounds, *Stochastics and Stochastics Reports* 57:71–87, 1996.

[5] G.B.Di Masi, and W.J. Runggaldier, On approximation methods for nonlinear filtering, *In: S.K. Mitter and A.Moro (eds.), Nonlinear Filtering and Stochastic Control, Lecture Notes in Mathematics Vol. 972*, Springer Verlag Berlin, 249–259, 1982.

[6] S.N. Ethier, and T.G. Kurtz, *Markov Processes : Characterization and Convergence*, John Wiley, New York, 1986.

[7] R. Elliott, C. Lahaie, and D. Madan, Filtering Derivative Security Valuations from Market Prices, *In: M. Dempster and S. Pliska (eds.), Mathematics of Derivative Securities*, Cambridge University Press pp. 141–161, 1997.

[8] P. Fisher, E. Platen, and W. Runggaldier, Risk-minimizing hedging strategies under partial information. To appear in *Proceedings of the 1996 Monte Veritá Seminar on Stochastic Analysis, Random Fields and Applications*, Birkhäuser Verlag.

[9] R. Frey, Derivative asset analysis in models with level-dependent and stochastic volatility. *CWI Quarterly, Amsterdam*, 10:1–34, 1997.

[10] R. Frey and W. Runggaldier, Risk-minimizing hedging strategies under restricted information: the case of stochastic volatility models observed only at discrete random times. *Mathematical Methods of Operations Research*, 50:339-350.

[11] M.Fujisaki, G.Kallianpur, and H.Kunita, Stochastic differential equations for the nonlinear filtering problem, *Osaka J. Math.*, 1:19–40, 1972.

[12] H. Geman, D. Madan and M. Yor, Asset prices are Brownian motion: only in business time. Preprint, University of Paris, 1998.

[13] D.M. Guillaume, M. Dacorogna, R. Davé, U. Müller, R. Olsen, and P. Pictet. From the bird's eye to the microscope: A survey of new stylized facts of the intra-daily foreign exchange markets. *Finance and Stochastics*, 1:95–129, 1997.

[14] H.J.Kushner, *Probability Methods for Approximations in Stochastic Control and for Elliptic Equations*, Academic Press, New York, 1977.

[15] H.J.Kushner and P.G. Dupuis, *Numerical Methods for Stochastic Control Problems in Continuous Time*, Springer Verlag, 1993.

[16] R.S. Liptser and A.N.Shiryaev, *Statistics of Random Processes I*, Springer Verlag, New York, 1977.

[17] L.C.G. Rogers and O. Zane. Designing models for high-frequency data. Preprint, University of Bath, 1998.

[18] T. Rydberg and N. Shephard. A modelling framework for prices and trades made at the New York stock exchange," Nuffield College working paper series 1999-W14, Oxford, forthhcoming in *Nonlinear and nonstationary signal processing*, edited by W.J. Fitzgerald et. al. Cambridge University Press, 2000.