Credit Risk and Incomplete Information: a Nonlinear-Filtering Approach

Rüdiger Frey
Department of Mathematics
University of Leipzig
ruediger.frey@math.uni-leipzig.de*

Wolfgang Runggaldier
Department of Mathematics
University of Padua
runggal@math.unipd.it†

August 2, 2007

Abstract

We study reduced-form portfolio credit risk models where the default intensities of the firms in a given portfolio depend on a common state variable process $\mathbf{X}$. In line with market practice, we assume that this state variable process is not directly observable; instead we assume that the information set of investors contains only the default history of the portfolio and noisy price observations of traded credit derivatives. This incomplete-information setup is a convenient way of modelling information-driven default contagion. In this context we study the pricing of credit derivatives and the computation of other economically interesting quantities. It turns out that this leads to a nonlinear filtering problem in a natural way. This filtering problem is studied in a general jump-diffusion model with common jumps of the state process $\mathbf{X}$ and the jump process $\mathbf{Y}$ associated with the default times of the firms. We provide recursive updating rules for the filter and derive a finite-dimensional filter for the case where $\mathbf{X}$ follows a finite-state Markov chain. Finally, we provide a filter-approximation result which justifies the use of the filter obtained for the finite-state Markov-chain case as a computational tool for more general models.

Keywords: Credit risk, nonlinear filtering, marked point processes, default contagion.

1 Introduction

This paper is concerned with reduced-form credit risk models under incomplete information. We consider models where the default intensities of the firms under consideration depend on a latent state variable process $\mathbf{X}$, but where the information set of investors contains only the default history of the portfolio and noisy price observations of traded credit derivatives. This incomplete-information setup is in line with market practice. Moreover, when the factors cannot be observed directly but their distribution has to be updated successively on the basis of past and current observations, one encounters the phenomenon of *information-driven default contagion*; see for instance Schönbucher (2004), Collin-Dufresne, Goldstein & Helwege (2003), Giesecke (2004) or Section 9.8 of McNeil, Frey & Embrechts (2005).

*Postal address: Fakultät für Mathematik und Informatik, Universität Leipzig, 04081 Leipzig, Germany.

†Postal address: Dipartimento di Matematica Pura ed Aplicata, Universitá di Padova, Via Trieste 63, 35121 Padova, Italy. Financial support from the European Science Foundation via the programme Advanced Mathematical Methods for Finance (AMaMeF) is gratefully acknowledged.

Our contribution is twofold: From a financial point of view we propose a novel two-step procedure for model building and the pricing of credit derivatives via martingale methods in this incomplete-information context. In Step 1 we consider an artificial model, where the state variable process $\mathbf{X}$ driving default intensities is assumed to be observable. This so-called *full-information-model* can be studied with Markov-process techniques. In Step 2 we study the pricing problem in the realistic *incomplete-information model* which is constructed by projecting default intensities and price dynamics from the full-information model onto the information $(\mathcal{F}_t^I)$ actually available to investors. We assume that $(\mathcal{F}_t^I)$ contains the default history $(\mathcal{H}_t)$ of the portfolio under consideration and noisy price-observations for traded credit derivatives, modelled via observations $\mathbf{Z}$ of nonlinear functions of the state process $\mathbf{X}$ in additive Gaussian noise. It is shown that Step 2 leads in a natural way to a nonlinear filtering problem: one has to determine the conditional distribution of $\mathbf{X}_t$ given the investor information $\mathcal{F}_t^I$.

The analysis of this nonlinear filtering problem forms our probabilistic contribution. Financial considerations explained in detail in the body of the paper lead us to consider fairly general models where the state process $\mathbf{X}$ and the default indicator process $\mathbf{Y}$ have common jumps. The presence of common jumps between $\mathbf{X}$ and the component $\mathbf{Y}$ of the observations makes the filtering problem quite involved. In particular, $\mathbf{X}$ and $\mathbf{Y}$ cannot be made independent by a change-of-measure and this complicates especially the derivation of filter-approximation results, as will become apparent in Section 6. Related filtering problems with common jumps of the unobserved state process and the observations have previously been discussed in the literature. First results can be found in Grigelionis (1972). The papers Kliemann, Koch & Marchetti (1990) and Ceci & Gerardi (2006) consider the case of common jumps between $\mathbf{X}$ and the observations, but are concerned with scalar observations described by a pure jump process. The recent papers Cvitanic, Liptser & Rozovski (2006) and Cvitanic, Rozovski & Zalyapin (2006) on the other hand treat the filtering problem for a very general model with marked point process observations but without common jumps between the state and the observation processes. All these papers correspond to the innovations approach and the so-called Kushner-Stratonovich equations. In the present paper we present a different derivation of the filter exploiting the recursive structure of the default history $(\mathcal{H}_t)$. In this way we obtain alternative general filter formulas. These formulas lead to explicit recursive solutions via a finite-dimensional filter in the case when $\mathbf{X}$ is a finite-state continuous-time Markov chain as well as in the so-called *affine case.* Here we concentrate on the former; affine models are discussed in detail in Frey, Prosdocimi & Runggaldier (2006). Finally, we derive a novel filter-approximation result, showing that the filter for finite-state Markov chains can be used as a computational tool for evaluating the filter for more general state variable processes of jump-diffusion type. The proof of this result relies heavily on our alternative general filter formulas.

The paper is organized as follows. The full information model is discussed in Section 2. In Section 3 we introduce the incomplete-information setup and show how the computation of various economically interesting quantities leads to a nonlinear filtering problem in a natural way. The ensuing filtering problem is then studied in Sections 4 to 6.

2 The Model under Full Information

We work on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$; all stochastic processes considered will be $(\mathcal{F}_t)$-adapted. Throughout we consider a fixed credit portfolio consisting of a set of m firms; the $(\mathcal{F}_t)$-stopping time τ_i denotes the default time of firm i. The current default state of the portfolio is summarized by the default indicator process $\mathbf{Y} = (Y_{t,1}, \dots, Y_{t,m})_{t \geq 0}$ with $Y_{t,i} = 1_{\{\tau_i \leq t\}}$

or, conversely, $\tau_i = \inf\{t \geq 0, Y_{t,i} = 1\}$. Depending on the context, P will represent either the historical probability measure or a pricing/martingale measure.

We assume that the factor process $\mathbf{X} = (X_{t,1}, \ldots, X_{t,d})_{t\geq 0}$ and default indicator process $\mathbf{Y}$ solve the following SDE

$$\begin{aligned}\mathbf{X}_t = \mathbf{X}_0 + \int_0^t b(\mathbf{X}_{s-}, \mathbf{Y}_{s-})ds + \int_0^t \sigma(\mathbf{X}_{s-}, \mathbf{Y}_{s-})d\mathbf{W}_s \\ + \int_0^t \int_E K^{\mathbf{X}}(\mathbf{X}_{s-}, \mathbf{Y}_{s-}, u)\mathcal{N}(ds, du),\end{aligned} \tag{1}$$

$$Y_{t,j} = Y_{0,j} + \int_0^t \int_E (1 - Y_{s-,j})K_j^{\mathbf{Y}}(\mathbf{X}_{s-}, \mathbf{Y}_{s-}, u)\mathcal{N}(ds, du), \quad 1 \leq j \leq m. \tag{2}$$

Here $\mathbf{W}$ is a standard k-dimensional Brownian motion; drift $b = (b_1, \ldots, b_d)$ and dispersion matrix $\sigma = (\sigma_{i,l})$, $1 \leq i \leq d, 1 \leq l \leq k$ are functions from $S^{\mathbf{X}} \times \{0,1\}^m$ to $\mathbb{R}^d$ and $\mathbb{R}^{d\times k}$ respectively, $S^{\mathbf{X}} \subset \mathbb{R}^d$ is the state space of $\mathbf{X}$; $\mathcal{N}(ds, du)$ denotes a $(P, (\mathcal{F}_t))$-standard Poisson random measure on $\mathbb{R}_+ \times E$, E some Euclidean space, with compensator measure $F_{\mathcal{N}}(du)ds$; $\mathbf{W}$ and $\mathcal{N}$ are independent; $\mathbf{X}_0$ is a random vector taking values in $S^{\mathbf{X}} \subset \mathbb{R}^d$; $\mathbf{Y}_0$ is a given element of $\{0,1\}^m$. We assume that $K_j^{\mathbf{Y}}(\boldsymbol{x}, \boldsymbol{y}, u) \in \{0,1\}$ for all $\boldsymbol{x}, \boldsymbol{y}, u$ and all $1 \leq j \leq m$, so that the solution of (2) is in fact of the form $Y_{t,j} = 1_{\{\tau_j \leq t\}}$. By $\Sigma_{\mathbf{X}}(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^{d\times d}$ we denote the matrix $\sigma(\boldsymbol{x}, \boldsymbol{y})\sigma'(\boldsymbol{x}, \boldsymbol{y})$. Define

$$D_i^{\mathbf{X}}(\boldsymbol{x}, \boldsymbol{y}) := \{u \in E : K_i^{\mathbf{X}}(\boldsymbol{x}, \boldsymbol{y}, u) \neq 0\}, \quad 1 \leq i \leq d, \tag{3}$$

$$D_j^{\mathbf{Y}}(\boldsymbol{x}, \boldsymbol{y}) := \{u \in E : K_j^{\mathbf{Y}}(\boldsymbol{x}, \boldsymbol{y}, u) \neq 0\}, \quad 1 \leq j \leq m. \tag{4}$$

We make the following assumptions.

A1. For every pair $(\boldsymbol{x}, \boldsymbol{y}) \in S^{\mathbf{X}} \times \{0,1\}^m$ the SDE system (1), (2) has a global solution with $\mathbf{X}_0 = \boldsymbol{x}, \mathbf{Y}_0 = \boldsymbol{y}$; moreover, pathwise uniqueness holds. (Sufficient conditions are discussed in Remark 2.1 below).

A2. For all $1 \leq i \leq d$, $1 \leq j \leq m$ and all $T \geq 0$ we have $E\left(\int_0^T F_{\mathcal{N}}(D_i^{\mathbf{X}}(\mathbf{X}_s, \mathbf{Y}_s))\, ds\right) < \infty$ and $E\left(\int_0^T F_{\mathcal{N}}(D_j^{\mathbf{Y}}(\mathbf{X}_s, \mathbf{Y}_s))ds\right) < \infty$.

A3. For all $1 \leq j_1 < j_2 \leq m$ and all $(\boldsymbol{x}, \boldsymbol{y}) \in S^{\mathbf{X}} \times \{0,1\}^m$ we have $F_{\mathcal{N}}(D_{j_1}^{\mathbf{Y}}(\boldsymbol{x}, \boldsymbol{y}) \cap D_{j_2}^{\mathbf{Y}}(\boldsymbol{x}, \boldsymbol{y})) = 0$.

Assumption A2 ensures that the expected number of jumps of $\mathbf{X}$ on every time interval $[0, T]$ is finite and that $\tau_i > 0$ for all firms i such that $Y_{0,i} = 0$. Assumption A3 ensures that for $j_1 \neq j_2$ the processes Y_{j_1} and Y_{j_2} have no common jumps so that there are no joint defaults. Note however, that the model (1), (2) allows for common jumps of $\mathbf{X}$ and $\mathbf{Y}$. More precisely, there is a strictly positive probability that the factor process $\mathbf{X}$ jumps at τ_j, if

$$F_{\mathcal{N}}\big(D_j^{\mathbf{Y}}(\mathbf{X}_{\tau_j-}, \mathbf{Y}_{\tau_j-}) \cap D_i^{\mathbf{X}}(\mathbf{X}_{\tau_j-}, \mathbf{Y}_{\tau_j-})\big) > 0 \text{ for some } 1 \leq i \leq d. \tag{5}$$

A simple model with common jumps of $\mathbf{X}$ and $\mathbf{Y}$ is presented in Example 2.2 below.

Remark 2.1 (Sufficient conditions for A1). There are various types of conditions ensuring strong existence and uniqueness of the system (1), (2). In Theorem 2.2 of Kliemann et al. (1990) strong existence and uniqueness is proved under growth conditions on $b(\boldsymbol{x}, \boldsymbol{y})$, $\Sigma_X(x, \boldsymbol{y})$, $F_{\mathcal{N}}(D_i^{\mathbf{X}}(\boldsymbol{x}, \boldsymbol{y}))$, $F_{\mathcal{N}}(D_j^{\mathbf{Y}}(\boldsymbol{x}, \boldsymbol{y}))$ and under the additional assumption that for every fixed $\boldsymbol{y}$ the SDE $d\mathbf{X}_t = b(\mathbf{X}_t, \boldsymbol{y})ds + \sigma(\mathbf{X}_t, \boldsymbol{y})d\mathbf{W}_s$ has a unique weak solution which is moreover a Feller process. Alternatively, one can impose growth and Lipschitz conditions on the data of the problem; see for instance Appendix 1 (Section 4) of Ceci & Gerardi (2006).

Default intensities. By definition of the compensator of a Poisson random measure,

$$Y_{t,j} - \int_0^t \int_E (1 - Y_{s-,j}) K_j^{\mathbf{Y}}(\mathbf{X}_{s-}, \mathbf{Y}_{s-}, u)\, F_{\mathcal{N}}(du) ds = Y_{t,j} - \int_0^{t\wedge\tau_j} F_{\mathcal{N}}(D_j^{\mathbf{Y}}(\mathbf{X}_{s-}, \mathbf{Y}_{s-}))\, ds$$

is a $(\mathcal{F}_t)$-martingale, so that $\lambda_j(\mathbf{X}_{t-}, \mathbf{Y}_{t-}) := F_{\mathcal{N}}(D_j^{\mathbf{Y}}(\mathbf{X}_{t-}, \mathbf{Y}_{t-}))$ is the $(\mathcal{F}_t)$-default intensity of firm j.

Some notation. Typically we take $\mathbf{Y}_0 = \mathbf{0}$. In that case Assumption A2 permits us to introduce the *ordered default times* $0 = T_0 < T_1 < \cdots < T_m < T_{m+1} := \infty$ and the counting process $N_t := \max\{n \le m \colon T_n \le t\}$; for $n \ge 1$ the rv ξ_n denotes the identity of the firm defaulting at T_n. The sequence $(T_n, \xi_n)_{1\le n\le m}$ gives a representation of $\mathbf{Y}$ as marked point process. The σ-field $\mathcal{H}_t = \sigma(\mathbf{Y}_s : s \le t) = \sigma\big(\{(T_n, \xi_n) : n = 1, \ldots, N_t\}\big)$ is the internal filtration of $\mathbf{Y}$ or, in economic terms, the default history of the portfolio at time t. For further use we define the functions

$$\bar\lambda_j(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{j} (1 - y_i)\lambda_i(\boldsymbol{x}, \boldsymbol{y}),\ 0 \le j \le m,\ \text{and}\ \bar\lambda(\boldsymbol{x}, \boldsymbol{y}) := \bar\lambda_m(\boldsymbol{x}, \boldsymbol{y}). \tag{6}$$

Example 2.2 (extended Davis Lo model). This model is an extended version of the model proposed in Davis & Lo (2001). It is of interest in our context, as it is a simple model where $\mathbf{X}$ and $\mathbf{Y}$ do have common jumps. We begin with an intuitive description. The state process X is modeled as a finite-state Markov chain with state space $S^X = \{1, \ldots, K\} \subset \mathbb{R}$; the default intensity of firm j is given by $\lambda_j(X_t)$ for increasing functions $\lambda_j : S^X \to \mathbb{R}^+$. At a default time T_n, X jumps upward by one unit with probability p_{ξ_n} which might depend on the identity ξ_n of the nth defaulting firm, and remains constant with probability $1 - p_{\xi_n}$ (unless, of course, if $X_{T_n-} = K$, where we assume that X remains constant). If the system is in an "ignited state", i.e. if $X_t \ge 2$, X_t jumps to $X_t - 1$ with intensity $\gamma(X_t)$; these downward jumps occur independently of the default history. An upward jump in the state process of X as consequence of a default can be viewed as contagious effect of the default, as the default probability of the remaining firms is increased. Note that our model is more flexible than the (Davis & Lo 2001)-model in that the contagious effect may occur randomly.

In the context of the general SDE-model (1), (2), these dynamics of X can be modeled as follows: Put $E = \mathbb{R}$, let $F_{\mathcal{N}} = \nu$, the Lebesgue-measure on $\mathbb{R}$, define for $k = 1, \cdots, K-1$ and $\boldsymbol{y} \in \{0,1\}^m$ the sets

$$I_{k,k+1}^{\boldsymbol{y}} = \bigcup_{j=1}^{m} \left[\bar\lambda_{j-1}(k, \boldsymbol{y}), \bar\lambda_{j-1}(k, \boldsymbol{y}) + (1 - y_j) p_j \lambda_j(k)\right]\,,$$

$\bar\lambda$ and $\bar\lambda_j$ as in (6), and put, for $k = 2, \cdots, K$, $I_{k,k-1}^{\boldsymbol{y}} = [-\gamma(k), 0]$. Then let

$$K^X(k, \boldsymbol{y}, u) = 1_{I_{k,k+1}^{\boldsymbol{y}}}(u) - 1_{I_{k,k-1}^{\boldsymbol{y}}}(u)\,, \quad k = 1, \cdots, K,$$

and $K_j^{\mathbf{Y}}(k, \boldsymbol{y}, u) = 1_{[\bar\lambda_{j-1}(k,\boldsymbol{y}), \bar\lambda_j(k,\boldsymbol{y})]}(u)$. This structure reflects the fact that a jump of X from k to $k+1$ can occur only at one of the default times, while between defaults only jumps from k to $k-1$ are possible. Moreover, the default intensities are given by $\nu(D_j^{\mathbf{X}}(x, \boldsymbol{y})) = (1 - y_j)\lambda_j(x)$. Both facts are of course in line with the informal description of the model.

Example 2.3 (conditionally independent defaults). Next we consider a model with conditionally independent doubly-stochastic default times (see eg. Section 9.6 of McNeil et al. (2005)).

Following the literature such as Duffie & Garleanu (2001) we assume that $\mathbf{X}$ follows a jump diffusion model of the form

$$d\mathbf{X}_t = b(\mathbf{X}_t)dt + \sigma(\mathbf{X}_t)d\mathbf{W}_t + d\mathbf{J}_t,$$

where $\mathbf{J}$ denotes a $\mathbb{R}^d$-valued compound Poisson process with compensator measure $F_{\mathbf{J}}(d\boldsymbol{x})ds$. Moreover, the default intensity of firm j at time t is equal to $\lambda_j(\mathbf{X}_t)$. A possible choice for $\mathcal{N}$, $K^{\mathbf{X}}$ and $K^{\mathbf{Y}}$ is as follows. Take $E = \mathbb{R}^d \times \mathbb{R}$, $F_{\mathcal{N}} = F_{\mathbf{J}} \times \nu$, ν Lebesgue-measure on $\mathbb{R}$, and put

$$K_j^{\mathbf{Y}}(\boldsymbol{x}, \boldsymbol{u}) = 1_{\left[\sum_{i=1}^{j-1}\lambda_i(\boldsymbol{x}), \sum_{i=1}^{j}\lambda_i(\boldsymbol{x})\right]}(u_{d+1}), \quad 1 \le j \le m, \text{ and} \tag{7}$$

$$K_i^{\mathbf{X}}(\boldsymbol{x}, \boldsymbol{u}) = u_i 1_{[-1,0)}(u_{d+1}), \quad 1 \le i \le d. \tag{8}$$

Note that $K^{\mathbf{X}}$ and $K^{\mathbf{Y}}$ have been chosen so that $F_{\mathcal{N}}(D_i^{\mathbf{X}}(\boldsymbol{x}) \cap D_j^{\mathbf{Y}}(\boldsymbol{x})) = 0$ for all $1 \le i \le d$, $1 \le j \le m$ and all $\boldsymbol{x}$ in $S^{\mathbf{X}}$.

3 Incomplete Information

The investor filtration. We assume that investors observe the default history of the portfolio under consideration; moreover they observe a function of the state variable $\mathbf{X}$ in additive Gaussian noise. We make the following formal assumption on the investor filtration $(\mathcal{F}_t^I)$:

A4. $\mathcal{F}_t^I = \mathcal{H}_t \vee \mathcal{F}_t^{\mathbf{Z}}$, $t \ge 0$, where the l-dimensional process $\mathbf{Z}$ solves the SDE

$$d\mathbf{Z}_t = \boldsymbol{a}_t(\mathbf{X}_t)dt + v\, d\boldsymbol{\beta}_t\,. \tag{9}$$

Here $\boldsymbol{\beta}$ is an l-dimensional standard Brownian motion on $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$, independent of $\mathbf{X}$ and $\mathbf{Y}$; v denotes an *invertible* $l \times l$ matrix of constants; $\boldsymbol{a}_t(\cdot) \colon \Omega \times S^{\mathbf{X}} \to \mathbb{R}^l$ is an $\mathcal{H}_t$-measurable random function. Note that $\boldsymbol{a}_t(\cdot)$ is of the form

$$\boldsymbol{a}_t(\omega; \boldsymbol{x}) = \sum_{n=0}^{m} 1_{\{T_n(\omega) \le t < T_{n+1}(\omega)\}}\, a_n\big(t, \boldsymbol{x};\, T_j(\omega), \xi_j(\omega) : 1 \le j \le n\big), \quad \boldsymbol{x} \in S^{\mathbf{X}}; \tag{10}$$

we moreover assume that the functions functions $a_n \colon [0,\infty) \times S^{\mathbf{X}} \times \big([0,\infty) \times \{1, \dots, m\}\big)^n \to \mathbb{R}^l$ are continuous and bounded.

The economic interpretation of $\mathbf{Z}$ is discussed below. Note that by the independence of $\mathbf{X}$ and $\boldsymbol{\beta}$, the factor process $\mathbf{X}$ is not $(\mathcal{F}_t^I)$-adapted.

Next we show that under incomplete information, the computation of many economically relevant quantities leads to a of nonlinear filtering problem in a natural way.

Pricing credit derivatives and nonlinear filtering. Credit derivatives are securities whose payoff at maturity T depends on the default history of some underlying reference portfolio; in abstract terms their payoff is hence given by some $\mathcal{H}_T$-measurable random variable H. We study the pricing of these securities using the popular martingale modeling approach (Björk 1998). Hence we work directly under some risk-neutral pricing measure - denoted by P - and assume that the P-dynamics of the processes $\mathbf{X}$ and $\mathbf{Y}$ are given by (1), (2). We assume that investor information is given by the filtration $(\mathcal{F}_t^I)$ introduced in Assumption A4. In line with the literature on portfolio credit derivatives we moreover take the default-free short rate $r_t > 0$

to be deterministic. In this context the price of a credit derivative with maturity $T > t$ and $\mathcal{H}_T$-measurable payoff H is given by

$$H_t := E\Big(\exp\big(- \int_t^T r_s ds\big) H \mid \mathcal{F}_t^I \Big). \tag{11}$$

In order to simplify the computation of H_t we define the *full-information value* $\tilde{H}_t$ of H as

$$\tilde{H}_t := E\big(\exp\big(- \int_t^T r_s ds\big) H \mid \mathcal{F}_t \big), \quad t \le T. \tag{12}$$

By the Markovianity of the pair $(\mathbf{X}, \mathbf{Y})$, we have $\tilde{H}_t = h_t(\mathbf{X}_t)$ for some $\mathcal{H}_t$-measurable random function $h_t : \Omega \times S^{\mathbf{X}} \to \mathbb{R}$. For instance, we get in case of a defaultable zero-coupon bond on firm j with zero recovery rate, maturity T and payoff $H = 1 - Y_{T,j}$

$$E\Big(e^{-\int_t^T r_s ds}\,(1 - Y_{T,j}) \mid \mathcal{F}_t\Big) = \big(1 - Y_{t,j}\big) E_{(\mathbf{X}_t, \mathbf{Y}_t)}\Big(e^{-\int_0^{T-t} r_s ds}\,(1 - Y_{T-t,j})\Big). \tag{13}$$

Note that expression (13) is a function of t, $\mathbf{X}_t$ and $\mathbf{Y}_t$ and hence clearly of the form (10). Now we get by iterated conditional expectations

$$H_t = E\Big(E\big(e^{-\int_t^T r_s ds} H \mid \mathcal{F}_t\big) \mid \mathcal{F}_t^I\Big) = E(h_t(\mathbf{X}_t) \mid \mathcal{F}_t^I). \tag{14}$$

In order to compute the price H_t from the rhs of (14) we need to determine (in weak form) $\pi_{\mathbf{X}_t | \mathcal{F}_t^I}(d\boldsymbol{x})$, the conditional distribution of $\mathbf{X}_t$ given $\mathcal{F}_t^I$. This is a typical *nonlinear filtering problem* which will be studied in detail in Sections 4, 5 and 6.

Default intensities and default contagion. It is well-known that $(\mathcal{F}_t^I)$-default-intensities can be computed by projection, i.e. the $(\mathcal{F}_t^I)$-default intensity of firm j is given by the left-continuous version of

$$\hat{\lambda}_{t,j} = E\left(\lambda_j(\mathbf{X}_t, \mathbf{Y}_t) \mid \mathcal{F}_t^I\right) = \int_{\mathbb{R}^d} \lambda(\boldsymbol{x}, \mathbf{Y}_t)\, \pi_{\mathbf{X}_t | \mathcal{F}_t^I}(d\boldsymbol{x}), \quad t \le \tau_j. \tag{15}$$

Relation (15) illustrates nicely the notion of information-based default contagion that was already mentioned in the introduction: new default information such as the news that obligor $i \ne j$ has defaulted leads to an update in the conditional distribution $\pi_{\mathbf{X}_t | \mathcal{F}_t^I}(d\boldsymbol{x})$ and hence to a jump in the $(\mathcal{F}_t^I)$-default intensity of firm j.

Default probabilities. In our setup the computation of default and survival probabilities leads to a nonlinear filtering problem as well.[1] Since the pair $(\mathbf{X}_t, \mathbf{Y}_t)$ is Markov, the survival probability at time t of firm i beyond time $T > t$ is in the full information case given by

$$P(\tau_i > T \mid \mathcal{F}_t) = (1 - Y_{t,i}) E_{(\mathbf{X}_t, \mathbf{Y}_t)}(1 - Y_{T-t,i}) := \Pi(t, \mathbf{X}_t, \mathbf{Y}_t).$$

By iterated conditional expectations we get then for the incomplete information case

$$P\left(\tau_i > T \mid \mathcal{F}_t^I\right) = E\left(\Pi(t, \mathbf{X}_t, \mathbf{Y}_t) \mid \mathcal{F}_t^I\right).$$

To compute this quantity we again need the conditional distribution $\pi_{\mathbf{X}_t | \mathcal{F}_t^I}(d\boldsymbol{x})$.

[1] Note that in this context it is more natural to assume that P represents the historical probability measure.

Interpretation of Z. The process $\mathbf{Z}$ can be taken to represent cumulative noisy price-observations for traded credit derivatives. This interpretation is motivated by the following heuristic argument. Denote by $\boldsymbol{a}_t(\mathbf{X}_t) = \big(a_{t,1}(\mathbf{X}_t), \ldots, a_{t,l}(\mathbf{X}_t)\big)$ the full-information value of l traded credit derivative at time t. Suppose that investors observe market quotes for theses contract at equidistant time points $t_k = k\Delta$ for some small $\Delta > 0$ (eg. daily observations) and that the quotes actually observed by investors are of the form $\boldsymbol{z}_{t_k} = \boldsymbol{a}_{t_k}(\mathbf{X}_{t_k}) + \boldsymbol{\epsilon}_k$ for an iid sequence of $\mathbb{R}^l$-valued noise variables $(\boldsymbol{\epsilon}_k)_k$, independent of $\mathbf{X}$, with $E(\boldsymbol{\epsilon}_1) = \mathbf{0}$ and $\mathrm{cov}(\boldsymbol{\epsilon}_1) = \tilde{\Sigma}_{\boldsymbol{\epsilon}}$. Assume that $\tilde{\Sigma}_{\boldsymbol{\epsilon}}$ is strictly positive definite and choose an invertible root $\tilde{v}$. Define the scaled *cumulative observation process* $\mathbf{Z}_t^{\Delta} := \Delta \sum_{t_k \le t} \boldsymbol{z}_{t_k}$ and let $v = \sqrt{\Delta}\tilde{v}$. Then we have for Δ small, using Donsker's invariance principle,

$$\mathbf{Z}_t^{\Delta} = \sum_{t_k \le t} \boldsymbol{a}_{t_k}(\mathbf{X}_{t_k})\Delta + \Delta \sum_{t_k \le t} \boldsymbol{\epsilon}_k \approx \int_0^t \boldsymbol{a}_s(\mathbf{X}_s) ds + v\boldsymbol{\beta}_t. \tag{16}$$

Note that $\Sigma_{\mathbf{Z}} := vv' = \Delta\tilde{\Sigma}_{\boldsymbol{\epsilon}}$, so that the instantaneous covariance matrix of $\mathbf{Z}$ in (16) is proportional to the covariance matrix $\tilde{\Sigma}_{\boldsymbol{\epsilon}}$ of the observation noise and inversely proportional to the observation frequency $1/\Delta$. The noise - $(\boldsymbol{\epsilon}_k)_k$ respectively $v\boldsymbol{\beta}$ - represents classical observation errors such as bid-ask spreads and transmission errors as well as so-called *model errors.* The latter reflect the fact that the full-information model will typically not be able to fully explain observed price trajectories of traded credit derivatives. In applications the error covariance matrix v will be chosen by the modeler in an attempt to balance calibration errors (resulting from error variances which are too large) against potential instabilities of the filter (resulting from error variances which are too low). For an in-depth discussion of the choice of v in practical situations we refer to Jazwinski (1970). The idea of interpreting observed prices of derivatives as noisy observations of functions of the factor process is also pursued in Gombani, Jaschke & Runggaldier (2005).

Remark 3.1 (Common jumps of X and Y). The interpretation of $\mathbf{Z}$ as noisily observed full-information value of traded credit derivatives suggests to use a model where $\mathbf{X}$ and $\mathbf{Y}$ have common jumps, for the following reason: It is a well-established empirical fact that spreads of traded credit derivatives frequently increase in reaction to unexpected major default events; see for instance Collin-Dufresne et al. (2003). Under the above interpretation of the observation process $\mathbf{Z}$, this should of course be reflected in its dynamics, i.e. the model should allow for the possibility that the full-information values $\boldsymbol{a}_t(\mathbf{X}_t)$ appearing in the drift-term of (9) jump at default times. Common jumps of factor process $\mathbf{X}$ and default indicator $\mathbf{Y}$ - as present in our framework - are a convenient way to achieve this.

4 Filter Equations

We have seen that the problems of pricing and statistical analysis in credit risk models under incomplete information lead to the following filtering problem. Given a bounded and (for approximation purposes) continuous function $f : \mathbb{R}^d \to \mathbb{R}$ develop a recursive approach to computing

$$\pi_t f := E(f(\mathbf{X}_t) \mid \mathcal{F}_t^I)\,.$$

In mathematical terms this is a filtering problem with mixed observations of marked-point process and diffusion type and with common jumps of the state process $\mathbf{X}$ and the point-process observation $\mathbf{Y}$. The specific solution to the given filtering problem is derived in Subsections 4.3 and 4.4 below. In 4.3 we consider the case between two successive default times and in 4.4 the updating of $\pi_t f$ at a default time. Between default times the new information essentially comes from observing the process $\mathbf{Z}$ in (9). This leads to a classical nonlinear filtering problem with diffusion information for which it is convenient to use the so-called reference probability approach based on an absolutely continuous measure transformation; this measure transformation and its implications are discussed in Subsection 4.1. In Subsection 4.2 we study the dynamics of $\mathbf{X}$ between default events. As is usual in nonlinear filtering, the general filtering equations are difficult to solve explicitly. In Section 5 we therefore study the particular case when the factor process is finite-state Markov and show that in this case the filter equations can be solved explicitly. Since general (jump-diffusion) processes can be approximated by finite-state Markov chains, in Section 6 we finally discuss the approximations of the filter for a general factor process by that for finite-state Markov factors.

Notation. In the sequel we use the following pieces of notation. By $\Sigma_{\mathbf{Z}}^{-1} = (vv')^{-1}$ we denote the inverse of the instantaneous covariation matrix of $\mathbf{Z}$; for any vector $\boldsymbol{a} \in \mathbb{R}^l$ we define $\|\boldsymbol{a}\|^2_{\Sigma_{\mathbf{Z}}^{-1}} := \boldsymbol{a}'\Sigma_{\mathbf{Z}}^{-1}\boldsymbol{a}$.

4.1 Measure transformation and reduction to $(\mathcal{H}_t)$

It will be convenient to model the processes $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ on a product space $(\Omega, \mathcal{F}, (\mathcal{F}_t), R^0)$ so that $\mathbf{Z}$ is independent of $\mathbf{X}$ and $\mathbf{Y}$, and to introduce dependence via an equivalent change of measure. For this we denote by $(\Omega_2, \mathcal{F}_2, (\mathcal{F}_t^2), P^{0,l})$ the l-dimensional Wiener space with coordinate process $\boldsymbol{\beta}^0$, i.e. $\boldsymbol{\beta}_t^0(\omega_2) = \omega_2(t)$. Given some probability space $(\Omega_1, \mathcal{F}_1, (\mathcal{F}_t^1), P)$ supporting a solution $(\mathbf{X}, \mathbf{Y})$ of the SDE-system (1), (2), let $\Omega := \Omega_1 \times \Omega_2$, $\mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2$, $\mathcal{F}_t = \mathcal{F}_t^1 \otimes \mathcal{F}_t^2$, $R^0 := P \times P^{0,l}$, and put for $\omega = (\omega_1, \omega_2) \in \Omega$

$$\mathbf{X}_t(\omega) := \mathbf{X}_t(\omega_1),\ \mathbf{Y}_t(\omega) := \mathbf{Y}_t(\omega_1),\ \text{and}\ \boldsymbol{\beta}_t^0(\omega) := \boldsymbol{\beta}_t^0(\omega_2).$$

Note that this implies that under R^0, $\boldsymbol{\beta}^0$ is l-dimensional Brownian motion, independent of $\mathbf{X}$ and $\mathbf{Y}$. Define the process $\mathbf{Z}_t := v\boldsymbol{\beta}_t^0$. Introduce a Girsanov-type measure transformation of the form $\frac{dR}{dR^0}_{|\mathcal{F}_t} = L_t$ with

$$\begin{aligned} L_t &= \exp\left\{ \int_0^t \left(v^{-1}\boldsymbol{a}_s(\mathbf{X}_s)\right)' d\boldsymbol{\beta}_s^0 - \frac{1}{2}\int_0^t \|\boldsymbol{a}_s(\mathbf{X}_s)\|^2_{\Sigma_{\mathbf{Z}}^{-1}}\, ds \right\} \\ &= \exp\left\{ \int_0^t \boldsymbol{a}_s'(\mathbf{X}_s)\Sigma_{\mathbf{Z}}^{-1} d\mathbf{Z}_s - \frac{1}{2}\int_0^t \|\boldsymbol{a}_s(\mathbf{X}_s)\|^2_{\Sigma_{\mathbf{Z}}^{-1}}\, ds \right\}, \end{aligned} \tag{17}$$

and note that the process L is indeed a R^0-martingale of mean one as $\boldsymbol{a}_t(\cdot)$ is bounded by A4. Using the Girsanov theorem for Brownian motion we therefore obtain that, under R, the process $\mathbf{Z}$ has the original dynamics (9); moreover, since $\boldsymbol{\beta}^0$ is orthogonal to both $\mathbf{W}$ and the martingale that results from compensating the counting measure $\mathcal{N}$, the above measure transformation induces no changes in the law of $\mathbf{X}$ and $\mathbf{Y}$. Hence under R the triple of processes $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ has indeed the correct joint law. Notice finally that by (17), L can be expressed in terms of the observation $\mathbf{Z}$. By the following Bayes formula, known as Kallianpur-Striebel formula (Kallianpur & Striebel 1968), we then have

$$\pi_t f := E^R(f(\mathbf{X}_t) \mid \mathcal{F}_t^I) = \frac{E^{R^0}(f(\mathbf{X}_t)L_t \mid \mathcal{F}_t^I)}{E^{R^0}(L_t \mid \mathcal{F}_t^I)}, \tag{18}$$

so that, to compute $\pi_t f$, it suffices to compute the numerator on the right-hand side in (18).

Recall that $\mathcal{F}_t^I = \mathcal{H}_t \vee \mathcal{F}_t^{\mathbf{Z}}$. Next we reduce the conditioning on $\mathcal{F}_t^I$ to a conditioning on $\mathcal{H}_t$. Since $\mathbf{Z}_t = v\boldsymbol{\beta}_t^0$ with v invertible, we have $\mathcal{F}_t^{\mathbf{Z}} = \mathcal{F}_t^{\boldsymbol{\beta}^0}$, i.e. given $\mathcal{F}_t^{\mathbf{Z}}$, $\omega_2(s), s \le t$ is "known." Using the Fubini theorem and the product-structure of $(\Omega, \mathcal{F}, (\mathcal{F}_t), R^0)$ we therefore get

$$E^{R^0}(f(\mathbf{X}_t)L_t \mid \mathcal{H}_t \vee \mathcal{F}_t^{\mathbf{Z}})(\omega) = E^P(f(\mathbf{X}_t)L_t(\cdot, \omega_2) \mid \mathcal{H}_t)(\omega_1)\,, \text{ with} \tag{19}$$
$$L_t(\omega_1, \omega_2) = \exp\left\{ \int_0^t \boldsymbol{a}_s'(\omega_1; \mathbf{X}_s(\omega_1))\Sigma_{\mathbf{Z}}^{-1} d\mathbf{Z}_s(\omega_2) - \frac{1}{2}\int_0^t \|\boldsymbol{a}_s(\omega_1; \mathbf{X}_s(\omega_1)\|^2_{\Sigma_{\mathbf{Z}}^{-1}}\, ds \right\}.$$

In order to compute $\pi_t f$ we thus have to evaluate the conditional expectation on the right hand side of (19). Note that this involves only the first component $(\Omega_1, \mathcal{F}_1, (\mathcal{F}_t^1), P)$ of the underlying probability space and hence only the joint law of $\mathbf{X}$ and $\mathbf{Y}$; expectations with respect to that law will be simply denoted by E (instead of E^P).

4.2 Dynamics of X between default times

Next we discuss the dynamics of $\mathbf{X}$ for $t \in [T_{n-1}, T_n)$ i.e. between default times. This is a prerequisite for the filtering equations in the next subsections and for the derivation of approximation results in Section 6. For this purpose we define the new kernel

$$\bar{K}^{\mathbf{X}}(\boldsymbol{x}, \boldsymbol{y}, u) := \begin{cases} 0, & \text{if} \quad u \in \bar{D}^{\mathbf{Y}}(\boldsymbol{x}, \boldsymbol{y}) := \bigcup_{\{j\,:\,\boldsymbol{y}_j = 0\}} D_j^{\mathbf{Y}}(\boldsymbol{x}, \boldsymbol{y})\,, \\ K^{\mathbf{X}}(\boldsymbol{x}, \boldsymbol{y}, u) & \text{else.} \end{cases} \tag{20}$$

We will see shortly that the kernel $\bar{K}^{\mathbf{X}}$ governs the jumps of $\mathbf{X}$ between default times. For instance, in case of the extended Davis Lo model (Example 2.2), $K^X(k, \boldsymbol{y}, u) = -1_{[-\gamma(k),0]}(u)$, $k = 1, \cdots, K$, reflecting the fact that between defaults X can jump only downwards.

Consider, for $t > T_{n-1}$, the SDE system

$$\begin{aligned} \widetilde{\mathbf{X}}_t = \mathbf{X}_{T_{n-1}} &+ \int_{T_{n-1}}^t b(\widetilde{\mathbf{X}}_s, \mathbf{Y}_{T_{n-1}})ds + \int_{T_{n-1}}^t \sigma(\widetilde{\mathbf{X}}_s, \mathbf{Y}_{T_{n-1}})d\mathbf{W}_s \\ &+ \int_{T_{n-1}}^t \int_E \bar{K}^{\mathbf{X}}(\widetilde{\mathbf{X}}_{s-}, \mathbf{Y}_{T_{n-1}}, u)\mathcal{N}(ds, du), \end{aligned} \tag{21}$$
$$\widetilde{Y}_{t,j} = \mathbf{Y}_{T_{n-1},j} + \int_{T_{n-1}}^t \int_E (1 - \widetilde{Y}_{s-,j})K_j^{\mathbf{Y}}(\widetilde{\mathbf{X}}_{s-}, \widetilde{\mathbf{Y}}_{T_{n-1}}, u)\mathcal{N}(ds, du), \quad 1 \le j \le m. \tag{22}$$

Comparing the system (21), (22) with the original model dynamics (1), (2), we see that $K^{\mathbf{X}}$ has been replaced with $\bar{K}^{\mathbf{X}}$; moreover, in the coefficients of the equation the "initial value" $\mathbf{Y}_{T_{n-1}}$ replaces $\mathbf{Y}_{t-}$. Since, for $T_{n-1} \le t < T_n$, there are no atoms of $\mathcal{N}(ds, du)$ in $\{t\} \times \bar{D}^{\mathbf{Y}}(\mathbf{X}_{t-}, \mathbf{Y}_{T_{n-1}})$, by definition of $\bar{K}^{\mathbf{X}}$ in (20) we have the equality

$$\int_{T_{n-1}}^t \int_E K^{\mathbf{X}}(\mathbf{X}_{s-}, \mathbf{Y}_{s-}, u)\mathcal{N}(ds, du) = \int_{T_{n-1}}^t \int_E \bar{K}^{\mathbf{X}}(\mathbf{X}_{s-}, \mathbf{Y}_{T_{n-1}}, u)\mathcal{N}(ds, du)\,,\ T_{n-1} \le t < T_n.$$

Strong uniqueness of (21) and (22) - implied by Assumption A1 - therefore yields that a.s.

$$\mathbf{X}_t = \widetilde{\mathbf{X}}_t \ \text{ for } T_{n-1} \le t < T_n, \quad \text{ and } \ \mathbf{Y}_t = \widetilde{\mathbf{Y}}_t \ \text{ for } T_{n-1} \le t \le T_n \ (T_n \text{ included})\,. \tag{23}$$

Denote by $\bar{P}_{(\boldsymbol{x},\boldsymbol{y})}$ the law of the solution $(\bar{\mathbf{X}}, \bar{\mathbf{Y}})$ to the SDE system (21), (22) starting at $t = 0$ in the point $(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^d \times \{0,1\}^m$. Now the law of $\big(\widetilde{\mathbf{X}}_{t+T_{n-1}}, \widetilde{\mathbf{Y}}_{t+T_{n-1}}\big)_{t \ge 0}$ is obviously equal to

$\bar{P}_{(\mathbf{X}_{T_{n-1}},\mathbf{Y}_{T_{n-1}})}$ and so $\bar{P}_{(\mathbf{X}_{T_{n-1}},\mathbf{Y}_{T_{n-1}})}$ also governs the evolution of the original process $(\mathbf{X}_{t-},\mathbf{Y}_t)$ for $T_{n-1} < t \le T_n$.

Define the stopping time $\bar{T}_1 := \inf\{t \ge 0 : \Delta \bar{\mathbf{Y}}_t \neq 0\}$ and denote by $\bar{\xi}_1 \in \{1, \cdots, m\}$ the identity of the first jump firm in the "bar-model". Note that that $\bar{T}_1$ is a standard doubly stochastic random time. Hence we have

$$\bar{P}_{(\boldsymbol{x},\boldsymbol{y})}\Big(\bar{T}_1 > t \mid \mathcal{F}_\infty^{\bar{\mathbf{X}}}\Big) = \exp\Big\{ - \int_0^t \bar{\lambda}(\bar{\mathbf{X}}_s, \boldsymbol{y}) ds\Big\} \text{ and} \tag{24}$$

$$h_{\bar{T}_1,\bar{\xi}_1|\mathcal{F}_\infty^{\bar{\mathbf{X}}}}(t, i \mid \bar{\mathbf{X}}) = \lambda_i(\bar{\mathbf{X}}_t, \boldsymbol{y}) \exp\Big\{ - \int_0^t \bar{\lambda}(\bar{\mathbf{X}}_s, \boldsymbol{y}) ds\Big\}, \tag{25}$$

where $h_{\bar{T}_1,\bar{\xi}_1|\mathcal{F}_\infty^{\bar{\mathbf{X}}}}$ is the conditional density of $\bar{T}_1, \bar{\xi}_1$ under $\bar{P}_{(\boldsymbol{x},\boldsymbol{y})}$, given $\mathcal{F}_\infty^{\bar{\mathbf{X}}}$ (see e.g. Section 9.6.2 of McNeil et al. (2005)). Properties (24) and (25) will be essential for the derivation of the filter equations.

4.3 Filtering between defaults

We denote the distribution of $\mathbf{X}_0$ given $\mathcal{F}_0^I$ (the initial filter distribution) by π_0. The main result of this subsection is

Theorem 4.1. *Given two successive default times T_{n-1}, T_n, we have for $t \in [T_{n-1}, T_n)$*

$$\begin{aligned}\pi_t f \propto \int_{\mathbb{R}^d} \bar{E}_{(\boldsymbol{x},\mathbf{Y}_{T_{n-1}})}\Big(f\left(\bar{\mathbf{X}}_{t-T_{n-1}}\right) \exp\Big\{ - \int_0^{t-T_{n-1}} \bar{\lambda}(\bar{\mathbf{X}}_s, \mathbf{Y}_{T_{n-1}}) ds\Big\} \\ \cdot \exp\Big\{ \int_0^{t-T_{n-1}} \boldsymbol{a}'_{T_{n-1}+s}(\bar{\mathbf{X}}_s)\Sigma_{\mathbf{Z}}^{-1}\, d\mathbf{Z}_{s+T_{n-1}}(\omega_2) - \frac{1}{2}\int_0^{t-T_{n-1}} \big\|\boldsymbol{a}_{s+T_{n-1}}(\bar{\mathbf{X}}_s)\big\|^2_{\Sigma_{\mathbf{Z}}^{-1}}\, ds\Big\}\Big)\, \pi_{T_{n-1}}(d\boldsymbol{x})\,.\end{aligned} \tag{26}$$

Here $\bar{E}_{(\boldsymbol{x},\boldsymbol{y})}$ denotes expectation under the measure $\bar{P}_{(\boldsymbol{x},\boldsymbol{y})}$ introduced in Section 4.2 and $\pi_{T_{n-1}}(d\boldsymbol{x})$ is the filter distribution at $t = T_{n-1}$.

The proportionality factor is given by the right hand side of (26) evaluated at $f(\boldsymbol{x}) \equiv 1$.

Proof. Recall that by (18) and (19), $\pi_t f \propto E\big(f(\mathbf{X}_t)L_t(\cdot,\omega_2) \mid \mathcal{H}_t\big)$. Let $Y_t^{T_n} = 1_{\{T_n \le t\}}$ denote the indicator of the random time T_n and denote by $\mathcal{F}_t^{Y^{T_n}} = \sigma\big(Y_s^{T_n} : s \le t\big)$ the corresponding filtration. Now note that for $T_{n-1} \le t < T_n$, $\mathcal{H}_t = \mathcal{H}_{T_{n-1}} \vee \mathcal{F}_t^{Y^{T_n}}$. By the so-called Dellacherie-formula (see for instance Lemma 3.1 in Elliott, Jeanblanc & Yor (2000)) we get for any integrable, $\mathcal{F}_\infty$-measurable random variable U that

$$E\big(U 1_{\{T_n > t\}} \mid \mathcal{H}_t\big) = 1_{\{T_n>t\}} \frac{E\big(U 1_{\{T_n>t\}} \mid \mathcal{H}_{T_{n-1}}\big)}{P\big(T_n > t \mid \mathcal{H}_{T_{n-1}}\big)}\,. \tag{27}$$

With $U = f(\mathbf{X}_t)L_t(\cdot, \omega_2)$ we therefore obtain for $t \in [T_{n-1}, T_n)$

$$E(f(\mathbf{X}_t)L_t(\cdot,\omega_2) \mid \mathcal{H}_t) \propto E(f(\mathbf{X}_t)L_t(\cdot,\omega_2)1_{\{T_n>t\}} \mid \mathcal{H}_{T_{n-1}})\,. \tag{28}$$

By double conditioning we get

$$\begin{aligned}&E\big(f(\mathbf{X}_t)L_t(\cdot,\omega_2)1_{\{T_n>t\}} \mid \mathcal{H}_{T_{n-1}}\big) \\ &\quad = E\Big(L_{T_{n-1}}(\cdot,\omega_2)E\Big(f(\mathbf{X}_t)\frac{L_t(\cdot,\omega_2)}{L_{T_{n-1}}(\cdot,\omega_2)}\, 1_{\{T_n>t\}} \mid \mathcal{F}_{T_{n-1}}\Big) \mid \mathcal{H}_{T_{n-1}}\Big)\,.\end{aligned} \tag{29}$$

By the Markov property of $(\mathbf{X},\mathbf{Y})$ and the equality in law discussed in the previous subsection below equation (23), the inner conditional expectation in (29) equals

$$
\begin{aligned}
\bar{E}_{(\mathbf{X}_{T_{n-1}},\mathbf{Y}_{T_{n-1}})}\Big(f(\bar{\mathbf{X}}_{t-T_{n-1}})\exp\Big\{\int_0^{t-T_{n-1}}&\boldsymbol{a}'_{T_{n-1}+s}(\bar{\mathbf{X}}_s)\Sigma_{\mathbf{Z}}^{-1}\,d\mathbf{Z}_{s+T_{n-1}}(\omega_2)\\
-\frac{1}{2}\int_0^{t-T_{n-1}}\big\|\boldsymbol{a}_{s+T_{n-1}}(\bar{\mathbf{X}}_s)\big\|^2_{\Sigma_{\mathbf{Z}}^{-1}}\,ds\Big\}1_{\{\bar{T}_1>t-T_{n-1}\}}\Big)\,.
\end{aligned}
\tag{30}
$$

Using the survival function of $\bar{T}_1$ as given in (24) and double conditioning on $\mathcal{F}^{\bar{\mathbf{X}}}_\infty$, (30) is equal to

$$
\begin{aligned}
&\bar{E}_{(\mathbf{X}_{T_{n-1}},\mathbf{Y}_{T_{n-1}})}\Big(f(\bar{\mathbf{X}}_{t-T_{n-1}})\exp\Big\{\int_0^{t-T_{n-1}}\boldsymbol{a}'_{T_{n-1}+s}(\bar{\mathbf{X}}_s)\Sigma_{\mathbf{Z}}^{-1}\,d\mathbf{Z}_{s+T_{n-1}}(\omega_2)\\
&\quad-\frac{1}{2}\int_0^{t-T_{n-1}}\big\|\boldsymbol{a}_{s+T_{n-1}}(\bar{\mathbf{X}}_s)\big\|^2_{\Sigma_{\mathbf{Z}}^{-1}}\,ds\Big\}\cdot\exp\Big\{-\int_0^{t-T_{n-1}}\bar{\lambda}(\bar{\mathbf{X}}_s,\mathbf{Y}_{T_{n-1}})\,ds\Big\}\Big)\,.
\end{aligned}
$$

By the Kallianpur Striebel formula (18), expression (29) is therefore proportional to

$$
\begin{aligned}
&\int_{\mathbb{R}^d}\bar{E}_{(\boldsymbol{x},\mathbf{Y}_{T_{n-1}})}\Bigg(f(\bar{\mathbf{X}}_{t-T_{n-1}})\exp\Big\{\int_0^{t-T_{n-1}}\boldsymbol{a}'_{T_{n-1}+s}(\bar{\mathbf{X}}_s)\Sigma_{\mathbf{Z}}^{-1}\,d\mathbf{Z}_{s+T_{n-1}}(\omega_2)\\
&\quad-\frac{1}{2}\int_0^{t-T_{n-1}}\big\|\boldsymbol{a}_{s+T_{n-1}}(\bar{\mathbf{X}}_s)\big\|^2_{\Sigma_{\mathbf{Z}}^{-1}}\,ds\Big\}\cdot\exp\Big\{-\int_0^{t-T_{n-1}}\bar{\lambda}(\bar{\mathbf{X}}_s,\mathbf{Y}_{T_{n-1}})\,ds\Big\}\Bigg)\,\pi_{T_{n-1}}(d\boldsymbol{x})\,,
\end{aligned}
$$

which coincides with the right hand side in (26). □

4.4 Filtering at a default time T_n

Again by (18) and (19), at a generic default time T_n we have $\pi_{T_n}f \propto E(f(\mathbf{X}_{T_n})L_{T_n}(\cdot,\omega_2)\mid\mathcal{H}_{T_n})$. Notice now that, due to the possibility of common jumps between $\mathbf{X}$ and $\mathbf{Y}$, the expressions $E(f(\mathbf{X}_{T_n})L_{T_n}(\cdot,\omega_2)\mid\mathcal{H}_{T_n})$ and $E(f(\mathbf{X}_{T_n^-})L_{T_n}(\cdot,\omega_2)\mid\mathcal{H}_{T_n})$ do not necessarily coincide. We shall therefore proceed along two steps. In Step 1 we show that one can obtain the conditional expectation $E(f(\mathbf{X}_{T_n})L_{T_n}(\cdot,\omega_2)\mid\mathcal{H}_{T_n})$ once one is able to compute $E(g(\mathbf{X}_{T_n^-})L_{T_n}(\cdot,\omega_2)\mid\mathcal{H}_{T_n})$ for a generic function $g(\cdot)$. In this step we use the joint distribution of the jumps $\Delta\mathbf{X}_{T_n}$ and $\Delta\mathbf{Y}_{T_n}$ and hence the particular structure of the given model. In Step 2 we then compute the latter of those two quantities via Bayesian updating.

Step 1 (Reduction to the filter distribution of $\mathbf{X}_{T_n-}$). Here we show

Proposition 4.2. *We have the relation*

$$
E(f(\mathbf{X}_{T_n})L_{T_n}(\cdot,\omega_2)\mid\mathcal{H}_{T_n}) = E(g(\mathbf{X}_{T_n^-};Y_{T_{n-1}},\xi_n)L_{T_n}(\cdot,\omega_2)\mid\mathcal{H}_{T_n}),
$$

where the function g is given by

$$
g(\boldsymbol{x};\boldsymbol{y},j)=\begin{cases}F_{\mathcal{N}}\big(D_j^{\mathbf{Y}}(\boldsymbol{x},\boldsymbol{y})\big)^{-1}\displaystyle\int_{D_j^{\mathbf{Y}}(\boldsymbol{x},\boldsymbol{y})}f\big(\boldsymbol{x}+K^{\mathbf{X}}(\boldsymbol{x},\boldsymbol{y},u)\big)F_{\mathcal{N}}(du)\,, & \textit{if } F_{\mathcal{N}}\big(D_j^{\mathbf{Y}}(\boldsymbol{x},\boldsymbol{y})\big)>0\,,\\ f(\boldsymbol{x})\,, & \textit{else.}\end{cases}
\tag{31}
$$

Note that without common jumps of $\mathbf{X}$ and $\mathbf{Y}$, i.e. if $F_{\mathcal{N}}(D_j^{\mathbf{Y}}(\boldsymbol{x},\boldsymbol{y}) \cap D_i^{\mathbf{X}}(\boldsymbol{x},\boldsymbol{y})) = 0$ for all $i, j, \boldsymbol{x}, \boldsymbol{y}$, we have that $K^{\mathbf{X}}(\boldsymbol{x},\boldsymbol{y},u)$ is zero on $D_j^{\mathbf{Y}}(\boldsymbol{x},\boldsymbol{y})$ $F_{\mathcal{N}}-$a.s., and hence

$$\int_{D_j^{\mathbf{Y}}(\boldsymbol{x},\boldsymbol{y})} f\big(\boldsymbol{x} + K^{\mathbf{X}}(\boldsymbol{x},\boldsymbol{y},u)\big) F_{\mathcal{N}}(du) = F_{\mathcal{N}}\big(D_j^{\mathbf{Y}}(\boldsymbol{x},\boldsymbol{y})\big) f(\boldsymbol{x}) \,.$$

Consequently $g(\boldsymbol{x},\boldsymbol{y},j) \equiv f(\boldsymbol{x})$, even for $F_{\mathcal{N}}\big(D_j^{\mathbf{Y}}(\boldsymbol{x},\boldsymbol{y})\big) > 0$ and so, for the case without common jumps, Step 1 becomes superfluous.

Proof. By (1), (2) we have for $t \in (T_{n-1}, T_n]$ and for j with $Y_{T_{n-1},j} = 0$ the equality

$$\big\{(T_n,\xi_n) = (t,j)\big\} = \big\{\mathcal{N}\left(\{t\} \times D_j^{\mathbf{Y}}(\mathbf{X}_{t^-}, \mathbf{Y}_{T_{n-1}})\right) = 1\big\} \ ,$$

or, in words, $(T_n,\xi_n) = (t,j)$ if and only if $\mathcal{N}(ds,du)$ has an atom in the set $\{t\} \times D_j^{\mathbf{Y}}(\mathbf{X}_{t^-}, \mathbf{Y}_{T_{n-1}})$. Hence

$$\mathcal{H}_{T_n} \subset \widetilde{\mathcal{F}}_{T_n^-} := \mathcal{F}_{T_n^-} \vee \sigma\Big(1_{\big\{\mathcal{N}\big(\{T_n\} \times D_{\xi_n}^{\mathbf{Y}}(\mathbf{X}_{T_n^-}, \mathbf{Y}_{T_{n-1}})\big)=1\big\}}\Big) \,. \tag{32}$$

Since moreover $\mathbf{X}_{T_n} = \mathbf{X}_{T_n^-} + \int_E K^{\mathbf{X}}(\mathbf{X}_{T_n^-}, \mathbf{Y}_{T_{n-1}}, u)\, \mathcal{N}(\{T_n\}, du)$, we get by double conditioning, using (32),

$$\begin{aligned}
&E(f(\mathbf{X}_{T_n}) L_{T_n}(\cdot,\omega_2) \mid \mathcal{H}_{T_n}) \\
&\quad = E\left(E\Big(L_{T_n}(\cdot,\omega_2) \int_E f\big(\mathbf{X}_{T_n^-} + K^{\mathbf{X}}(\mathbf{X}_{T_n^-}, \mathbf{Y}_{T_{n-1}}, u)\big)\, \mathcal{N}(\{T_n\}, du) \mid \widetilde{\mathcal{F}}_{T_n^-}\Big) \mid \mathcal{H}_{T_n}\right) .
\end{aligned}$$

Now note that L_{T_n} and $\mathbf{X}_{T_n^-}$ are $\mathcal{F}_{T_n^-}$ measurable whereas $\mathcal{N}(\{T_n\}, du)$ is independent of $\mathcal{F}_{T_n^-}$ with compensator measure $F_{\mathcal{N}}(du)$. Moreover, given $\mathcal{F}_{T_n^-}$, conditioning on $\widetilde{\mathcal{F}}_{T_n^-}$ is equivalent to conditioning on the fact that $\mathcal{N}(\{T_n\}, du)$ has an atom in the set $\{T_n\} \times D_{\xi_n}^{\mathbf{Y}}\big(\mathbf{X}_{T_n^-}, \mathbf{Y}_{T_{n-1}}\big)$. Hence the inner conditional expectation is equal to $g(\mathbf{X}_{T_n^-}; Y_{T_{n-1}}, \xi_n) L_{T_n}(\cdot,\omega_2)$, and the result follows. $\square$

Step 2 (Updating of the conditional distribution of $\mathbf{X}_{T_n^-}$). Here we have

Theorem 4.3. *Given the information that a default has actually occurred at $t = T_n$ and given the identity ξ_n of the defaulting firm, for a generic function $g : \mathbb{R}^d \to \mathbb{R}$ we have*

$$\begin{aligned}
&E\big(g(\mathbf{X}_{T_n^-}) L_{T_n}(\cdot,\omega_2) \mid \mathcal{H}_{T_n}\big) \propto \\
&\quad \int_{\mathbb{R}^d} \bar{E}_{(\boldsymbol{x},\mathbf{Y}_{T_{n-1}})}\bigg(g(\bar{\mathbf{X}}_{T_n - T_{n-1}}) \lambda_{\xi_n}\left(\bar{\mathbf{X}}_{T_n - T_{n-1}}, \mathbf{Y}_{T_{n-1}}\right) \exp\Big\{ - \int_0^{T_n - T_{n-1}} \bar{\lambda}(\bar{\mathbf{X}}_s, \mathbf{Y}_{T_{n-1}})\, ds \Big\} \\
&\quad \cdot \exp\Big\{ \int_0^{T_n - T_{n-1}} \boldsymbol{a}'_{s+T_{n-1}}(\bar{\mathbf{X}}_s) \Sigma_{\mathbf{Z}}^{-1} d\mathbf{Z}_{s+T_{n-1}}(\omega_2) - \frac{1}{2} \int_0^{T_n - T_{n-1}} \big\| \boldsymbol{a}_{s+T_{n-1}}(\bar{\mathbf{X}}_s) \big\|^2_{\Sigma_{\mathbf{Z}}^{-1}}\, ds \Big\}\bigg) \pi_{T_{n-1}}(d\boldsymbol{x}) \,.
\end{aligned} \tag{33}$$

Proof. The proof goes in two steps: first we reduce the claim to a statement with respect to $\bar{P}_{(\boldsymbol{x},\mathbf{Y}_{T_{n-1}})}$; secondly we apply Bayesian updating. Since $\mathcal{H}_{T_n} = \mathcal{H}_{T_{n-1}} \vee \sigma(T_n, \xi_n)$ we get

$$\begin{aligned}
&E\big(g(\mathbf{X}_{T_n^-}) L_{T_n}(\cdot,\omega_2) \mid \mathcal{H}_{T_n}\big) \\
&\quad = E\Big(L_{T_{n-1}}(\cdot,\omega_2) E\Big(g(\mathbf{X}_{T_n^-}) \frac{L_{T_n}(\cdot,\omega_2)}{L_{T_{n-1}}(\cdot,\omega_2)} \mid \mathcal{F}_{T_{n-1}}, T_n, \xi_n\Big) \mid \mathcal{H}_{T_{n-1}}, T_n, \xi_n\Big) \,.
\end{aligned} \tag{34}$$

We now concentrate on the inner expectation. Using (23) we get that

$$
\begin{aligned}
&E\Big(g(\mathbf{X}_{T_n^-})\frac{L_{T_n}(\cdot,\omega_2)}{L_{T_{n-1}}(\cdot,\omega_2)} \mid \mathcal{F}_{T_{n-1}}, T_n=t, \xi_n=i\Big)\\
&= E\Big(g(\widetilde{\mathbf{X}}_{T_n^-})\exp\Big\{\int_{T_{n-1}}^{T_n} \boldsymbol{a}_s'(\widetilde{\mathbf{X}}_s)\Sigma_{\mathbf{Z}}^{-1}\,d\mathbf{Z}_s(\omega_2)-\frac{1}{2}\int_{T_{n-1}}^{T_n}\Big\|\boldsymbol{a}_s(\widetilde{\mathbf{X}}_s)\Big\|^2_{\Sigma_{\mathbf{Z}}^{-1}}\,ds\Big\}\;\Big|\;\mathcal{F}_{T_{n-1}}, T_n=t,\xi_n=i\Big)\,.
\end{aligned}
$$

Now note that, due to the equality in law as discussed after relation (23), given $\mathcal{F}_{T_{n-1}}$, the joint law of $\big((\mathbf{X}_s)_{T_{n-1}\le s<T_n}, T_n-T_{n-1},\xi_n\big)$ equals the law of $\big((\bar{\mathbf{X}}_s)_{0\le s<\bar T_1}, \bar T_1, \bar\xi_1\big)$ under $\bar P_{(\mathbf{X}_{T_{n-1}},\mathbf{Y}_{T_{n-1}})}$. Moreover, $\Delta\bar{\mathbf{X}}_{\bar T_1}=0$ a.s. Hence the last term equals

$$
\begin{aligned}
\bar E_{(\mathbf{X}_{T_{n-1}},\mathbf{Y}_{T_{n-1}})}\Big(g(\bar{\mathbf{X}}_{\bar T_1})\exp\Big\{&\int_0^{\bar T_1}\boldsymbol{a}'_{s+T_{n-1}}(\bar{\mathbf{X}}_s)\Sigma_{\mathbf{Z}}^{-1}\,d\mathbf{Z}_{s+T_{n-1}}(\omega_2)\\
&-\frac{1}{2}\int_0^{\bar T_1}\big\|\boldsymbol{a}_{s+T_{n-1}}(\bar{\mathbf{X}}_s)\big\|^2_{\Sigma_{\mathbf{Z}}^{-1}}\,ds\Big\}\;\Big|\;\bar T_1=t-T_{n-1},\bar\xi_1=i\Big)\,.
\end{aligned}
\tag{35}
$$

We are now in a position to do the Bayesian updating. Recall from (25) the form of the conditional density $h_{\bar T_1,\bar\xi_1|\mathcal{F}^{\bar{\mathbf{X}}}_\infty}$. By double conditioning on $\mathcal{F}^{\bar{\mathbf{X}}}_\infty\vee\sigma\big(\bar T_1,\bar\xi_1\big)$ and the Bayes formula, (35) is easily seen to be proportional to

$$
\begin{aligned}
&\bar E_{(\mathbf{X}_{T_{n-1}},\mathbf{Y}_{T_{n-1}})}\Big(g(\bar{\mathbf{X}}_{t-T_{n-1}})\lambda_i(\bar{\mathbf{X}}_{t-T_{n-1}},\mathbf{Y}_{T_{n-1}})\exp\Big\{\int_0^{t-T_{n-1}}\bar\lambda(\bar{\mathbf{X}}_s,\mathbf{Y}_{T_{n-1}})ds\Big\}\\
&\quad\cdot\exp\Big\{\int_0^{t-T_{n-1}}\boldsymbol{a}'_{s+T_{n-1}}(\bar{\mathbf{X}}_s)\Sigma_{\mathbf{Z}}^{-1}\,d\,\mathbf{Z}_{s+T_{n-1}}(\omega_2)-\frac{1}{2}\int_0^{t-T_{n-1}}\big\|\boldsymbol{a}_{s+T_{n-1}}(\bar{\mathbf{X}}_s)\big\|^2_{\Sigma_{\mathbf{Z}}^{-1}}\,ds\Big\}\Big)\,.
\end{aligned}
$$

Combining this with (34), thereby using (18) and (19), gives the result. □

5 Finite-state Markov factor process

Consider now the case when the factor process is a scalar finite-state Markov chain X. These models are meaningful in themselves; relevant examples include the model of Graziano & Rogers (2006) or the generalized Davis-Lo model (Example 2.2). Moreover, the results for the finite state Markov case can be used to construct a filter approximation for more general situations, as will be shown in Section 6 below. In Subsection 5.1 we derive the particular form that the filter obtained in Sections 4.3 and 4.4 takes in the finite-state Markov case; in Subsection 5.2 we discuss computational aspects.

5.1 Filter equations for a finite-state Markov chain

The model for X. Without loss of generality we may assume that the chain X takes values in the set $\{1,\dots,K\}$. The transition intensities of X at time t may depend on the current state $\mathbf{Y}_t$ of the credit portfolio; we denote the generator matrix of X, given that the credit portfolio is in the state $\boldsymbol{y}\in\{0,1\}^m$, by $Q^{\boldsymbol{y}}=(q^{\boldsymbol{y}}_{k,h})$, $k,h=1,\cdots,K$. In this context the filter distribution can be summarized by the K-dimensional process $\boldsymbol{\pi}_t=(\pi^1_t,\dots,\pi^K_t)$ with $\pi^i_t:=P(X_t=i\mid\mathcal{F}^I_t)$.

We begin by modeling the dynamics of X in a way that fits into the general model framework as given in (1), (2). For the drift and diffusion coefficients we obviously have $b(x,\boldsymbol{y})\equiv\sigma(x,\boldsymbol{y})\equiv 0$. Since X is scalar, we may take $E=\mathbb{R}$ and let $F_{\mathcal{N}}(du)=\nu(du)$, ν Lebesgue measure. Concerning the kernels $K^X(\cdot)$, $K_j^{\mathbf{Y}}(\cdot)$ we may proceed as follows. For each $\boldsymbol{y}$ consider a family of sets $I_{k,h}^{\boldsymbol{y}} \subset \mathbb{R}$, $k,h = 1,\cdots,K$, such that

$$\nu\left(I_{k,h}^{\boldsymbol{y}}\right) = q_{k,h}^{\boldsymbol{y}} \text{ for all } k \neq h \quad \text{and} \quad I_{k,h_1}^{\boldsymbol{y}} \cap I_{k,h_2}^{\boldsymbol{y}} = \emptyset \text{ for all } k \text{ and all } h_1 \neq h_2\,. \tag{36}$$

Put then

$$K^X(k,\boldsymbol{y},u) = h-k \quad \text{for} \quad u \in I_{k,h}^{\boldsymbol{y}}, \tag{37}$$

so that at a generic jump time T the process X jumps to the (uniquely defined) new state h for which $\mathcal{N}\big(\{T\} \times I_{X_{T-},h}^{\boldsymbol{y}}\big) = 1$. The kernels $K_j^{\mathbf{Y}}(k,\boldsymbol{y},u)$, $1 \leq j \leq m$, have to be chosen in order to ensure that the pair of processes $(X,\mathbf{Y})$ has the desired joint law. In Example 2.2 we have shown how this can be achieved for the specific case of the extended Davis-Lo model.

For the filter equations we shall also need the particular expression of the kernel $\bar{K}^{\mathbf{X}}(\cdot)$ defined in (20) that here becomes

$$\bar{K}^{\mathbf{X}}(k,\boldsymbol{y},u) := \begin{cases} h-k\,, & \text{if} \quad u \in I_{k,h}^{\boldsymbol{y}} \setminus \bar{D}^{\mathbf{Y}}(k,\boldsymbol{y}) =: \bar{I}_{k,h}^{\boldsymbol{y}}, \quad h \neq k.\,, \\ 0\,, & \text{else.} \end{cases} \tag{38}$$

We shall denote by $\bar{Q}^{\boldsymbol{y}} = \big(\bar{q}_{k,h}^{\boldsymbol{y}}\big)$ the corresponding transition intensity matrix, for which $\bar{q}_{k,h}^{\boldsymbol{y}} = \nu\big(\bar{I}_{k,h}^{\boldsymbol{y}}\big)$.

The filter equations. The key step in applying the filtering results of Section 4, in particular Theorem 4.1 and 4.3, is the evaluation of expressions of the form

$$\sigma_t g[n,\boldsymbol{y}](\omega_2) := \sum_{i=1}^{K} \bar{E}_{(i,\boldsymbol{y})}\left(g\left(\bar{X}_t\right) L_t^n(\cdot,\omega_2) \exp\Big\{-\int_0^t \bar{\lambda}(\bar{X}_s,\boldsymbol{y})\,ds\Big\}\right) \pi_{T_{n-1}}(\{i\}) \tag{39}$$

for generic $g\colon \{1,\dots,K\} \to \mathbb{R}$, $\boldsymbol{y} \in \{0,1\}$, and with

$$L_t^n(\cdot,\omega_2) = \exp\Big\{\int_0^t (\boldsymbol{a}_s^n)'(\bar{X}_s)\Sigma_{\mathbf{Z}}^{-1}\,d\mathbf{Z}_s^n(\omega_2) - \frac{1}{2}\int_0^t \big\|\boldsymbol{a}^n(\bar{X}_s)\big\|_{\Sigma_{\mathbf{Z}}^{-1}}^2\,ds\Big\},$$

where $\mathbf{Z}_s^n := \mathbf{Z}_{s+T_{n-1}}$ and $\boldsymbol{a}_s^n(\cdot) := \boldsymbol{a}_{s+T_{n-1}}(\cdot)$. Put for $h \in \{1,\dots,K\}$

$$\sigma_t^h[n,\boldsymbol{y}](\omega_2) := \sum_{i=1}^{K} \bar{E}_{(i,\boldsymbol{y})}\left(1_{\{\bar{\mathbf{X}}_t=h\}} L_t^n(\cdot,\omega_2) \exp\Big\{-\int_0^t \bar{\lambda}(\bar{X}_s,\boldsymbol{y})\,ds\Big\}\right) \pi_{T_{n-1}}(\{i\}), \tag{40}$$

so that $\sigma_t[n,\boldsymbol{y}]g = \sum_{h=1}^{K} \sigma_t^h[n,\boldsymbol{y}]g(h)$. In the next proposition we derive a Zakai-type SDE for the vector process $\boldsymbol{\sigma}_t[n,\boldsymbol{y}] = (\sigma_t^1[n,\boldsymbol{y}],\dots,\sigma_t^K[n,\boldsymbol{y}])$ that, as follows immediately from the previous development, represents a vector of unnormalized conditional probabilities.

Proposition 5.1. *The process $\boldsymbol{\sigma}_t = \boldsymbol{\sigma}_t[n,\boldsymbol{y}]$ solves the SDE*

$$d\sigma_t^i = \Big(\sum_{k=1}^{K} \bar{q}_{k,i}^{\boldsymbol{y}}\sigma_t^k - \bar{\lambda}(i,\boldsymbol{y})\sigma_t^i\Big)\,dt + \sigma_t^i\,(\boldsymbol{a}_t^n)'(i)\Sigma_{\mathbf{Z}}^{-1}\,d\mathbf{Z}_t^n\,, \quad 1 \leq i \leq K\,, \tag{41}$$

with initial condition $\sigma_0^i = \pi_{T_{n-1}}(\{i\})$. Here $(\bar{q}_{k,h}^{\boldsymbol{y}})$, $1 \leq k,h \leq K$, denotes the intensity matrix corresponding to the kernel $\bar{K}^X$ from (38).

Proof. A similar reasoning as in Section 4.1 yields $\sigma_t^h \propto \bar{R}_\pi\left(\bar{\mathbf{X}}_t = h \mid \mathcal{F}_t^{\mathbf{Z}^n}\right)$ where under $\bar{R}_\pi$, $\bar{X}$ is a Markov chain with generator matrix $\bar{Q}^{\boldsymbol{y}}$, initial distribution π, and where $\mathbf{Z}_t^n$ solves the SDE $d\mathbf{Z}_t^n = \boldsymbol{a}_t^n(\bar{X}_t)dt + vd\boldsymbol{\beta}_t^n$, with $\boldsymbol{\beta}_t^n = \left(\boldsymbol{\beta}_{t+T_{n-1}}\right)_{t\geq 0}$.

The statement can now be derived from general results in Corollary 3.9 of Elliott (1993). For this purpose we need to define an appropriate process H that we choose as $H_t := \exp\big\{-\int_0^t \bar{\lambda}(X_s, \boldsymbol{y})ds\big\}$, so that $dH_t = -\bar{\lambda}(X_t, \boldsymbol{y})H_t dt$. With this choice of H the coefficients in (3.1) of Elliott (1993) become $\beta \equiv \delta \equiv 0$ and $\alpha_t = -\bar{\lambda}(X_t, \boldsymbol{y})H_t$. Equation (41) now corresponds to relation (3.18) in Elliott (1993). □

The filter distribution at a default time T_n. Next we specialize Proposition 4.2 to the case where the factor process is a finite-state Markov chain X. Using the construction of X as given in (36) and (37) and the convention $0/0 = 0$, a similar reasoning as in the proof of Proposition 4.2 gives

Corollary 5.2.

$$\begin{aligned}\pi_{T_n}^i := \pi_{T_n}(\{i\}) = \sum_{h\neq i} P\left(X_{T_n^-} = h \mid \mathcal{F}_{T_n}^I\right) \frac{\nu\big(D_{\xi_n}^{\mathbf{Y}}(h, \mathbf{Y}_{T_{n-1}}) \cap I_{h,i}\big)}{\nu(D_{\xi_n}^{\mathbf{Y}}(h, \mathbf{Y}_{T_{n-1}}))} + \\ + P\left(X_{T_n^-} = i \mid \mathcal{F}_{T_n}^I\right) \frac{\nu\Big(D_{\xi_n}^{\mathbf{Y}}(i, \mathbf{Y}_{T_{n-1}}) \setminus \bigcup_{h\neq i} I_{i,h}\Big)}{\nu(D_{\xi_n}^{\mathbf{Y}}(i, Y_{T_{n-1}}))}, \quad 1 \leq i \leq K\,.\end{aligned} \tag{42}$$

For illustrative purposes we note that in the special case of the generalized Davis-Lo model (Example 2.2) relation (42) becomes

$$\begin{aligned}\pi_{T_n}^1 &= (1 - p_{\xi_n})P(X_{T_n^-} = 1 \mid \mathcal{F}_{T_n}^I)\,, \\ \pi_{T_n}^k &= (1 - p_{\xi_n})P(X_{T_n^-} = k \mid \mathcal{F}_{T_n}^I) + p_{\xi_n}P(X_{T_n^-} = k-1 \mid \mathcal{F}_{T_n}^I),\ 2 \leq k \leq K-1, \\ \pi_{T_n}^K &= P(X_{T_n^-} = K \mid \mathcal{F}_{T_n}^I) + p_{\xi_n}P(X_{T_n^-} = K-1 \mid \mathcal{F}_{T_n}^I)\,.\end{aligned} \tag{43}$$

5.2 Computational aspects

If X follows a finite-state Markov chain, Theorems 4.1 and 4.3 together with Propositions 4.2 and 5.1 lead to a finite-dimensional recursive procedure for computing the filter distribution $\boldsymbol{\pi}_t$. We summarize this procedure in the following algorithm. Recall that $\left(\boldsymbol{\sigma}_s[n, \boldsymbol{y}]\right)_{s\geq 0}$ denotes the solution of (41) with initial value $\boldsymbol{\sigma}_0^i[n, \boldsymbol{y}] = \pi_{T_{n-1}}(\{i\})$, $\mathbf{Z}_s^n = \mathbf{Z}_{s+T_{n-1}}$ and $\boldsymbol{a}_s^n(\cdot) = \boldsymbol{a}_{s+T_{n-1}}(\cdot)$.

Algorithm 5.3 (Filtering for a finite-state Markov chain).

1. Set $n = 1$, $T_0 = 0$, $\mathbf{Y}_0 = 0$, and denote the initial filter distribution by π_0.

2. Compute $\sigma_t^i[n, \mathbf{Y}_{T_{n-1}}]$ for $t \geq T_{n-1}$ according to (41) (for the actual computation see the next subsection).

3. For $t \in [T_{n-1}, T_n)$, set (see Theorem 4.1) $\pi_t^i := \dfrac{\sigma_{t-T_{n-1}}^i[n, \mathbf{Y}_{T_{n-1}}]}{\sum_{k=1}^K \sigma_{t-T_{n-1}}^k[n, \mathbf{Y}_{T_{n-1}}]}$

4. For $t = T_n$ compute first (see Theorem 4.3)

$$P\left(X_{T_{n^-}} = i \mid \mathcal{F}_{T_n}^I\right) := \frac{\lambda_{\xi_n}(i, \mathbf{Y}_{T_{n-1}})\sigma_{T_n-T_{n-1}}^i[n, \mathbf{Y}_{T_{n-1}}]}{\sum_{k=1}^K \lambda_{\xi_n}(k, \mathbf{Y}_{T_{n-1}})\sigma_{T_n-T_{n-1}}^k[n, \mathbf{Y}_{T_{n-1}}]}$$

and determine then $\pi_{T_n} = \left[\pi^1_{T_n}, \ldots, \pi^K_{T_n}\right]$ according to (42). Replace n by $n+1$ and return to Step 2.

Solving the Zakai-equation (41). In order to apply Algorithm 5.3, we need to solve the SDE (41). For better illustration we distinguish here two cases: first we consider the case where $(\mathcal{F}^I_t) = (\mathcal{H}_t)$ or, equivalently, $\boldsymbol{a}^n_t(\cdot) \equiv 0$; second we consider the case with noisily observed derivative prices, where $\boldsymbol{a}^n_t(\cdot)$ is generally different from zero.

The case $\boldsymbol{a}^n_t(\cdot) \equiv 0$. In this case equation (41) reduces to the ODE-system

$$\frac{d}{dt}\sigma^i_t = \sum_{k=1}^K \bar{q}^{\boldsymbol{y}}_{k,i}\sigma^k_t - \bar{\lambda}(i,\boldsymbol{y})\sigma^i_t\,, \quad 1 \le i \le K. \tag{44}$$

In vector notation the solution of this equation is given by the matrix exponential

$$\boldsymbol{\sigma}_t[n,\boldsymbol{y}] = \boldsymbol{\sigma}_0[n,\boldsymbol{y}] \exp\left\{t\left((\bar{Q}^{\boldsymbol{y}})' - \operatorname{diag}(\bar{\lambda}(1,\boldsymbol{y})\cdots,\bar{\lambda}(K,\boldsymbol{y}))\right)\right\}.$$

This matrix exponential can be computed by diagonalizing the matrix $\bar{Q}^{\boldsymbol{y}}$. Alternatively, one can of course apply numerical schemes for ODEs to equation (44).

The case $\boldsymbol{a}^n_t(\cdot) \neq 0$. In this case (41) is a stochastic differential equation. For its solution we have two possibilities : a numerical approach or an explicit analytical solution.

1. *Numerical approach :* One may proceed with time discretization, e.g. according to the Euler-Maruyama scheme. This makes particularly sense if only discrete time observations of $\mathbf{Z}$ are available. For a discussion on various ways to implement the time discretization and for the convergence properties we refer to Clark (1978).

2. *Analytical solution.* Here the key idea, due to Clark (1978), is to reduce the stochastic differential equation (41) to a deterministic one via a well-chosen factorization. To this effect notice that a straightforward application of Itô's formula allows to show the following

Proposition 5.4. *We have $\sigma^i_t := A_{t,i} \cdot r_i(t,\mathbf{A})$ where*

$$A_{t,i} := \exp\left\{\int_0^t (\boldsymbol{a}^n_s)'(i)\Sigma_{\mathbf{Z}}^{-1}\, d\mathbf{Z}^n{}_s - \frac{1}{2}\int_0^t \|\boldsymbol{a}^n_s(i)\|^2_{\Sigma_{\mathbf{Z}}^{-1}}\, ds\right\}, \quad i = 1,\ldots,K \tag{45}$$

and

$$\frac{d}{dt}r_i(t,\mathbf{A}) = A^{-1}_{t,i}\sum_{k=1}^K A_{t,k}\big(\bar{q}^{\boldsymbol{y}}_{k,i} - \bar{\lambda}(i,\boldsymbol{y})\delta_{ik}\big)r_k(t,\mathbf{A})\,, \quad \boldsymbol{r}(0,\mathbf{A}) = \boldsymbol{\sigma}_0\,. \tag{46}$$

To actually use the results of this proposition one has to be able to determine a trajectory of $\mathbf{A} = (A_{t,1},\cdots .A_{t,K})_{|t\ge 0}$ given the observed trajectory of the process $\mathbf{Z}^n_t$. This can be accomplished by using stochastic partial integration to allow for a pathwise evaluation of the stochastic integral in (45). For this purpose one has first to strengthen assumption A4 by requiring that for given $\boldsymbol{x}$, the function $\boldsymbol{a}_t(\boldsymbol{x})$ is $\mathcal{C}^1$ in t. On the basis of this assumption and denoting by $\dot{\boldsymbol{a}}^n_t(i)$ the time derivative of $\boldsymbol{a}^n_t(i)$, it is easily seen that the stochastic integral with respect to $\mathbf{Z}^n_t$ in (45) can be given the following pathwise representation

$$\int_0^t (\boldsymbol{a}^n_s)'(i)\Sigma_{\mathbf{Z}}^{-1} d\mathbf{Z}^n_s = (\boldsymbol{a}^n_t)'(i)\Sigma_{\mathbf{Z}}^{-1}\mathbf{Z}^n_t - \int_0^t (\mathbf{Z}^n_s)'\Sigma_{\mathbf{Z}}^{-1}\,\dot{\boldsymbol{a}}^n_s(i)ds\,. \tag{47}$$

For a deeper discussion of pathwise non-linear filtering we refer to Davis (1978).

6 Filter Approximations

In the previous section we have seen that for a finite state Markov chain the conditional probabilities π_t^k can actually be computed in a recursive fashion. While such models are meaningful in themselves, there are often good reasons to work with models where $\mathbf{X}$ is given by a general jump-diffusion model with an infinite or continuous state space. For instance, for so-called affine jump-diffusion models (see e.g.Duffie, Pan & Singleton (2000)), the full-information price of many important credit derivatives can be computed explicitly. To compute the incomplete-information price it is then natural to consider approximations of $\mathbf{X}$ by a sequence $\mathbf{X}^m$ of finite-state Markov chains as a computational tool. In this section we provide the necessary theoretical basis for this and show the convergence of the filter computed for a finite-state approximation $\mathbf{X}^m$ to the filter corresponding to the original state variable process $\mathbf{X}$.

An alternative representation of the filter. We begin with an alternative expression for the filter; while more abstract than the results from Section 4, this expression is well-suited for deriving approximation results. Consider a pair of processes $(\mathbf{X},\mathbf{Y})$ solving the original SDE system (1) (2); in particular, the dynamics of $\mathbf{X}$ are given in terms of the drift vector $b(\boldsymbol{x},\boldsymbol{y})$, the dispersion matrix $\sigma(\boldsymbol{x},\boldsymbol{y})$ or equivalently the matrix $\Sigma_{\mathbf{X}}(\boldsymbol{x},\boldsymbol{y}) = \sigma(\boldsymbol{x},\boldsymbol{y})\sigma(\boldsymbol{x},\boldsymbol{y})'$ and the kernel $K^{\mathbf{X}}(\boldsymbol{x},\boldsymbol{y},u)$. Fix some $t>0$. Put as before $N_t := \max\{n \le m : T_n \le t\}$ and recall that the sequence (T_n,ξ_n), $n=1,\dots,N_t$ or equivalently the process $(\mathbf{Y}_s)_{s\le t}$ represents the default history up to time t. In the sequel we will always work with respect to given and fixed default observations $(\hat{T}_n,\hat{\xi}_n)$, $n=0,\dots,\hat{N}_t$ respectively $(\hat{\mathbf{Y}}_s)_{s\le t}$; the "hat-notation" is meant to indicate that the default observations are fixed and can hence be considered deterministic.

Recall now the definition (20) of the kernel $\bar{K}^{\mathbf{X}}$. Denote by $(\Omega^d,\mathcal{F}^d,(\mathcal{F}_t^d))$ the Skorokhod space $D^d([0,\infty))$ with its canonical filtration and denote the coordinate process on Ω^d by $\bar{\mathbf{X}}$. For reasons that will become apparent in the sequel, we define on $(\Omega^d,\mathcal{F}^d,(\mathcal{F}_t^d))$ a predictable vector process by

$$
\begin{aligned}
B_t \;=\; & \int_0^t b(\bar{\mathbf{X}}_s,\hat{\mathbf{Y}}_s)ds + \int_0^t\int_E \bar{K}^{\mathbf{X}}(\bar{\mathbf{X}}_{s-},\hat{\mathbf{Y}}_{s-},u)F_{\mathcal{N}}(du)\,ds \\
&+ \sum_{n=1}^{\hat{N}_t} F_{\mathcal{N}}\big(D^{\mathbf{Y}}_{\hat{\xi}_n}(\bar{\mathbf{X}}_{\hat{T}_n^-},\hat{\mathbf{Y}}_{\hat{T}_{n-1}})\big)^{-1}\int_{D^{\mathbf{Y}}_{\hat{\xi}_n}(\bar{\mathbf{X}}_{\hat{T}_n^-},\hat{\mathbf{Y}}_{\hat{T}_{n-1}})} \bar{K}^{\mathbf{X}}(\bar{\mathbf{X}}_{\hat{T}_n^-},\hat{\mathbf{Y}}_{\hat{T}_{n-1}},u)\,F_{\mathcal{N}}(du);
\end{aligned} \tag{48}
$$

a predictable $\mathbb{R}^{d\times d}$-valued process $\widetilde{C}_t^{ij}$, $1\le i,j\le d$, by

$$
\begin{aligned}
\widetilde{C}_t^{ij} = & \int_0^t \Sigma_{\mathbf{X}}(\bar{\mathbf{X}}_s,\hat{\mathbf{Y}}_s)ds + \int_0^t\int_E \bar{K}_i^{\mathbf{X}}(\bar{\mathbf{X}}_{s-},\hat{\mathbf{Y}}_{s-},u)\bar{K}_j^{\mathbf{X}}(\bar{\mathbf{X}}_{s-},\hat{\mathbf{Y}}_{s-},u)F_{\mathcal{N}}(du)\,ds \\
&+\sum_{n=1}^{\hat{N}_t} F_{\mathcal{N}}\big(D^{\mathbf{Y}}_{\hat{\xi}_n}(\bar{\mathbf{X}}_{\hat{T}_n^-},\hat{\mathbf{Y}}_{\hat{T}_{n-1}})\big)^{-1}\int_{D^{\mathbf{Y}}_{\hat{\xi}_n}(\bar{\mathbf{X}}_{\hat{T}_n^-},\hat{\mathbf{Y}}_{\hat{T}_{n-1}})} \bar{K}_i^{\mathbf{X}}(\bar{\mathbf{X}}_{s-},\hat{\mathbf{Y}}_{s-},u)\bar{K}_j^{\mathbf{X}}(\bar{\mathbf{X}}_{s-},\hat{\mathbf{Y}}_{s-},u)F_{\mathcal{N}}(du);
\end{aligned} \tag{49}
$$

and finally a predictable random measure ν on $[0,\infty)\times\mathbb{R}^d$ given for bounded and measurable

$\gamma : \mathbb{R}^d \to \mathbb{R}$ by

$$\begin{aligned}\int_0^t \gamma(\boldsymbol{x})\nu(ds, d\boldsymbol{x}) &= \int_0^t \int_E \gamma\left(\bar{K}^{\mathbf{X}}(\bar{\mathbf{X}}_{\hat{T}_n^-}, \hat{\mathbf{Y}}_{\hat{T}_{n-1}}, u)\right) F_{\mathcal{N}}(du)ds \\ &+ \sum_{n=1}^{\hat{N}_t} F_{\mathcal{N}}\big(D^{\mathbf{Y}}_{\hat{\xi}_n}(\bar{\mathbf{X}}_{\hat{T}_n^-}, \hat{\mathbf{Y}}_{\hat{T}_{n-1}})\big)^{-1} \int_{D^{\mathbf{Y}}_{\hat{\xi}_n}(\bar{\mathbf{X}}_{\hat{T}_n^-}, \hat{\mathbf{Y}}_{\hat{T}_{n-1}})} \gamma\left(\bar{K}^{\mathbf{X}}(\bar{\mathbf{X}}_{\hat{T}_n^-}, \hat{\mathbf{Y}}_{\hat{T}_{n-1}}, u)\right) F_{\mathcal{N}}(du)\ ,\end{aligned} \tag{50}$$

The triple $(B, \widetilde{C}, \nu)$ has the typical form of *modified semimartingale characteristics* (Definition II.2.16. in Jacod & Shiryaev (2003)). We assume that

A5. The martingale problem associated with $(B, \widetilde{C}, \nu)$ and initial distribution π_0 is well-posed, i.e. there is a unique probability measure $\bar{R}$ on Ω^d such that $\bar{X}$ is a semimartingale with modified characteristics $(B, \widetilde{C}, \nu)$ and initial law π_0.

Furthermore, denote as in Section 4.1 by $(\Omega_2, \mathcal{F}_2, P^{0,l})$ the l-dimensional Wiener space with coordinate process $\boldsymbol{\beta}^0$ and let $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{R}^0)$ be the product space

$$(\bar{\Omega}, \bar{\mathcal{F}}, \bar{R}^0) = (\Omega^d \times \Omega_2, \mathcal{F}^d \otimes \mathcal{F}_2, \bar{R} \times P^{0,l})\,. \tag{51}$$

According to the results from Section 4 (see, in particular, (18), (19) as well as Theorems 4.1, 4.3 and Proposition 4.2) we obtain that, for a given time t, for a given default history $\{(\hat{T}_n, \hat{\xi}_n) \colon n \le \hat{N}_t\}$, and for a bounded and continuous function $f : S^{\mathbf{X}} \to \mathbb{R}$, the filter $\pi_t f$ can be expressed in the form

$$\pi_t f(\omega_2) \propto E^{\bar{R}}\left(f(\bar{\mathbf{X}}_t) L_t^1 L_t^2(\cdot, \omega_2)\right) \tag{52}$$

where

$$L_t^1 = \prod_{n=1}^{\hat{N}_t} \left\{\lambda_{\hat{\xi}_n}(\bar{\mathbf{X}}_{\hat{T}_n^-}, \hat{\mathbf{Y}}_{\hat{T}_{n-1}}) \exp\Big(-\int_{\hat{T}_{n-1}}^{\hat{T}_n} \bar{\lambda}(\bar{\mathbf{X}}_s, \hat{\mathbf{Y}}_{\hat{T}_{n-1}})ds\Big)\right\} \exp\Big(-\int_{\hat{T}_{N_t}}^t \bar{\lambda}(\bar{\mathbf{X}}_s, \hat{\mathbf{Y}}_{\hat{T}_{N_t}})ds\Big), \tag{53}$$

and where, in analogy with (17) or (19),

$$L_t^2 = \exp\Big\{\int_0^t \boldsymbol{a}_s'(\bar{\mathbf{X}}_s)\Sigma_{\mathbf{Z}}^{-1} d\mathbf{Z}_s - \frac{1}{2}\int_0^t \left\|\boldsymbol{a}_s(\bar{\mathbf{X}}_s)\right\|^2_{\Sigma_{\mathbf{Z}}^{-1}}\, ds\Big\} \text{ for } \mathbf{Z}_t = v\boldsymbol{\beta}_t^0. \tag{54}$$

Approximating filter sequence. Consider a sequence of processes $(\mathbf{X}^m, \mathbf{Y}^m)_{m\in\mathbb{N}}$, solving the SDE system

$$\mathbf{X}_t^m = \mathbf{X}_0^m + \int_0^t \int_E K^{\mathbf{X},m}(\mathbf{X}^m_{s-}, \mathbf{Y}^m_{s-}, u)\mathcal{N}(ds, du) \tag{55}$$

$$Y^m_{t,j} = Y_{0,j} + \int_0^t \int_E (1 - Y^m_{s-,j}) K_j^{\mathbf{Y}}(\mathbf{X}^m_{s-}, \mathbf{Y}^m_{s-}, u)\mathcal{N}(ds, du). \tag{56}$$

Note that $K_j^{\mathbf{Y}}$ is independent of m. We assume that the kernels $K^{\mathbf{X},m}(\boldsymbol{x}, \boldsymbol{y}, u)$ are of the form as discussed in Section 5.1, so that $\mathbf{X}^m$ is in fact a finite state Markov chain with state space given by a finite set $S^{\mathbf{X},m} \subset \mathbb{R}^d$.

Given the default observation $(\hat{T}_n, \hat{\xi}_n)$, $n = 1, \ldots, \hat{N}_t$ respectively $(\hat{\mathbf{Y}}_s)_{s\le t}$, introduce the modified semimartingale characteristics $(B^m, \widetilde{C}^m, \nu^m)$ given by (48), (49) and (50) with $b = \Sigma = 0$ and $\bar{K}^{\mathbf{X},m}$ instead of $\bar{K}^{\mathbf{X}}$. Choose a sequence of initial distributions π_0^m on the state space of

$\mathbf{X}^m$ such that π_0^m converges weakly to π_0 and denote by the measure $\bar{R}^m$ on $(\Omega^d, \mathcal{F}^d, (\mathcal{F}_t^d))$ the solution of the martingale problem associated with the characteristics $(B^m, \widetilde{C}^m, \nu^m)$ and the initial distribution π_0^m (we assume that this martingale problem is well-posed for every m). The filter $\pi_t^m f$ in the approximating model can be expressed in the form

$$\pi_t^m f(\omega_2) \propto E^{\bar{R}^m} \left(f(\bar{\mathbf{X}}_t) L_t^1 L_t^2(\cdot, \omega_2) \right) . \tag{57}$$

Below we shall give conditions on $(B^m, \widetilde{C}^m, \nu^m)$ so that the sequence of measures $\bar{R}^m$ converges weakly to $\bar{R}$ (denoted $\bar{R}^m \Rightarrow \bar{R}$) as $m \to \infty$. Assuming for a moment such a convergence, here we show first the ensuing convergence of the filters. We need the following additional assumption.

A6. The default intensities $\lambda_j(\boldsymbol{x}, \boldsymbol{y})$ are bounded and continuous in $\boldsymbol{x}$.

Theorem 6.1. *Fix some $t > 0$ and a default history $\{(\hat{T}_n, \hat{\xi}_n) \colon n = 1, \ldots, \hat{N}_t\}$. Suppose that Assumptions A1 to A6 hold and that $\bar{R}^m \Rightarrow \bar{R}$. Then $\pi_t^m f(\omega_2)$ converges $P^{0,l}$-stochastically to $\pi_t f(\omega_2)$, i.e. $P^{0,l} - \lim_{m\to\infty} \pi_t^m f = \pi_t f$.*

Remark 6.2. 1.) Although the convergence is in a weaker form than a.s. convergence, it still implies that, for m sufficiently large, the probability that $\pi_t^m f$ differs from $\pi_t f$ by a given amount can be made arbitrarily small.
2.) Note that Theorem 6.1 is a filter-approximation result for a model where signal and observation cannot be made independent via a measure transformation. This sets the result apart from filter approximation results as in Zeng (2003) which are based on general results T. Kurtz and E. Goggins concerning weak convergence of conditional expectations. By the same token, Zeng (2003) obtains only weak convergence of the approximating filters, while here we obtain convergence in probability.

Proof of Theorem 6.1. Denote by $\delta_d(\boldsymbol{x}, \boldsymbol{y})$ the Prokhorov metric on $D^d([0,\infty))$ (see Jacod & Shiryaev (2003), Chapter VI, (1.26)). By Skorokhod embedding, the weak convergence $\bar{\mathbb{R}}^m \Rightarrow \bar{R}$ implies that there is some probability space - denoted again by $(\Omega^d, \mathcal{F}^d, \bar{R})$ for simplicity - and processes $\bar{\mathbf{X}}^m$ and $\bar{\mathbf{X}}$ with laws $\bar{R}^m$ and $\bar{R}$ respectively such that $\lim_{m\to\infty} \delta_d(\bar{\mathbf{X}}^m, \bar{\mathbf{X}}) = 0$, $\bar{R}$ a.s. and hence also $\bar{R}^0 = \bar{R} \times P^{0,l}$ a.s. Now we have the following two Lemmas.

Lemma 6.3. *Consider bounded and continuous functions $f(\cdot)$ and $\lambda(\cdot)$ We get for processes $\bar{\mathbf{X}}^m, \bar{\mathbf{X}}$ as above that*

$$\lim_{m\to\infty} f(\bar{\mathbf{X}}_t^m) = f(\bar{\mathbf{X}}_t) \quad \bar{R}^0 - a.s. \tag{58}$$

$$\lim_{m\to\infty} \lambda(\bar{\mathbf{X}}^m_{\hat{T}_n^-}) = \lambda(\bar{\mathbf{X}}_{\hat{T}_n^-}), \quad \bar{R}^0 - a.s., \ \ n = 1, \ldots, \hat{N}_t \tag{59}$$

$$\lim_{m\to\infty} \int_{t_1}^{t_2} \bar{\lambda}(\bar{\mathbf{X}}_s^m, \hat{\mathbf{Y}}_s) ds = \int_{t_1}^{t_2} \bar{\lambda}(\bar{\mathbf{X}}_s, \hat{\mathbf{Y}}_s) ds \, ; \quad t_1 < t_2 < t; \ \bar{R}^0 - a.s. \tag{60}$$

Proof of Lemma 6.3. Note that by the form of the characteristics of $\bar{\mathbf{X}}$, for $t \notin J_t := \{\hat{T}_n : n = 1, \ldots, \hat{N}_t\}$ we have $\Delta \bar{\mathbf{X}}_t = 0$, $\bar{R}^0-$a.s. Given the $\bar{R}^0-$a.s. convergence of $\bar{\mathbf{X}}^m$ to $\bar{\mathbf{X}}$ in the Prokhorov metric, (58) follows from Proposition VI.2.1 (b.5) in Jacod & Shiryaev (2003). On the other hand, if $\nu(\{\hat{T}_n\} \times \mathbb{R}^d) > 0$ then also $\nu^m(\{\hat{T}_n\} \times \mathbb{R}^d) > 0$ for m sufficiently large, so that Proposition VI.2.1 (b.6) in Jacod & Shiryaev (2003) implies that for $\hat{T}_n \le t$ we have $\bar{R}^0-$a.s.

$$\lim_{m\to\infty} \bar{\mathbf{X}}^m_{\hat{T}_n^-} = \bar{\mathbf{X}}_{\hat{T}_n^-} \text{ and } \lim_{m\to\infty} \bar{\mathbf{X}}^m_{\hat{T}_n} = \bar{\mathbf{X}}_{\hat{T}_n}. \tag{61}$$

Relation (61) then implies (59) and, for $t \in J_t$, also (58). Relation (60) is obvious from the definition of the Skorokhod topology. $\square$

Lemma 6.4. *Let $L_t^{2,m}$ denote the process L_t^2 defined in (54), but with $\bar{\mathbf{X}}^m$ replacing $\bar{\mathbf{X}}$ there. Then, for processes $\bar{\mathbf{X}}^m, \bar{\mathbf{X}}$ as above, one has $\bar{R}^0 - \lim_{m\to\infty} L_t^{2,m} = L_t^2$.*

Proof of Lemma 6.4. We first consider the stochastic integral terms. Since $\mathbf{Z}_t = v\boldsymbol{\beta}_t^0$, we get

$$\begin{aligned} E^{\bar{R}^0}&\left(\left(\int_0^t \boldsymbol{a}_s'(\bar{\mathbf{X}}_s^m)\Sigma_{\mathbf{Z}}^{-1}d\mathbf{Z}_s - \int_0^t \boldsymbol{a}_s'(\bar{\mathbf{X}}_s)\Sigma_{\mathbf{Z}}^{-1}d\mathbf{Z}_s\right)^2\right) \\ &= E^{\bar{R}^0}\left(\int_0^t \left\|(\boldsymbol{a}_s(\bar{\mathbf{X}}_s^m) - \boldsymbol{a}_s(\bar{\mathbf{X}}_s))\right\|_{\Sigma_{\mathbf{Z}}^{-1}}^2 ds\right) \to 0 \quad \text{for } m\to\infty, \end{aligned} \tag{62}$$

where the convergence follows from the assumption of continuity and boundedness of the function $\boldsymbol{a}_t(\cdot)$ (see Assumption A4) as well as from bounded convergence. Next, always by the Assumption A4 as regards $\boldsymbol{a}_t(\cdot)$ as well as by the triangle inequality we also have, as $m\to\infty$,

$$\int_0^t \left(\left\|\boldsymbol{a}_s(\bar{\mathbf{X}}_s^m)\right\|_{\Sigma_{\mathbf{Z}}^{-1}}^2 - \left\|\boldsymbol{a}_s(\bar{\mathbf{X}}_s)\right\|_{\Sigma_{\mathbf{Z}}^{-1}}^2\right) ds \le 2\,\|\boldsymbol{a}\| \int_0^t \left\|(\boldsymbol{a}_s(\bar{\mathbf{X}}_s^m) - \boldsymbol{a}_s(\bar{\mathbf{X}}_s))\right\|_{\Sigma_{\mathbf{Z}}^{-1}} ds \to 0 \quad \bar{R}^0 - \text{a.s.} \tag{63}$$

Relations (62) and (63) imply immediately that for $m\to\infty$,

$$\int_0^t \boldsymbol{a}_s'(\bar{\mathbf{X}}_s^m)\Sigma_{\mathbf{Z}}^{-1}d\mathbf{Z}_{s^-} - \frac{1}{2}\int_0^t \left\|\boldsymbol{a}_s(\bar{\mathbf{X}}_s^m)\right\| 2_{\Sigma_{\mathbf{Z}}^{-1}} ds \xrightarrow{\bar{R}^0} \int_0^t \boldsymbol{a}_s'(\bar{\mathbf{X}}_s)\Sigma_{\mathbf{Z}}^{-1}d\mathbf{Z}_s - \frac{1}{2}\int_0^t \left\|a_s(\bar{\mathbf{X}}_s)\right\|_{\Sigma_{\mathbf{Z}}^{-1}}^2 ds$$

and therefore also $\bar{R}^0 - \lim_{m\to\infty} L_t^{2,m} = L_t^2$ for arbitrary fixed t. □

Now we return to the proof of Theorem 6.1. From the boundedness of f and $L^{1,m}$ (see A6), the definition of $L^{1,m}$ according to (53) (with $\bar{\mathbf{X}}^m$ instead of $\mathbf{X}$) and the fact that $E^{\bar{R}^0}\left((L^{2,m})^2\right) \le C < \infty$ (due to the boundedness of $\boldsymbol{a}_t(\cdot)$) we obtain uniform integrability for the sequence $(f(\bar{\mathbf{X}}_t^m)\,L_t^{1,m}L_t^{2,m})_{m\in\mathbb{N}}$. From Lemma 6.3 as well as from Lemma 6.4 we then obtain that

$$f(\bar{\mathbf{X}}_t^m)\,L_t^{1,m}L_t^{2,m} \to f(\bar{\mathbf{X}}_t)\,L_t^1L_t^2 \quad \text{in} \quad \mathcal{L}^1(\bar{\Omega}, \bar{\mathcal{F}}, \bar{R}^0),\ m\to\infty. \tag{64}$$

Using the product-form of $\bar{R}^0$, this $\mathcal{L}^1$-convergence can be written more explicitly as

$$\lim_{m\to\infty}\int_{\Omega_2} E^{\bar{R}}\left(\left|f(\bar{\mathbf{X}}_t^m)\,L_t^{1,m}L_t^{2,m}(\cdot,\omega_2) - f(\bar{\mathbf{X}}_t)\,L_t^1L_t^2(\cdot,\omega_2)\right|\right) P^{0,l}(d\omega_2) = 0. \tag{65}$$

Hence the inner expectation in (65) converges to zero in $\mathcal{L}^1(\Omega_2, \mathcal{F}_2, P^{0,l})$ and therefore also $P^{0,l}$-stochastically, which proves the theorem. □

Weak convergence. In the remaining part of this subsection, under a couple of additional assumptions on the model, we give conditions on the modified characteristics of $\bar{\mathbf{X}}^m$ for which we obtain weak convergence of $\bar{\mathbf{X}}^m$ to $\bar{\mathbf{X}}$. For this purpose we base ourselves on Theorems 3.35 and 2.11 in chapter IX of Jacod & Shiryaev (2003). The additional assumptions on the model are:

A7.1 *(a strengthening of A2).* There exists $A \subset E$ with A compact such that $D_i^{\mathbf{X}}(\boldsymbol{x},\boldsymbol{y}) \subset A$ for all $\boldsymbol{x},\boldsymbol{y}$ and all $i = 1,\cdots,d$.

A7.2 There exists a positive constant H such that for all $\boldsymbol{x},\boldsymbol{y}$, all $i = 1,\cdots,d$ and all $u \in E$,

$$\sum_{i=1}^d |\, b_i(\boldsymbol{x},\boldsymbol{y},u)\,| \le H\,; \quad \sum_{i=1}^d |\, \bar{K}_i^{\mathbf{X}}(\boldsymbol{x},\boldsymbol{y},u)\,| \le H\,; \quad \sum_{i=1}^d |\, \Sigma_{\mathbf{X}}^{ii}(\boldsymbol{x},\boldsymbol{y})\,| \le H\,.$$

Following Jacod & Shiryaev (2003) we introduce the following class of test functions: $\mathcal{C}_1 := \{\gamma : \mathbb{R}^d \to \mathbb{R} \colon \gamma \text{ continuous}\,, |\gamma(\boldsymbol{x})| < 1,\, \gamma(0) = 0\}$. Furthermore, for generic d, $\delta_d(\boldsymbol{x}, \boldsymbol{y})$ denotes the Prokhorov metric on the d-dimensional Skorohod space.

Proposition 6.5. *Let Assumption A1 to A7 hold and suppose that for the characteristics of the approximating Markov chains $\bar{\mathbf{X}}^m$ one has*

$$\begin{cases} \delta_d(B^m, B \circ \bar{\mathbf{X}}) & \xrightarrow{\bar{R}^m} 0\,, \\ \delta_{d^2}(\tilde{C}^m, \tilde{C} \circ \bar{\mathbf{X}}) & \xrightarrow{\bar{R}^m} 0\,, \\ \delta_1(\gamma * \nu^m, (\gamma * \nu) \circ \bar{\mathbf{X}}) & \xrightarrow{\bar{R}^m} 0 \text{ for all } \gamma \in \mathcal{C}_1\,. \end{cases} \tag{66}$$

where $\bar{R}^m$ is the sequence of measures making the coordinate process a semimartingale with modified characteristics $(B^m, \tilde{C}^m, \nu^m)$ and initial distribution π_0^m with $\pi_0^m \Longrightarrow \pi_0$. Then, for a given default sequence $(\hat{T}_n, \hat{\xi}_n)$, we have the weak convergence $\bar{R}^m \Longrightarrow \bar{R}$ for $m \to \infty$.

The conditions (66) can be taken as guidelines when choosing the approximating sequence of finite state Markov chains. While given in a somewhat abstract form here, they assume a more specific form for a given problem at hand.

The proof of the proposition is based on the following two lemmas.

Lemma 6.6. *Under the assumptions of Proposition 6.5 the sequence of measures $\bar{R}^m$, $m \in \mathbb{N}$, is tight.*

Proof. The proof is based on that for Theorem 3.35 in Chapter IX of Jacod & Shiryaev (2003). We show that, under our assumptions, Conditions i),ii),v) and vi) of that theorem are satisfied.

Condition i) *(Strong majorization hypothesis II)* can, on the basis of Assumption A7.1 and A7.2, be seen to be satisfied if in the definition of the strong majorization condition (Definition 3.11 in Chapter IX of Jacod & Shiryaev (2003)) one takes as deterministic increasing càdlàg functions the following:

$$F_t = [2H + H(1+H)F_{\mathcal{N}}(A)]\, t + H(H+1)\, \hat{N}_t \tag{67}$$

with A and H from Assumptions A7.1 and A7.2 respectively; for $\gamma \in \mathcal{C}_1$ we take

$$F_t^\gamma = F_{\mathcal{N}}(A)\, t + \hat{N}_t\,. \tag{68}$$

Recall that we work with a given observed sequence $(\hat{T}_n, \hat{\xi}_n)$, so that the functions F_t and F_t^γ are deterministic functions of time. The remaining conditions are immediate: Condition ii) *(Condition on the big jumps)* is automatically satisfied under the given assumptions; Condition v) holds by assumption on the initial conditions; Condition vi) corresponds to (66).

The statement of the lemma follows now from the first part of the proof of Theorem 3.35 in Chapter IX of Jacod & Shiryaev (2003), which in turn is based on Theorem 3.20 in the same chapter. ☐

Lemma 6.7. *Under the hypotheses of Proposition 6.5 every weakly converging sequence $\bar{\mathbf{X}}^m$ has as limit a semimartingale process with modified characteristics $(B, \tilde{C}, \nu)$*

Proof. Here we rely on Theorem 2.11 in Chapter IX of Jacod & Shiryaev (2003). Condition i) in that Theorem is satisfied since the assumption made by requiring (66) is stronger than this condition. Condition ii) *(majorization condition)* is here a rather immediate consequence of assumptions A7.1 andA7.2.

There remains Condition iii), namely the *(Continuity condition)* for the modified characteristics $(B, \tilde{C}, \nu)$ in the Skorokhod topology. This condition has to hold a.s. with respect to the limit measure, in our case $\bar{R}$. Next recall that each of the characteristics is composed of terms expressed as an integral with respect to time and one term given by a sum over $\hat{T}_n$. The time integrals are automatically continuous. The term given by the sum over $\hat{T}_n$ can on the other hand be treated by analogy to Lemma 6.3 noticing that we have $\lambda_j(\mathbf{X}_{t^-}, \hat{\mathbf{Y}}_{t^-}) = F_{\mathcal{N}}(D_j^{\mathbf{Y}}(\mathbf{X}_{t^-}, \hat{\mathbf{Y}}_{t^-})$ and for $\lambda_j(\cdot, \boldsymbol{y})$ we have the continuity assumption A6. Having now all conditions of Theorem 2.11 in Chapter IX of Jacod & Shiryaev (2003) satisfied, the statement here follows from that same Theorem. □

We can now conclude with the

Proof of Proposition 6.5. By the tightness result of Lemma 6.6 we have that every sequence $\bar{R}^m$ has a weakly converging subsequence. By Lemma 6.7 each of these converging sequences has a weak limit that corresponds to the same modified characteristics $(B, \tilde{C}, \nu)$, namely those of the original process $\bar{\mathbf{X}}$. By Assumption A5 the weak limit is unique and this gives us the statement of the proposition. □

References

Björk, T. (1998), *Arbitrage Theory in Continuous Time*, Oxford University Press, Oxford.

Ceci, C. & Gerardi, A. (2006), 'A model for high frequency data under partial information: a filtering approach', *International Journal of Theoretical and Applied Finance (IJTAF)* pp. 1–22.

Clark, J. (1978), The design of robust approximations to the stochastic differential equations of nonlinear filtering, *in* J. Skwirzynski, ed., 'Communication Systems and Random Process Theory', Sijthoff and Noordhoff, pp. 721–734.

Collin-Dufresne, P., Goldstein, R. & Helwege, J. (2003), 'Is credit event risk priced? modeling contagion via the updating of beliefs', Preprint, Carnegie Mellon University.

Cvitanic, J., Liptser, R. & Rozovski, B. (2006), 'A filtering approach to tracking volatility from prices observed at random times', *The Annals of Applied Probability* **16**, 1633–1652.

Cvitanic, J., Rozovski, B. & Zalyapin, I. (2006), 'Numerical estimation of volatility values from discretely observed diffusion data', *Journal of Computational Finance* **9**, 1–36.

Davis, M. (1978), Pathwise non-linear filtering, *in* M. Hazewinkel & J. Willems, eds, 'Stochastic Systems: The Mathematics of Filtering and Identification and Applications', D. Reidel Publishing Company, pp. 505–528.

Davis, M. & Lo, V. (2001), 'Infectious defaults', *Quant. Finance* **1**, 382–387.

Duffie, D. & Garleanu, N. (2001), 'Risk and valuation of collateralized debt obligations', *Financial Analyst's Journal* **57**(1), 41–59.

Duffie, D., Pan, J. & Singleton, K. (2000), 'Transform analysis and asset pricing for affine jump diffusions', *Econometrica* **68**, 1343–1376.

Elliott, R. (1993), 'New finite-dimensional filters and smoothers for noisily observed markov chains', *IEEE Trans. Info. theory* **IT-39**, 265–271.

Elliott, R., Jeanblanc, M. & Yor, M. (2000), 'On models of default risk', *Mathematical Finance* **10**, 179–195.

Frey, R., Prosdocimi, C. & Runggaldier, W. (2007), 'Affine credit risk models under incomplete information', *in* J. Akahori, S. Ogawa, S. Watanabe, eds., Stochastic Processes and Application to Mathematical Finance, World Scientific, pp. 97–113.

Giesecke, K. (2004), 'Correlated defaults with incomplete information', *Journal of Banking and Finance* **28**, 1521–1545.

Gombani, A., Jaschke, S. & Runggaldier, W. (2005), 'A filtered no arbitrage model for term structures with noisy data', *Stochastic Processes and Applications* **115**, 381–400.

Graziano, G. & Rogers, C. (2006), 'A dynamic approach to the modelling of correlation credit derivatives using Markov chains', working paper, Statistical Laboratory, University of Cambridge.

Grigelionis, B.I. (1972), 'On stochastic equations for nonlinear filtering problem of stochastic processes', *Lietuvos Matematikos Rinkinys* **12**, 37–51.

Jacod, J. & Shiryaev, A. (2003), *Limit Theorems for Stochastic Processes*, 2nd edn, Springer Verlag, Berlin.

Jazwinski, A. (1970), *Stochastic Processes and Filtering Theory*, Academic Press.

Kallianpur, G. & Striebel, C. (1968), 'Estimation of stochastic systems: arbitrary systems process with additive white noise observation errors', *Annals of Mathematical Statistics* **39**, 785–801.

Kliemann, W., Koch, G. & Marchetti, F. (1990), 'On the unnormalized solution of the filtering problem with counting process observations', *IEEE* **IT-36**, 1415–1425.

McNeil, A., Frey, R. & Embrechts, P. (2005), *Quantitative Risk Management: Concepts, Techniques and Tools*, Princeton University Press, Princeton, New Jersey.

Schönbucher, P. (2004), 'Information-driven default contagion', Preprint, Department of Mathematics, ETH Zürich.

Zeng, Y. (2003), 'A partially observed model for micromovement of asset prices with bayes estimation via filtering', *Mathematical Finance* **13**, 411–444.