

Credit risk and incomplete information: filtering and EM parameter estimation

Claudio Fontana and Wolfgang J. Runggaldier

Department of Pure and Applied Mathematics, University of Padova,
Via Trieste, 63, I-35121 Padova, Italy
e-mails: fontana@math.unipd.it, runggal@math.unipd.it

December 31, 2009

Abstract

We consider a reduced-form credit risk model where default intensities and interest rate are functions of a not fully observable Markovian factor process, thereby introducing an information-driven default contagion effect among defaults of different issuers. We determine arbitrage-free prices of *OTC* products coherently with information from the financial market, in particular yields and credit spreads and this can be accomplished via a filtering approach coupled with an EM-algorithm for parameter estimation.

1 Introduction

In recent years intensity-based models have become quite popular for the modeling of credit risk. In these models one specifies the default intensities that affect the default probabilities and the prices of credit derivatives. A critical point in this context is the modeling of the correlation among the default events of different issuers.

To this effect, we propose a model in an incomplete information framework, that is, we assume that both the default intensities and the interest rate are given as functions with a special structure of a multivariate, not fully observable stochastic factor process. These functions are in fact linear in the exponentials of the factors and with positive coefficients. Besides leading to positive rates and intensities, this special structure will turn out to be convenient for the filtering procedure to be developed below. The factors themselves may be very general: some may have an economic interpretation (macroeconomic, sectorial, idiosyncratic, etc.), other may not. Above all, they are not necessarily directly observable, namely they may be latent factors. Besides enhancing the model flexibility, latent factors can indeed have a meaningful economic interpretation as documented e.g. in Das et al. (2007), Duffie et al. (2009) and Bhar & Handzic (2008) where, on the basis of empirical evidence, it is shown that unobservable stochastic factor processes driving the default intensities are needed on top of observable covariates in order to explain clustering of defaults in historical data and large comovements of credit spreads. In addition, the formulation of a model under incomplete information has the advantage of avoiding a possibly inadequate specification of the factor model. This modeling approach generalizes the *frailty*-based approach proposed e.g. by Schönbucher (2003B), Duffie et al. (2009) and Azizpour & Giesecke (2008) in the sense that the factors can be seen as dynamic frailty variables which, especially when not fully observable, allow one to introduce an *information-driven* default contagion effect among the defaults of several issuers.

However, in this setting we have to face at least two important problems. The first one lies in the valuation of *Over the Counter (OTC)* products, the values of which can be expressed as functionals of the stochastic factors, but now some of them may not be observable/known. The second one is represented by the estimation of the parameters which characterize the model. Indeed, a standard calibration procedure, which is based on solving an inverse problem, could be quite difficult to implement under incomplete information; moreover, it may lead to unstable parameter values in the sense that it is strongly data-dependent.

The main contribution of our study is to propose a coherent and unifying methodology to solve the three problems: filtering the unobserved factors, estimating the parameters, pricing *OTC* products. The main tool is stochastic filtering that allows to effectively deal with these problems and leads to a more statistical type of parameter estimation as opposed to calibration.

More precisely, we propose a filtering approach to dynamically update the information on the unobserved factors, based on observations of market data. As market data we shall take, in addition to the default history, the short rate of interest and yields and credit spreads computed on default-free as well as defaultable bonds, which can be considered as meaningful representatives of more general market data. Indeed, we shall show that a large part of the market quotes for credit risky products can be expressed by means of these basic elements. Notice that the default indicator process is implicitly contained in the information coming from the credit spreads. An approach that is based exclusively on the information coming from the default indicator process can be found in Frey et al. (2006) (see also Schönbucher (2003B) in a related context). Since our observations are thus market quotes for credit risky products, on the other hand we are mainly interested in computing prices of *OTC* financial products, we shall formulate our filtering problem under a martingale/pricing measure and determine the prices of *OTC* products as conditional expected values of functions of the unobserved factors (thereby setting possibly observed factors equal to their observed values). Such prices will thus be coherent with the observed term structures, since the latter are the “input” of the filtering system. They can furthermore be shown to be arbitrage-free prices.

We shall deal in particular with the problem of parameter estimation and show how, by means of the *Expectation Maximization (EM)* algorithm, this problem is naturally linked to filtering. The estimates then evolve according to the filter solution and this allows the model to continuously adapt to the actual market situation. This fact is one of the major advantages of our methodology over the existing literature that deals with partially observed state-observation models where parameter estimation is typically done by maximizing the likelihood of the innovations obtained from the Extended Kalman Filter (EKF). Relevant contributions in this latter sense include Baadsgaard et al. (2000), Chen & Scott (2003), Duan & Simonato (1999), Geyer & Pichler (1999), Lund (1997) in the context of term structure models and, more recently, Bhar & Handzic (2008), Jacobs & Li (2008), Feldhütter & Lando (2008) in the context of credit risk. In addition, our results extend those of Elliott & Krishnamurthy (1999) and Elliott & Hyndman (2006) to models driven by a multivariate stochastic factor process with nonlinear dynamics.

In Section 2 we describe our model setup that we shall then cast in the context of incomplete information in Section 3, where we present also the filtering approach. Parameter estimation via the EM algorithm, which is related to the filtering approach, is then dealt with in Section 4, where we also describe an iterative algorithm for the actual estimation of the parameters. In Section 5 we then present some simulation results to show the performance of the algorithm. Finally, an Appendix contains two technical results (the proof of a Lemma and an additional Remark).

2 Complete information

2.1 Model setup

We consider a market with a certain number m of firms, each of which may default. Let $\tau_j > 0$ ($j = 1, \dots, m$) be the random time denoting the default of firm j and let $H_t^j := \mathbf{1}_{\{\tau_j \leq t\}}$ be the corresponding default indicator process that jumps from zero to one at $t = \tau_j$. The current default state of the firms is then described by $H_t := (H_t^1, \dots, H_t^m)$ and the default history up to time t is described by the filtration $(\mathcal{H}_t)_{0 \leq t \leq T^*}$ where T^* is a given horizon and $\mathcal{H}_t := \sigma\{H_s, s \leq t\}$. We consider an underlying probability space (Ω, \mathcal{G}, P) , equipped with a filtration $(\mathcal{G}_t)_{0 \leq t \leq T^*}$ such that $\mathcal{G} = \mathcal{G}_{T^*}$ and where, since here we consider mainly pricing of *OTC* products, the measure P will be supposed to be a martingale (pricing) measure for a numeraire given by the money market account $B(t) = B(0) \exp\left[\int_0^t r_s ds\right]$ with r_t the short rate of interest. The filtration $(\mathcal{G}_t)_{0 \leq t \leq T^*}$ is defined by $\mathcal{G}_t = \mathcal{F}_t \vee \mathcal{H}_t$, where $(\mathcal{F}_t)_{0 \leq t \leq T^*}$ is a given background filtration and $(\mathcal{G}_t)_{0 \leq t \leq T^*}$ represents the full information filtration to which all processes will be adapted. We also recall the definition of the (\mathcal{G}_t) -martingale intensity λ_t^j of the (\mathcal{G}_t) -stopping time τ_j (see McNeil et al. (2005), Definition 9.16) according to which λ_t^j has to be such that

$$M_t^j := H_t^j - \int_0^{t \wedge \tau_j} \lambda_s^j ds \quad (1)$$

is a (\mathcal{G}_t) -martingale.

We shall assume that the random default times τ_j are conditionally independent, doubly stochastic random times with respect to the background filtration $(\mathcal{F}_t)_{0 \leq t \leq T^*}$ (see e.g. McNeil et al. (2005), sections 9.2 and 9.6). Under these assumptions the (\mathcal{G}_t) -martingale intensity λ_t^j of τ_j is given by what is called its (\mathcal{F}_t) -conditional hazard rate process (in Bielecki et al. (2004) it is called (\mathcal{F}_t) -intensity of τ_j). To define such a process, and thus the (\mathcal{G}_t) -intensities, assume given a certain number n of stochastic factors. In many studies dealing with term structure modeling n is taken as $n = 3$ (see Baadsgaard et al. (2000), Chen & Scott (2003)) while more recent empirical studies from the credit risk literature suggest $n = 3$ (Bhar & Handzic (2008)), $n = 4$ (Jacobs & Li (2008)), or $n = 6$ (Feldhütter & Lando (2008)). The values of the factors are described by an n -dimensional process Ψ_t that is (\mathcal{F}_t) -adapted and that we assume satisfies the following dynamics

$$d\Psi_t = \text{diag}(e^{-\Psi_t}) \left[A e^{\Psi_t} + b - \frac{1}{2} \mathbf{1} \right] dt + \text{diag}(e^{-\frac{1}{2} \Psi_t}) dw_t \quad (2)$$

where w_t is an n -dimensional (\mathcal{F}_t, P) -Wiener process, $\text{diag}(V)$ denotes, for a given column vector V , a diagonal matrix with the elements of V on the diagonal, e^{Ψ_t} is the column vector with elements $e^{\Psi_t^i}, i = 1, \dots, n$, $A \in \mathbb{R}^{n \times n}$, $b \in \mathbb{R}^n$ and $\mathbf{1}$ is a column vector with entries equal to 1. The vector Ψ_t will later on be extended to include components that represent observation noises. Before specifying the assumptions on A and b , we make the following Remark.

Remark 1. Define

$$\Phi_t := \exp(\Psi_t) \quad (3)$$

with the meaning that $\Phi_t^i = \exp(\Psi_t^i)$ for $i = 1, \dots, n$; $0 \leq t \leq T^*$ and $\Phi_t := (\Phi_t^1, \dots, \Phi_t^n)'$. Then by Itô's formula we obtain:

$$d\Phi_t = (A\Phi_t + b) dt + \text{diag}(\sqrt{\Phi_t}) dw_t \quad (4)$$

namely Φ_t satisfies a multivariate CIR model in canonical form in the sense of Dai & Singleton (2000). The reason why we do not consider Φ_t itself as factors but rather $\Psi_t = \log(\Phi_t)$ will become apparent below, in particular from the expressions of the filter at a time of default (see Proposition 8) and the fact that we then do not have to impose nonnegativity constraints in the filter distribution for the factors.

In line with the so-called admissibility conditions (see Dai & Singleton (2000)), the Feller test for explosions (see Karatzas & Shreve (1991)) and condition A in Duffie & Kan (1996) we make now the following Assumption.

Assumption 2. With $A = (a_{ij})_{i,j=1,\dots,n}$, $b = (b_i)_{i=1,\dots,n}$ we require $a_{ij} \geq 0$ for $j \neq i$ and $b_i > \frac{1}{2}$, for $i = 1, \dots, n$.

Remark 3 (On the existence and uniqueness of a strong solution to the SDE (2)). Notice first that the drift and the diffusion terms in (2), being convex functions on \mathbb{R}^n , are also locally Lipschitz. It follows that if a (weak or strong) solution to (2) exists then it is also unique (in the sense of pathwise uniqueness, see e.g. Karatzas & Shreve (1991), Theorem 5.2.5). Therefore, since existence of a weak solution and pathwise uniqueness already imply existence and uniqueness of a strong solution, it is enough to show only the existence of a weak solution of (2). To this effect we assume that our probability space is rich enough to support an unique strong and strictly positive solution to (4). These requirements are satisfied by imposing the parameter restrictions specified in Assumption 2 (see Duffie & Kan (1996) and Dai & Singleton (2000)). Under these conditions, a (weak) solution to (2) can be obtained by applying Itô's formula to the logarithm of Φ_t .

We now specify the short rate of interest r_t and the default intensities λ_t^j as affine functions of the exponentials of the factors Ψ_t with coefficients that are positive so as to guarantee strict positivity of rates and intensities. More precisely

$$\begin{cases} r_t = a + be^{\Psi_t} = a + b\Phi_t \\ \lambda_t^j = c^j + d^j e^{\Psi_t} = c^j + d^j \Phi_t ; \quad j = 1, \dots, m \end{cases} \quad (5)$$

with a, c^j positive constants and b, d^j n -dimensional row vectors of positive constants. This setup allows for correlation between interest rate and default intensities, which (see Schönbucher (2003A)) is a desirable property for a stochastic credit risk model. It allows also for direct correlation among the intensities themselves; to this effect see e.g. Yu (2005) that points at the importance of having several common factors driving the intensities.

2.2 Prices of credit risky products

With the model of the previous subsection we are in the context of affine credit risk models (see e.g. Lando (1998), Duffie & Singleton (1999), Frey et al. (2006)), more precisely it is affine in e^{Ψ_t} , equivalently in Φ_t , and so we can obtain explicit formulae for the arbitrage-free prices of 0-coupon default-free bonds and 0-coupon 0-recovery defaultable bonds in a full information setting, that is, with full access to the information contained in the filtration $(\mathcal{G}_t)_{0 \leq t \leq T^*}$. In particular, for $t \leq T$, the price at time t of a default-free 0-coupon T -bond with unitary face value that is given by (E denotes expectation with respect to the measure P that we had assumed to be a martingale measure)

$$\Pi_{DF}(t, T) = E \left[e^{-\int_t^T r_s ds} | \mathcal{G}_t \right], \quad (6)$$

can be expressed as (see Duffie & Kan (1996))

$$\Pi_{DF}(t, T) = \exp \left[A(t, T) - B(t, T) e^{\Psi_t} \right] \quad (7)$$

where the scalar $A(\cdot, T)$ and vector $B(\cdot, T)$ satisfy each a well-known first order ordinary differential equation that for $B(\cdot, T)$ turns out to be of the Riccati type (the coefficients in these equations depend on the parameters in (2) and (5)). Analogously, the price of a generic 0-coupon defaultable bond with zero recovery, maturity T and unitary face value, which is given by (see McNeil et al. (2005), Theorem 9.23)

$$\Pi(t, T) = E \left[e^{-\int_t^T r_s ds} \mathbf{1}_{\{\tau > T\}} | \mathcal{G}_t \right] = \mathbf{1}_{\{\tau > t\}} E \left[e^{-\int_t^T (r_s + \lambda_s) ds} | \mathcal{F}_t \right] \quad (8)$$

with λ_t the default intensity of the issuer of the bond, can be expressed as

$$\Pi(t, T) = \mathbf{1}_{\{\tau > t\}} \exp \left[\tilde{A}(t, T) - \tilde{B}(t, T) e^{\Psi_t} \right] \quad (9)$$

where, again, $\tilde{A}(\cdot, T)$, $\tilde{B}(\cdot, T)$ satisfy each one a well-known first order ordinary differential equation that for $\tilde{B}(\cdot, T)$ is of the Riccati type and the coefficients in these equations depend again on the parameters in (2) and (5). Notice that the 0-coupon default free bonds and the 0-coupon 0-recovery defaultable bonds can be considered as “building blocks” for more complex instruments/products as will be shown below. By (7) we have that the yield (for $t < T$) of a 0-coupon default-free bond is given by

$$YL(t, T) := -\frac{1}{T-t} \log \Pi_{DF}(t, T) = -\frac{A(t, T)}{T-t} + \frac{B(t, T)}{T-t} e^{\Psi_t} \quad (10)$$

and results in an affine function of e^{Ψ_t} . Furthermore, by (7) and (9), the credit spread of a 0-coupon 0-recovery defaultable bond, computed with respect to a default free bond with the same face value and maturity T , is given (up to τ) by

$$\begin{aligned} CS(t, T) &:= -\frac{1}{T-t} \log \left[\frac{\Pi(t, T)}{\Pi_{DF}(t, T)} \right] \\ &= -\frac{1}{T-t} \left(\tilde{A}(t, T) - A(t, T) \right) + \frac{1}{T-t} \left(\tilde{B}(t, T) - B(t, T) \right) e^{\Psi_t}; \quad t < \tau \wedge T \end{aligned} \quad (11)$$

again an affine function of e^{Ψ_t} .

We conclude this subsection by showing that more general credit risky products, in particular *corporate bonds* and spreads of *Credit Default Swaps (CDS)*, can be expressed in terms of the prices $\Pi(t, T)$ of 0-coupon 0-recovery defaultable bonds. In fact, for corporate bonds and CDS the link between their values and the default event of a given firm is much clearer than for other products. As a preliminary to this purpose we show how the value of the *recovery payment* of a corporate bond or of a CDS can be expressed in terms of $\Pi(t, T)$. We first recall that a *recovery payment* consists of a non-negative (\mathcal{F}_t) -adapted process Z_t which represents the payoff given to the holder of the contract in case of default of the issuer (for a corporate bond) or of the reference entity (for a CDS) at the time τ of default, if τ occurs prior to the maturity T of the contract. Its arbitrage-free value at time $t < T$ is given by

$$R_t = E \left[\mathbf{1}_{\{\tau > t\}} e^{-\int_t^\tau r_u du} Z_\tau \mathbf{1}_{\{\tau \leq T\}} \middle| \mathcal{G}_t \right] \quad (12)$$

Expected values of the type (12) can be evaluated by using the next Lemma, of which the proof can be found in the Appendix. In the following we denote by P^T the forward measure associated to the maturity T , defined by $\frac{dP^T}{dP} = \frac{1}{B(T)\Pi_{DF}(0, T)}$, and we write $P_t^T(A) := P^T(A | \mathcal{G}_t)$ for any $A \in \mathcal{G}_T$.

Lemma 4. *Assuming (Recovery of Par scheme) that the recovery payment Z_τ at the default time τ is given by a fraction $\delta \in (0, 1)$ of the (unitary) nominal value, we have*

$$E \left[\mathbf{1}_{\{\tau > t\}} e^{-\int_t^\tau r_u du} \delta \mathbf{1}_{\{\tau \leq T\}} \middle| \mathcal{G}_t \right] = \delta \left(\mathbf{1}_{\{\tau > t\}} - \Pi(t, T) - \int_t^T \Pi(t, s) \left(\tilde{f}(t, s) - \tilde{h}(t, s) \right) ds \right) \quad (13)$$

where $\tilde{f}(t, T)$ denotes the defaultable instantaneous forward rate, defined on the set $\{\tau > t\}$ for any $T > t$ as $\tilde{f}(t, T) := -\frac{\partial}{\partial T} \log \Pi(t, T)$ and $\tilde{h}(t, T)$ is the forward hazard rate conditioned on survival, formally defined as $\tilde{h}(t, T) := \lim_{\Delta \searrow 0} \frac{P_t^T(T < \tau \leq T + \Delta | \tau > T)}{\Delta}$.

Furthermore, if r_t and λ_t follow independent stochastic processes, (13) simplifies to:

$$E \left[\mathbf{1}_{\{\tau > t\}} e^{-\int_t^\tau r_u du} \delta \mathbf{1}_{\{\tau \leq T\}} \middle| \mathcal{G}_t \right] = \delta \left(\mathbf{1}_{\{\tau > t\}} - \Pi(t, T) - \int_t^T \Pi(t, s) f(t, s) ds \right) \quad (14)$$

where $f(t, T)$, for $T > t$, denotes the default-free instantaneous forward rate.

Remark 5. Notice that also T -forward survival probabilities can be expressed in terms of bond prices. In fact, for any $T > t$:

$$\begin{aligned} P_t^T(\tau > T) &:= P^T(\tau > T | \mathcal{G}_t) = \mathbf{1}_{\{\tau > t\}} \frac{P^T(\tau > T | \mathcal{F}_t)}{P^T(\tau > t | \mathcal{F}_t)} = \mathbf{1}_{\{\tau > t\}} E^{P^T} \left[e^{-\int_t^T \lambda_u du} | \mathcal{F}_t \right] \\ &= \mathbf{1}_{\{\tau > t\}} \frac{E^P \left[e^{-\int_t^T (r_u + \lambda_u) du} | \mathcal{F}_t \right]}{E^P \left[e^{-\int_t^T r_u du} | \mathcal{F}_t \right]} = \frac{\Pi(t, T)}{\Pi_{DF}(t, T)} \end{aligned} \quad (15)$$

where the first two equalities follow from McNeil et al. (2005) (see Corollary 9.10 in Section 9.2.) and the last one from (6) and (8).

Considering now a **corporate bond** with a *Recovery of Par* scheme as above and with periodic coupon payments up to default of a fixed fraction c of the (unitary) face value at n time points t_i with $t \leq t_0 < t_1 < \dots < t_n \leq T$, one has that its value is given by

$$\begin{aligned} \Pi^{\delta, c}(t, T) &= \sum_{i=1}^n c(t_i - t_{i-1}) \Pi(t, t_i) + \Pi(t, T) + \\ &+ \delta \left(\mathbf{1}_{\{\tau > t\}} - \Pi(t, T) - \int_t^T \Pi(t, s) \left(\tilde{f}(t, s) - \tilde{h}(t, s) \right) ds \right) \end{aligned} \quad (16)$$

This expression simplifies in the case of independence of r_t and λ_t , as pointed out in Lemma 4. Moreover, an analogous simpler formula can be derived for the *Recovery of Treasury* scheme (see Fontana (2007)).

Concerning the **CDS spreads** consider first the value of the premium payment leg (for a premium S paid in arrears), for which one has immediately

$$V_P = \sum_{k=1}^n S(t_k - t_{k-1}) \Pi(t, t_k) \quad (17)$$

On the other hand, the value of the default protection leg (fraction $(1 - \delta)$ of the unitary face value) is by Lemma 4

$$\begin{aligned} V_D &= E \left[\mathbf{1}_{\{\tau > t\}} e^{-\int_t^\tau r_s ds} (1 - \delta) \mathbf{1}_{\{\tau \leq T\}} | \mathcal{G}_t \right] \\ &= (1 - \delta) \left(\mathbf{1}_{\{\tau > t\}} - \Pi(t, T) - \int_t^T \Pi(t, s) \left(\tilde{f}(t, s) - \tilde{h}(t, s) \right) ds \right) \end{aligned} \quad (18)$$

For the (fair) CDS spread it then follows (on $\{\tau > t\}$)

$$S = \frac{(1 - \delta) \left(1 - \Pi(t, T) - \int_t^T \Pi(t, s) \left(\tilde{f}(t, s) - \tilde{h}(t, s) \right) ds \right)}{\sum_{k=1}^n (t_k - t_{k-1}) \Pi(t, t_k)} \quad (19)$$

Notice that the *accrued payment* between the last premium payment date prior to default and the time of default may be neglected if the payment dates are close to one another (see details in Fontana (2007)). Finally, as with the corporate bond prices, also here an analogous simpler formula exists for the *Recovery of Treasury* scheme (see again Fontana (2007)) as well as when r_t and λ_t follow independent processes.

On one hand the representations (16) and (19) allow to reconstruct 0-coupon 0-recovery bond prices from corporate bond prices and CDS spreads and use these reconstructed values to determine, via the relation $CS(t, T) = -\frac{1}{T-t} \log \left[\frac{\Pi(t, T)}{\Pi_{DF}(t, T)} \right]$, observed values for $CS(t, T)$ that, within our modeling framework are considered as affine functions of e^{Ψ_t} (see (11)). Since later on we shall consider the observed values of $CS(t, T)$ mostly as reconstructed from corporate bond prices and CDS spreads, we shall consider these latter values as noisy perturbations of their true values.

On the other hand, since from (9) one can express $\Pi(t, T)$ as (nonlinear) function of the factors Ψ_t , (16) and (19) show that also corporate bond prices and CDS spreads can in our model be expressed as (nonlinear) functions of Ψ_t .

3 Incomplete information and filtering setup

3.1 Incomplete information

As already mentioned in the Introduction, some of the factors with an economic interpretation may be directly observable, others may not. Also the factors without economic interpretation are in general non observable. Without loss of generality we shall then assume in what follows that all the factors are latent factors that are not directly observable. On one hand this setup allows to consider what is called *information induced contagion* by letting the unobserved factors correspond to *dynamic frailty variables* (see Schönbucher (2003B), Duffie et al. (2009), Azizpour & Giesecke (2008)). On the other hand this implies that we cannot directly use formulas (7) and (9), and thus neither (16) nor (19) to compute prices of illiquid *OTC* bonds and related products.

Let now the flow of information, which is actually available to agents on the market, be characterized by a filtration $(\mathcal{Y}_t)_{0 \leq t \leq T^*}$ with $\mathcal{H}_t \subset \mathcal{Y}_t \subset \mathcal{G}_t$. While all the intensities λ_t^j will be considered as not directly observable, we shall however assume that r_t is observable (via a *proxy*), i.e. it is (\mathcal{Y}_t) -adapted and this implies by (5) that one can observe a linear combination of the exponentials of the factors. We suppose furthermore that at any time $t \leq T^*$ agents can observe a number p of yields $YL(t, T_i)$, ($i = 1, \dots, p$) corresponding to 0-coupon default-free bonds for p different maturities T_i , ($T_i > t$, $i = 1, \dots, p$), and a number q of credit spreads $CS(t, T_j)$, ($j = 1, \dots, q$) corresponding to 0-coupon 0-recovery defaultable bonds issued by a given firm for a set of q maturities T_j , ($T_j > t$, $j = 1, \dots, q$) (more generally, we could also consider several issuing firms, see the Example in Section 4.2). Notice that these latter observations take place only up to the (observed) default of the issuer and therefore contain also the information of the default history. In our context, we assume that a given firm defaults simultaneously on all the bonds it has issued and this affects furthermore all CDS for which it represents the reference entity.

Observe next that, by (10) and (11), in addition to r_t also yields and spreads are affine functions of e^{Ψ_t} . If then one has more observations than factors, namely if $1 + p + q \geq n$, then the factors can be reconstructed exactly from the observations. On the other hand, yields and spreads are in general not exactly observable. In fact, as mentioned at the end of subsection 2.2, they have mostly to be reconstructed from corporate bond prices and CDS spreads and, even if they could be observed directly, they are affected by bid-ask spreads, liquidity and tax effects, as well as other imprecisions. Following an approach proposed in Gombani et al. (2005), we introduce a number ℓ of further unobserved factors, on which r_t and λ_t^j do not depend, but which represent additive noise terms that affect the observations of $YL(t, T_i)$ and $CS(t, T_j)$. There is a good deal of flexibility in specifying these ℓ factors, apart from ensuring that the filtering problem defined below is non degenerate, namely that $n + \ell > p + q + 1$; it means that there are truly non observable factors. The noise factors may also be correlated among themselves, or with Ψ_t . Unless specified otherwise, for sake of simplicity in what follows we shall assume that they are simply independent (\mathcal{F}_t, P) -Wiener processes. Moreover, not all of the observations have to be affected by noise (r_t was in fact assumed to be exactly observable), we can also assume a specific noise term for each maturity/issuer. We define then by $\bar{\Psi}_t$ the vector of dimension ℓ given by the additional ℓ noise factors.

Summing up, the observations (interest rate, yields and credit spreads) are given by:

$$\begin{cases} r_t & = a + be^{\Psi_t} \\ YL(t, T_i) & = \alpha_t^i + \beta_t^i e^{\Psi_t} + \bar{\beta}_t^i \bar{\Psi}_t & i = 1, \dots, p \\ CS(t, T_j) & = \gamma_t^j + \delta_t^j e^{\Psi_t} + \bar{\delta}_t^j \bar{\Psi}_t & j = 1, \dots, q \end{cases} \quad (20)$$

where α_t^i, γ_t^j are deterministic functions of time, β_t^i, δ_t^j are n -dimensional and $\bar{\beta}_t^i, \bar{\delta}_t^j$ are ℓ -dimensional row vectors composed of deterministic functions of time that depend on the parameters in (2) and (5) to be “calibrated” to the market and $\bar{\Psi}_t$ is an ℓ -vector of standard Wiener processes.

We can also write more specifically

$$\mathcal{Y}_t := \mathcal{F}_t^Y \vee \mathcal{H}_t \quad \text{for } 0 \leq t \leq T^*$$

where $\mathcal{F}_t^Y := \sigma\{r_s, YL(s, T_i), CS(s, T_j) : s \leq t, i = 1, \dots, p; j = 1, \dots, q\}$ and notice that q is reduced every time there is a default. With the reduction of q also the number of noise factors may be correspondingly reduced depending on the specification of the latter.

3.2 The filtering problem

One of our purposes is to compute arbitrage-free prices of *OTC* financial products in the context of our incomplete information market model with information represented by the filtration $(\mathcal{Y}_t)_{0 \leq t \leq T^*}$. Since, as described in subsection 2.2, default-free and defaultable bond prices are building blocks of more general credit risk sensitive products and the former are, under full information concerning Ψ_t , given (see (7), (9)) as (exponentially) affine functions of e^{Ψ_t} , let us denote by $\Pi(t, T; \Psi_t)$ the generic expression that represents arbitrage-free prices of *OTC* credit risky products in the full-information filtration $(\mathcal{G}_t)_{0 \leq t \leq T^*}$.

Lemma 6. *Under the assumption that r_t is directly observable ($r_t \in \mathcal{Y}_t$), the arbitrage-free price corresponding to $\Pi(t, T; \Psi_t)$ in the investor filtration $(\mathcal{Y}_t)_{0 \leq t \leq T^*}$ is given by*

$$\hat{\Pi}(t, T) = E[\Pi(t, T; \Psi_t) \mid \mathcal{Y}_t] \quad \text{for } 0 \leq t \leq T \quad (21)$$

Proof. It suffices to show that $B^{-1}(t)\hat{\Pi}(t, T)$ is a (P, \mathcal{Y}_t) -martingale. Since $B^{-1}(t)\Pi(t, T; \Psi_t)$ is a (P, \mathcal{G}_t) -martingale we have, for $t_1 < t_2 \leq T$,

$$\begin{aligned} B^{-1}(t_1)\hat{\Pi}(t_1, T) &= E[B^{-1}(t_1)\Pi(t_1, T; \Psi_{t_1}) \mid \mathcal{Y}_{t_1}] \\ &= E[E[B^{-1}(t_2)\Pi(t_2, T; \Psi_{t_2}) \mid \mathcal{G}_{t_1}] \mid \mathcal{Y}_{t_1}] \\ &= E[B^{-1}(t_2)\Pi(t_2, T; \Psi_{t_2}) \mid \mathcal{Y}_{t_1}] = E[B^{-1}(t_2)E[\Pi(t_2, T; \Psi_{t_2}) \mid \mathcal{Y}_{t_2}] \mid \mathcal{Y}_{t_1}] \\ &= E[B^{-1}(t_2)\hat{\Pi}(t_2, T) \mid \mathcal{Y}_{t_1}] \end{aligned} \quad (22)$$

□

Notice that, besides being arbitrage-free, the prices $\hat{\Pi}(t, T)$ in (21) are also coherent with market data. In fact, they are functionals of the market data in \mathcal{Y}_t and, if $\Pi(t, T; \Psi_t)$ represents a yield or a spread, $\hat{\Pi}(t, T)$ coincides with $\Pi(t, T; \Psi_t)$ except for the observation noise (that may e.g. represent bid-ask spread).

In order to compute the expression in the right hand side in (21) we need the conditional distribution of Ψ_t , given \mathcal{Y}_t , which is what is called the *filter distribution* at time t and which will be discussed next.

Before coming to the filter distribution itself, notice that, by (5), the default intensities of the various firms are known up to the knowledge of Ψ_t . The filter distribution for Ψ_t will thus induce a continuously updated distribution for the various λ_t^j and one may consider as “filtered default intensity” the mean of this distribution (see also Remark 9.ii) below).

Coming now to the filter distribution itself, notice first that the observations in (20) are linear with respect to the the vector pair $(e^{\Psi_t}, \bar{\Psi}_t)$ determined by the unobserved factors and some of them may not be affected by any of the factors in $\bar{\Psi}_t$ that represent a noise term (or may be affected by identical noise terms). A direct formulation of the state-observation filtering system, that is, letting $[\Psi', \bar{\Psi}']'$ be the unobserved state, would cause the filtering problem to degenerate since then the values of the factors can be backed out from the observations. Such an approach was actually followed in the existing literature

(see e.g. Chen & Scott (1993), Pearson & Sun (1994), Duffie & Singleton (1997)) whereby the number of observables was postulated to exactly match the number of unobservable factors. In this paper we would like to consider a more general situation that admits observations of the type (20) and leads to a nondegenerate filter distribution. We point out that this then includes also models from the literature where all observations are supposed to be affected by noise (see Lund (1997), Duan & Simonato (1999), Geyer & Pichler (1999), Baadsgaard et al. (2000), Chen & Scott (2003)).

Our purpose can now be achieved due to the fact that it is possible to reduce the extended factor vector $[\Psi', \bar{\Psi}']'$ to an auxiliary unobserved state process X_t of dimension $(n + \ell) - (p + q + 1)$ and obtain a corresponding non-degenerate filtering problem that is as informative as the original problem with state Ψ_t . To this effect we follow Frey & Runggaldier (2009), section 3.2 (see also Fontana (2007)), letting $Y_t := [r_t, Y_t^1, \dots, Y_t^{p+q}]$ where $Y_t^i := YL(t, T_i)$ for $i = 1, \dots, p$ and $Y_t^{p+j} := CS(t, T_j)$ for $j = 1, \dots, q$. With the so defined process Y_t , and recalling (3) whereby for the original factors Ψ_t we have $\Psi_t^i = \log(\Phi_t^i)$, $i = 1, \dots, n$, one can rewrite system (20) as

$$Y_t = \mu_t + M_t \Phi_t^* \quad \text{with} \quad \Phi_t^* := \left(\Phi_t', \bar{\Psi}_t^1, \dots, \bar{\Psi}_t^\ell \right)' \quad (23)$$

for an appropriate $(p + q + 1)$ -dimensional time varying vector μ_t and a time varying $(p + q + 1, n + \ell)$ -matrix M_t which, recalling that $n + \ell > p + q + 1$ and considering only nonredundant observations in (20), can be assumed to be of full rank. We can now prove the following

Proposition 7. *One can choose a time-varying $(n + \ell - 1 - p - q, n + \ell)$ -matrix L_t such that the $(n + \ell)$ -square matrix $\begin{pmatrix} L_t \\ M_t \end{pmatrix}$ is invertible and this choice is not necessarily unique. Assume furthermore that $L_t \Phi_t^*$ is P -a.s. strictly positive for all t . Defining then a process $X_t = (X_t^1, \dots, X_t^{(n+\ell)-(p+q+1)})$ by*

$$e^{X_t^i} := L_t^{i,\cdot} \Phi_t^* \quad i = 1, \dots, (n + \ell) - (p + q + 1) \quad (24)$$

we have that, for appropriate matrices $\hat{\Gamma}_t, \hat{\Delta}_t$

$$\Phi_t^* = \hat{\Gamma}_t e^{X_t} + \hat{\Delta}_t (Y_t - \mu_t) \quad (25)$$

and

$$\Phi_t = \Gamma_t e^{X_t} + \Delta_t (Y_t - \mu_t) \quad (26)$$

where Γ_t, Δ_t are the submatrices of $\hat{\Gamma}_t, \hat{\Delta}_t$ formed by their first n rows (that correspond to Φ as subvector of Φ^*). Furthermore, the pair (X_t, Y_t) satisfies a system of the form

$$\left\{ \begin{array}{l} dX_t^i = \left[\frac{G_t^{1,i}(X_t, Y_t)}{G_t^{2,i}(X_t, Y_t)} + \frac{1}{2} \frac{G_t^{3,i}(X_t, Y_t)}{(G_t^{2,i}(X_t, Y_t))^2} \right] dt + \sum_{k=1}^n \frac{L_t^{i,k} \sqrt{\Gamma_t^{k,\cdot} e^{X_t} + \Delta_t^{k,\cdot} (Y_t - \mu_t)}}{G_t^{2,i}(X_t, Y_t)} dw_t^k + \\ \quad + \sum_{j=1}^{\ell} \frac{L_t^{i,n+j}}{G_t^{2,i}(X_t, Y_t)} d\bar{\Psi}_t^j \quad i = 1, \dots, n + \ell - (p + q + 1) \\ dY_t = (R_t + C_t e^{X_t} + K_t Y_t) dt + \sum_{k=1}^n M_t^{k,\cdot} \sqrt{\Gamma_t^{k,\cdot} e^{X_t} + \Delta_t^{k,\cdot} (Y_t - \mu_t)} dw_t^k + \sum_{j=1}^{\ell} M_t^{n+j,\cdot} d\bar{\Psi}_t^j \end{array} \right. \quad (27)$$

where Γ_t, Δ_t are as in (26), $G_t^{k,i}(X_t, Y_t)$, $k = 1, 2, 3$, are expressions of the form

$$G_t^{k,i}(X_t, Y_t) = R_t^{k,i} + C_t^{k,i} e^{X_t} + K_t^{k,i} Y_t$$

with $R_t^{k,i}, C_t^{k,i}, K_t^{k,i}$ (as well as R_t, C_t, K_t) appropriate constants and vectors (or matrices) and where, we recall, $\bar{\Psi}_t^i$, $i = 1, \dots, \ell$, are standard (\mathcal{F}_t, P) -Wiener processes.

Proof. For an arbitrary matrix A , let us denote by $A^{i\cdot}$ and $A^{\cdot j}$ the i -th row and the j -th column respectively. Similarly, $A^{\cdot[1:n]}$ denotes the first n columns and $A^{i,[1:n]}$ the first n elements of the i -th row. Since M_t is full rank, a matrix L_t can always be found such that $\begin{pmatrix} L_t \\ M_t \end{pmatrix}$ is invertible. Having chosen L_t , by (23) and (24) we have for appropriate matrices $\hat{\Gamma}_t, \hat{\Delta}_t$

$$\Phi_t^* = \begin{pmatrix} L_t \\ M_t \end{pmatrix}^{-1} \begin{pmatrix} e^{X_t} \\ Y_t - \mu_t \end{pmatrix} =: \hat{\Gamma}_t e^{X_t} + \hat{\Delta}_t (Y_t - \mu_t) \quad (28)$$

which proves (25) and (26). The system (27) follows by a direct application of Itô's formula to (23) and (24), whereby

$$\begin{aligned} dX_t^i &= \frac{1}{L_t^{i\cdot} \Phi_t^*} \left(\frac{d}{dt} L_t^{i\cdot} \Phi_t^* dt + L_t^{i,[1:n]} (A\Phi_t + b) dt + \sum_{k=1}^n L_t^{i,k} \sqrt{\Phi_t^k} dw_t^k + \sum_{j=1}^{\ell} L_t^{i,n+j} d\bar{\Psi}_t^j \right) + \\ &\quad - \frac{1}{2 \left(L_t^{i\cdot} \Phi_t^* \right)^2} \left(\sum_{k=1}^n \left(L_t^{i,k} \right)^2 \Phi_t^k + \sum_{j=1}^{\ell} \left(L_t^{i,n+j} \right)^2 \right) dt, \quad \text{for } i = 1, \dots, (n + \ell) - (p + q + 1) \\ dY_t &= \left(\frac{d}{dt} \mu_t + \frac{d}{dt} M_t \Phi_t^* \right) dt + M_t^{\cdot[1:n]} (A\Phi_t + b) dt + \sum_{k=1}^n M_t^{\cdot,k} \sqrt{\Phi_t^k} dw_t^k + \sum_{j=1}^{\ell} M_t^{\cdot,n+j} d\bar{\Psi}_t^j \end{aligned} \quad (29)$$

which both are of the type (27), due to (25) and (26). \square

System (27) is a nondegenerate nonlinear filter system to which we shall apply the well-known Extended Kalman Filter (EKF), leading to a Gaussian filter distribution for the factors. At this point one can notice that one of the advantages of having chosen as factors Ψ_t rather than Φ_t is that Ψ_t does not have to be restricted to take only positive values as is done in most of the existing related literature. As mentioned in the Introduction, the Gaussianity of the filter is the tool that allows us to treat in a unifying manner our three basic problems: filtering, parameter estimation via the EM method (that is rather intimately related to filtering) and pricing as we shall see shortly. We shall now prove that our choice of model for the factor dynamics allows us to have Gaussianity also of the outgoing filter at a default time provided the incoming filter is Gaussian with the latter property guaranteed by using the EKF. No filter approximation will thus be needed at a default time. The EKF leads to the only approximation in the paper but on one hand it is known that it is reliable (in fact, it has been extensively adopted in the literature), on the other hand the goodness of the approximation will also be verified by our numerical results below.

Define the filtration $(\tilde{\mathcal{Y}}_t)_{0 \leq t \leq T^*}$ by $\tilde{\mathcal{Y}}_t := \mathcal{F}_t^Y \vee \sigma \{ \tau^j : \tau^j < t; j = 1, \dots, m \}$ for $0 \leq t \leq T^*$. While between default times we have $p_{X_t | \tilde{\mathcal{Y}}_t} = p_{X_t | \mathcal{Y}_t}$, at a default time the conditional distribution of the unobserved state X_t has to be updated on the basis of the information of the default event itself and of the identity of the defaulted firm. Notice first that, from (5) and (26), the default intensity of firm j can be written as

$$\lambda_t^j(X_t, Y_t) = c^j + d^j (\Gamma_t e^{X_t} + \Delta_t (Y_t - \mu_t)) =: c_0^j(t) + \sum_{i=1}^{\tilde{n}} c_i^j(t) e^{X_t^i} \quad (30)$$

thereby implicitly defining the functions $c_i(t)$, $i = 0, \dots, \tilde{n} := (n + \ell) - (p + q + 1)$. Suppose now to be at time $t = \tau_j$ and to observe the default of firm j . The update of the conditional/filter distribution is then (see Frey & Runggaldier (2009), Theorem 6.4, for a more detailed justification)

$$p_{X_t | \mathcal{Y}_t}(dx) = \frac{\lambda_t^j(x, Y_t) p_{X_t | \tilde{\mathcal{Y}}_t}(dx)}{\int \lambda_t^j(x, Y_t) p_{X_t | \tilde{\mathcal{Y}}_t}(dx)} \quad \text{for } t = \tau_j \quad (31)$$

We can now show

Proposition 8. *Suppose that at a generic default time $t = \tau^j$ the incoming distribution is $p_{X_t|\mathcal{Y}_t} \sim \mathcal{N}(\mu_t, \Sigma_t)$. Then if the defaulted firm is the j -th firm, the outgoing distribution $p_{X_t|\mathcal{Y}_t}$ is a mixture of \tilde{n} Gaussian distributions having a moment generating function given by*

$$\int e^{u'x} p_{X_t|\mathcal{Y}_t}(dx) = \frac{c_0^j(t)}{c_0^j(t) + \sum_{i=1}^{\tilde{n}} c_i^j(t) e^{\mu_t^i + \frac{1}{2}(\sigma_t^i)^2}} \exp\left[u'\mu_t + \frac{1}{2}u'\Sigma_t u\right] + \sum_{i=1}^{\tilde{n}} \frac{c_i^j(t) \exp\left[\mu_t^i + \frac{1}{2}(\sigma_t^i)^2\right]}{c_0^j(t) + \sum_{i=1}^{\tilde{n}} c_i^j(t) e^{\mu_t^i + \frac{1}{2}(\sigma_t^i)^2}} \exp\left[u'(\mu_t + \Sigma_t e_i) + \frac{1}{2}u'\Sigma_t u\right] \quad (32)$$

where e_i is the i -th unit \tilde{n} -vector, μ_t^i and σ_t^i are the marginal i -th mean and variance corresponding to $\mathcal{N}(\mu_t, \Sigma_t)$ and $c_i(t)$, $i = 0, \dots, \tilde{n}$, are the functions implicitly defined in (30).

Proof. First notice that the denominator in (31) is given by

$$E_{\mathcal{N}(\mu_t, \Sigma_t)} \left[c_0^j(t) + \sum_{i=1}^{\tilde{n}} c_i^j(t) e^{x^i} \right] = c_0^j(t) + \sum_{i=1}^{\tilde{n}} c_i^j(t) \exp\left[\mu_t^i + \frac{1}{2}(\sigma_t^i)^2\right] \quad (33)$$

with obvious meaning for the symbol $E_{\mathcal{N}(\mu_t, \Sigma_t)}$. We next compute the numerator of $\int e^{u'x} p_{X_t|\mathcal{Y}_t}(dx)$. Using (30) and (31) we have that it is

$$\int \left(c_0^j(t) + \sum_{i=1}^{\tilde{n}} c_i^j(t) e^{x^i} \right) e^{u'x} \mathcal{N}(dx, \mu_t, \Sigma_t) = c_0^j(t) \exp\left[u'\mu_t + \frac{1}{2}u'\Sigma_t u\right] + \sum_{i=1}^{\tilde{n}} c_i^j(t) \exp\left[(u + e_i)'\mu_t + \frac{1}{2}(u + e_i)'\Sigma_t(u + e_i)\right] \quad (34)$$

Multiplying with the denominator as given in (33) we obtain the result. \square

Remark 9. i) While Gaussianity is now preserved at a default time, the price to be paid for it is that for each incoming Gaussian distribution the outgoing distribution is a mixture of \tilde{n} Gaussian distributions so that in the period up to the next default \tilde{n} parallel filters have to be run, each with an initial distribution given by one of the Gaussian components of the mixture.

ii) While between default times one has a continuous update of the “filtered default intensities”, these undergo a jump at a default time (see Figure 1 in Section 5) with the size of the jump depending on the riskiness of the defaulted firm, as measured e.g. by the parameters c^j and d^j in (5), and causing what is termed *information induced default contagion*.

Once we know the filter distribution of X_t , both between and at default times, we can solve the pricing problem by computing the expression in (21), more precisely, by computing

$$\begin{aligned} \hat{\Pi}(t, T) &= E \left[\Pi(t, T; \Gamma_t e^{X_t} + \Delta_t(Y_t - \mu_t)) \mid \mathcal{Y}_t \right] \\ &= \int \Pi(t, T; \Gamma_t e^x + \Delta_t(Y_t - \mu_t)) p_{X_t|\mathcal{Y}_t}(dx) \end{aligned}$$

4 Parameter estimation via the EM algorithm

4.1 Introductory description

Including parameters in a market model makes it more flexible and allows for its calibration to the market. In the context of our model the parameters are given by the coefficients in the factor dynamics (2) and in the affine representation of interest rate and default intensities in (5). Traditionally, parameters are calibrated by solving an inverse problem resulting from trying to match theoretical model prices with actually observed ones. The filtering approach to incomplete information allows for a more statistical type of parameter estimation that can be incorporated into the filtering procedure. Besides the so-called *combined filtering and parameter estimation*, an approach in this context is based on the EM-algorithm (for a general description see e.g. McLachlan & Krishnan (1997)) that we are going to discuss next. Notice that, even if we assume our parameters to be constant with respect to time t , through their continuous updating via the filter-based EM algorithm below, they acquire a form of time dependence thereby allowing to track the actual market situation unlike static calibration or maximum likelihood techniques.

Let θ be the vector composed of all the parameters characterizing the model. Denote by P^θ the probability measure induced by the model for a given parameter vector θ . We thus have a family of probability measures $\{P^\theta; \theta \in \Theta\}$ and we assume them all to be absolutely continuous with respect to a given reference measure \bar{P} . The EM algorithm is now based on the iterative maximization with respect to θ , for fixed θ' (the previously estimated vector of parameters), of the following expression

$$Q(\theta, \theta') := E_{\theta'} \left[\log \frac{dP^\theta}{dP^{\theta'}} \middle| \mathcal{F}_t^Y \right] \quad (35)$$

It iterates through the two steps

- i) Compute the function $Q(\theta, \theta')$ for given θ' (a conditional expected value)
- ii) Maximize $Q(\theta, \theta')$ with respect to θ

until a stopping criterion is met. The maximization step leads to a system of equations obtained from putting $\frac{\partial Q(\theta, \theta')}{\partial \theta} = 0$.

With the function $Q(\theta, \theta')$ as given in (35) it can be shown (see e.g. Chapter 3 in McLachlan & Krishnan (1997)) that, with every iteration, the values of θ obtained from the EM algorithm generate non-decreasing values of the (conditional) log-likelihood function

$$l(\theta) = \log E_{\bar{P}} \left[\frac{dP^\theta}{d\bar{P}} \middle| \mathcal{F}_t^Y \right] \quad (36)$$

and thus also of the likelihood function $L(\theta) = \exp\{l(\theta)\}$. Under appropriate regularity conditions (see Wu (1983), Theorem 5), which in our case are satisfied, it follows that the sequence of EM-estimates converges to a stationary point of the likelihood function $L(\theta) = \exp\{l(\theta)\}$. To avoid ending up in a local maximum, one may adapt various techniques that have been proposed to this effect such as regularization techniques (see e.g. Cont & Tankov (2004)); the most widely used practical rule consists in starting the algorithm from different initial conditions.

Our filtering problem concerns the state-observation system (27) but a direct application of the EM algorithm to the estimation of the parameters in (27) leads to the following problems:

- i) The diffusion coefficients cannot be estimated via EM since the measures P^θ and $P^{\theta'}$ have to be mutually absolutely continuous; these coefficients have thus to be estimated for instance by computing the empirical quadratic variation of Y_t and by determining the filtered estimate of the quadratic variation of X_t (equivalently, by determining the filtered estimate of the quadratic covariation of (X_t, Y_t)) (see e.g. Elliott & Krishnamurthy (1997), Fontana (2007)).

- ii) To compute the function $Q(\theta, \theta')$ (see (55) below) one needs to solve also the so-called *smoothing problem*, for instance computing $E[X_s | \mathcal{F}_t^Y]$ for $s < t$, and this is computationally more demanding than just the filtering problem.
- iii) A typical difficulty in the application of the EM algorithm, particularly in the maximization step, is caused by the fact that there may be many parameters to be estimated.

To overcome these difficulties, an alternating iterative algorithm has been proposed in Fontana (2007) and in the next subsection we are going to illustrate it at hand of an example. Since data are actually observed in discrete time, the starting point of this algorithm is a time discretization of the Euler-Maruyama type of the state-observation system (27). This leads to a consistent discretization in the sense that, as is well known, the interpolated solutions of the so discretized system converge weakly to the solution of the continuous-time system. In the time discretized model the diffusion coefficients can then be estimated by the EM algorithm and this allows already to overcome problem i). Also problem ii) can be overcome since, by using an extension of results from Elliott & Krishnamurthy (1999) to discrete time linear conditionally Gaussian models, the parameter estimates can be computed directly as functions of the filter solution. Finally, as described below for the example, we deal with iii) by alternating between several parallel state-observation systems and estimating only a subset of the components of θ for each of them. Notice that, in general, the difficulty related to iii) may be reduced by trying to combine traditional calibration with EM parameter estimation. For instance, certain parameters may be individually calibrated to prices of call and put options on default free bonds leaving only the remaining parameters to be estimated via EM.

4.2 Example

The main purpose of this example is to illustrate in a simple context the discrete time alternating algorithm that we are going to propose for the estimation of the parameters characterizing the model. We consider the case of a model as described above (Sections 2 and 3) where $n = 3, p = 1, q = 2, \ell = 2$, that is, we assume to observe (in noise), besides the interest rate and a default-free yield, the spread of two defaultable bonds issued by two different firms that we shall denote by A and B respectively. The yield and credit spreads are supposed to correspond to bonds of the same maturity T .

The components Ψ_t^i ($i = 1, 2, 3$) of the vector Ψ_t in (2) are supposed to satisfy

$$d\Psi_t^i = \left[a^i + e^{-\Psi_t^i} \left(b^i - \frac{1}{2} \right) \right] dt + e^{-\frac{1}{2}\Psi_t^i} dw_t^i \quad (37)$$

i.e. we consider the individual Ψ_t^i to form independent processes. In this specific case the Assumption 2 of admissibility is fulfilled by simply requiring $b^i > \frac{1}{2}$ ($i = 1, 2, 3$). The additional $\ell = 2$ components of the factor vector $\bar{\Psi}_t^j$ ($j = 1, 2$) are

$$\bar{\Psi}_t^j = \bar{w}_t^j \quad (38)$$

for independent (\mathcal{F}_t, P) -Wiener processes \bar{w}_t^j ($j = 1, 2$). Furthermore, we let (5) take the specific form

$$\begin{cases} r_t &= \Phi_t^1 + \Phi_t^2 \\ \lambda_t^A &= \lambda^A (\Phi_t^1 + \Phi_t^3) \\ \lambda_t^B &= \lambda^B (\Phi_t^2 + \Phi_t^3) \end{cases} \quad (39)$$

with $\lambda^A, \lambda^B > 0$. With this representation one obtains a direct dependence between interest rate and default intensities as well as between the default intensities themselves.

In order to write down the filtering system (27) for the specific case of the example, we need to specify a process X_t making explicit its dynamics as well as those of the observations. The observation dynamics in (27) derive, via (23) and (25), from (20) so that we obtain explicit dynamics for the observations once we have explicit coefficients in (20). Since in the example we have three independent processes for the factors, the explicit expressions for $\alpha_t^i, \beta_t^i, \gamma_t^j, \delta_t^j$ in (20) can easily be obtained according to Section 9.5.2. in McNeil et al. (2005) (see also Section 6.2.2 in Lamberton & Lapeyre (1995)) as functions of the coefficients a^i, b^i in (37) and of the two default intensity parameters λ^A, λ^B in (39). These expressions are rather lengthy, but well-known, and so in what follows we shall for simplicity keep using the shorthand notations $\alpha_t^i, \beta_t^i, \gamma_t^j, \delta_t^j$. Concerning the coefficients $(\bar{\beta}_t^1, \bar{\delta}_t^1, \bar{\delta}_t^2)$ of the additional noise factors, we assume them to be given by the following constant vectors

$$\bar{\beta}_t^1 = [\nu, 0] \quad , \quad \bar{\delta}_t^1 = [0, \rho^A] \quad , \quad \bar{\delta}_t^2 = [0, \rho^B] \quad (40)$$

with (ν, ρ^A, ρ^B) additional parameters to be estimated.

According to (23) the observations in (27) are given by the tuple $(r_t, YL(t, T), CSA(t, T), CSB(t, T))$. Notice however that, even though the spread derives from both, the default-free and the defaultable bonds, for our example and in particular by (39), the observations $CSA(t, T)$ do not depend on Ψ_t^2 and $CSB(t, T)$ do not depend on Ψ_t^1 . According to Proposition 7, with the choice of $\ell = 2$, the dimension of the process X_t becomes $(n + \ell) - (p + q + 1) = 1$ and so we may simply choose as X_t any of the components of $\Psi_t = [\Psi_t^1, \Psi_t^2, \Psi_t^3]'$. Furthermore, due to (26), we can rewrite e^{Ψ_t} in the observation equations (20) in terms of the chosen $X_t = \Psi_t^i$ and of the observations themselves, thus yielding a system which depends on only one single factor.

In addition to the filtering of the unobserved factors, we need also to estimate the eleven parameters, namely (a^i, b^i) for $i = 1, 2, 3$ as well as $(\lambda^A, \lambda^B, \nu, \rho^A, \rho^B)$. The simultaneous estimation of eleven parameters is a rather complex task. On the other hand we have the freedom of choosing as X_t any of the factors Ψ_t^i and the individual observation equations do not depend on all the eleven parameters.

This has led us to consider a set of parallel filter systems, in each of which one chooses as X_t a different factor Ψ_t^i and only nonredundant observations are considered. Furthermore, each of these systems depends on only a subset of all the parameters. In subsection 4.2.2 we shall then present an estimation algorithm that alternates between the individual filter systems.

In what follows we describe now explicitly the various systems that we are going to consider. Firstly there are the following three that result from choosing as X_t the various Ψ_t^i ($i = 1, 2, 3$).

$$\left\{ \begin{array}{l} d\Psi_t^1 = \left[a^1 + e^{-\Psi_t^1} (b^1 - \frac{1}{2}) \right] dt + e^{-\frac{1}{2}\Psi_t^1} dw_t^1 \\ YL(t, T) = (\alpha_t^1 + \beta_t^2 r_t) + (\beta_t^1 - \beta_t^2) e^{\Psi_t^1} + \nu \bar{w}_t^1 \\ CSA(t, T) = \left[\gamma_t^1 + \frac{\delta_t^{1,3}}{\delta_t^{2,3}} (CSB(t, T) - \gamma_t^2 - \delta_t^{2,2} r_t) \right] \\ \quad + \left[\delta_t^{1,1} + \frac{\delta_t^{1,3}}{\delta_t^{2,3}} \delta_t^{2,2} \right] e^{\Psi_t^1} + \left(\rho^A - \frac{\delta_t^{1,3}}{\delta_t^{2,3}} \rho^B \right) \bar{w}_t^2 \end{array} \right. \quad (41)$$

$$\left\{ \begin{array}{l} d\Psi_t^2 = \left[a^2 + e^{-\Psi_t^2} (b^2 - \frac{1}{2}) \right] dt + e^{-\frac{1}{2}\Psi_t^2} dw_t^2 \\ YL(t, T) = (\alpha_t^1 + \beta_t^1 r_t) + (\beta_t^2 - \beta_t^1) e^{\Psi_t^2} + \nu \bar{w}_t^1 \\ CSB(t, T) = \left[\gamma_t^2 + \frac{\delta_t^{2,3}}{\delta_t^{1,3}} (CSA(t, T) - \gamma_t^1 - \delta_t^{1,1} r_t) \right] \\ \quad + \left[\delta_t^{2,2} + \frac{\delta_t^{2,3}}{\delta_t^{1,3}} \delta_t^{1,1} \right] e^{\Psi_t^2} + \left(\rho^B - \frac{\delta_t^{2,3}}{\delta_t^{1,3}} \rho^A \right) \bar{w}_t^2 \end{array} \right. \quad (42)$$

$$\left\{ \begin{array}{l} d\Psi_t^3 = \left[a^3 + e^{-\Psi_t^3} \left(b^3 - \frac{1}{2} \right) \right] dt + e^{-\frac{1}{2}\Psi_t^3} dw_t^3 \\ CS^A(t, T) = \gamma_t^1 + \frac{\delta_t^{1,1}}{\beta_t^1 - \beta_t^2} (YL(t, T) - \alpha_t^1 - \beta_t^2 r_t) + \delta_t^{1,3} e^{\Psi_t^3} \\ \quad + \rho^A \bar{w}_t^2 - \frac{\delta_t^{1,1}}{\beta_t^1 - \beta_t^2} \nu \bar{w}_t^1 \\ CS^B(t, T) = \gamma_t^2 + \frac{\delta_t^{2,2}}{\beta_t^2 - \beta_t^1} (YL(t, T) - \alpha_t^1 - \beta_t^1 r_t) + \delta_t^{2,3} e^{\Psi_t^3} \\ \quad + \rho^B \bar{w}_t^2 - \frac{\delta_t^{2,2}}{\beta_t^2 - \beta_t^1} \nu \bar{w}_t^1 \end{array} \right. \quad (43)$$

In addition to these three systems we consider also the following three where the purpose is to focus specifically on the parameters other than (a^i, b^i) ($i = 1, 2, 3$) namely on $(\lambda^A, \lambda^B, \nu, \rho^A, \rho^B)$. They are

$$\left\{ \begin{array}{l} d\Psi_t^1 = \left[a^1 + e^{-\Psi_t^1} \left(b^1 - \frac{1}{2} \right) \right] dt + e^{-\frac{1}{2}\Psi_t^1} dw_t^1 \\ YL(t, T) = (\alpha_t^1 + \beta_t^2 r_t) + (\beta_t^1 - \beta_t^2) e^{\Psi_t^1} + \nu \bar{w}_t^1 \end{array} \right. \quad (44)$$

$$\left\{ \begin{array}{l} d\Psi_t^1 = \left[a^1 + e^{-\Psi_t^1} \left(b^1 - \frac{1}{2} \right) \right] dt + e^{-\frac{1}{2}\Psi_t^1} dw_t^1 \\ d\Psi_t^3 = \left[a^3 + e^{-\Psi_t^3} \left(b^3 - \frac{1}{2} \right) \right] dt + e^{-\frac{1}{2}\Psi_t^3} dw_t^3 \\ CS^A(t, T) = \gamma_t^1 + \delta_t^{1,1} e^{\Psi_t^1} + \delta_t^{1,3} e^{\Psi_t^3} + \rho^A \bar{w}_t^2 \end{array} \right. \quad (45)$$

$$\left\{ \begin{array}{l} d\Psi_t^2 = \left[a^2 + e^{-\Psi_t^2} \left(b^2 - \frac{1}{2} \right) \right] dt + e^{-\frac{1}{2}\Psi_t^2} dw_t^2 \\ d\Psi_t^3 = \left[a^3 + e^{-\Psi_t^3} \left(b^3 - \frac{1}{2} \right) \right] dt + e^{-\frac{1}{2}\Psi_t^3} dw_t^3 \\ CS^B(t, T) = \gamma_t^2 + \delta_t^{2,2} e^{\Psi_t^2} + \delta_t^{2,3} e^{\Psi_t^3} + \rho^B \bar{w}_t^2 \end{array} \right. \quad (46)$$

Notice that (44) is a subsystem of (41). It is introduced here for the only purpose of estimating the parameter ν .

We may now solve a filtering problem for each of the above six systems and estimate the parameters on which they depend. They are nonlinear systems with Gaussian disturbances and so, in line with Section 3.2, we shall apply the EKF for their solution. For this purpose we have to discretize them in time and then to linearize the nonlinear coefficients around the most recent available estimate of the factors. For the time discretization we apply the Euler-Maruyama scheme with constant step Δ and for what concerns the linearization we use the Taylor approximation

$$f(x) \approx f(\hat{x}) + f'(\hat{x})(x - \hat{x}) \quad (47)$$

for the observation equations and for the drift coefficient in the state equation and

$$f(x) \approx f(\hat{x}) \quad (48)$$

for the diffusion coefficient in the state equation.

Since the six systems are analogous to one another, we describe in detail the procedure only for the first one (41) that we rewrite in more compact form as

$$\begin{cases} d\Psi_t^1 &= \left[a^1 + e^{-\Psi_t^1} \left(b^1 - \frac{1}{2} \right) \right] dt + e^{-\frac{1}{2}\Psi_t^1} dw_t^1 \\ YL(t, T) &= (\alpha_t + \beta_t r_t) + \gamma_t e^{\Psi_t^1} + \nu \bar{w}_t^1 \\ CSA(t, T) &= (f_t + g_t r_t + h_t CS^B(t, T)) + k_t e^{\Psi_t^1} + \rho_t \bar{w}_t^2 \end{cases} \quad (49)$$

Following Section 8.2 in Anderson & Moore (1979), for the linearization of the exponentials in the equation for Ψ_t^1 we take as the most recent estimate of Ψ_t^1 its filtered value $\hat{\Psi}_{t|t}^1$ (the mean value of the filter distribution at time t) so that (47) and (48) become

$$e^{\Psi_t^1} \approx e^{\hat{\Psi}_{t|t}^1} + e^{\hat{\Psi}_{t|t}^1} \left(\Psi_t^1 - \hat{\Psi}_{t|t}^1 \right) \quad \text{and} \quad e^{\Psi_t^1} \approx e^{\hat{\Psi}_{t|t}^1} \quad (50)$$

In the observation equations the most recent estimate of Ψ_t^1 is its predicted value $\hat{\Psi}_{t|t-\Delta}^1$ (the mean value of the predictive distribution at time t). This leads to the following discrete time linear filter system

$$\begin{cases} \Psi_{t+\Delta}^1 &= \left[a^1 + e^{-\hat{\Psi}_{t|t}^1} \left(1 + \hat{\Psi}_{t|t}^1 \right) \left(b^1 - \frac{1}{2} \right) \right] \Delta \\ &+ \left[1 - e^{-\hat{\Psi}_{t|t}^1} \left(b^1 - \frac{1}{2} \right) \Delta \right] \Psi_t^1 + e^{-\frac{1}{2}\hat{\Psi}_{t|t}^1} \sqrt{\Delta} Z_{t+\Delta}^1 \\ YL(t, T) &= \left[\alpha_t + \beta_t r_t + \gamma_t e^{\hat{\Psi}_{t|t-\Delta}^1} \left(1 - \hat{\Psi}_{t|t-\Delta}^1 \right) \right] \\ &+ \gamma_t e^{\hat{\Psi}_{t|t-\Delta}^1} \Psi_t^1 + \nu \bar{w}_t^1 \\ CSA(t, T) &= \left[f_t + g_t r_t + h_t CS^B(t, T) + k_t e^{\hat{\Psi}_{t|t-\Delta}^1} \left(1 - \hat{\Psi}_{t|t-\Delta}^1 \right) \right] \\ &+ k_t e^{\hat{\Psi}_{t|t-\Delta}^1} \Psi_t^1 + \rho_t \bar{w}_t^2 \end{cases} \quad (51)$$

where Z_t^1 is a sequence of i.i.d. standard Gaussian random variables and equality is to be understood modulo the Taylor approximation.

For simplicity, in what follows we shall denote by (y_t^1, cs_t^A, cs_t^B) the noise perturbed yield and spread observations corresponding to the linearized model for a fixed maturity T . With some abuse of notation we then rewrite system (51) in a further compactified notation

$$\begin{cases} \Psi_{t+\Delta}^1 &= \bar{a}_t + \bar{b}_t \Psi_t^1 + \bar{c}_t Z_{t+\Delta}^1 \\ y_t^1 &= (\bar{\alpha}_t + \beta_t r_t) + \bar{\gamma}_t \Psi_t^1 + \nu \bar{w}_t^1 \\ cs_t^A &= (\bar{f}_t + g_t r_t + h_t cs_t^B) + \bar{k}_t \Psi_t^1 + \rho_t \bar{w}_t^2 \end{cases} \quad (52)$$

and notice that the coefficients depend not only on the parameters and the observations themselves but, due to the linearization procedure, also on the filtered and predicted estimates $\hat{\Psi}_{t|t}^1$ and $\hat{\Psi}_{t|t-\Delta}^1$.

4.2.1 The EM algorithm itself

Each of the individual parallel filter systems (41)-(46) can, via time discretization and linearization, be reduced to a system of the form (52) that is of the discrete time linear conditionally Gaussian type

so that the Kalman filter can be applied to determine the filter distribution that is Gaussian and thus characterized by its (conditional) mean and variance. We are now going to describe schematically the procedure followed in Fontana (2007) to use the results of Elliott & Krishnamurthy (1999) in order to obtain the EM parameter estimates as functions just of the filter solution.

System (52), and analogously the other parallel filter systems, may be conveniently rewritten in the following synthetic form where we now use $k \in \mathbb{N}$ to index the discrete time points.

$$\begin{cases} X_{k+1} &= V_k(\theta) + F_k(\theta)X_k + D_{k+1}(\theta) Z_{k+1} \\ Y_k &= R_k(\theta) + C_k(\theta)X_k + E_k(\theta) \bar{w}_k \end{cases} \quad (53)$$

In (53) the variables X_k denote the unobservable component of the state-observation system (corresponds to one or more components of Ψ_k), Y_k is a subvector of (yl, cs^A, cs^B) and Z_k and \bar{w}_k denote independent sequences of independent Gaussian variables/vectors. In the coefficients we have made explicit their dependence on the parameter vector denoted as θ and notice that these coefficients may depend also on the observation history $Y^k := (Y_0, \dots, Y_k)$ up to time k .

The fundamental quantity underlying the EM algorithm is $Q(\theta, \theta')$ in (35), where we may write

$$\frac{dP^\theta}{dP^{\theta'}} = \frac{dP^\theta}{d\Lambda} \cdot \frac{d\Lambda}{dP^{\theta'}} \quad (54)$$

with Λ denoting Lebesgue measure. Suppose then that we are standing at the generic time step k so that $\frac{dP^\theta}{d\Lambda}$ can be seen as the joint density of $(X_l, Y_l)_{l=0, \dots, k}$ for given values of the parameters in θ (analogously for $\frac{dP^{\theta'}}{d\Lambda}$ when the parameters are θ'). It can then be easily seen that

$$\begin{aligned} Q(\theta, \theta') &= -\sum_{l=1}^k \log |D_l(\theta)| - \sum_{l=0}^k \log |E_l(\theta)| \\ &- \frac{1}{2} E_{\theta'} \left[\sum_{l=0}^{k-1} (X_{l+1} - F_l(\theta) X_l - V_l(\theta))' (D_{l+1}(\theta) D_{l+1}'(\theta))^{-1} (X_{l+1} - F_l(\theta) X_l - V_l(\theta)) \mid \mathcal{F}_k^Y \right] \\ &- \frac{1}{2} E_{\theta'} \left[\sum_{l=0}^k (Y_l - C_l(\theta) X_l - R_l(\theta))' (E_l(\theta) E_l'(\theta))^{-1} (Y_l - C_l(\theta) X_l - R_l(\theta)) \mid \mathcal{F}_k^Y \right] \\ &+ E_{\theta'} [J_k(\theta') \mid \mathcal{F}_k^Y] \end{aligned} \quad (55)$$

where $J_k(\theta')$ includes all the terms that do not depend on θ and thus plays no role in the maximization of $Q(\theta, \theta')$ with respect to θ .

The maximization of $Q(\theta, \theta')$ with respect to θ can be accomplished by differentiating $Q(\theta, \theta')$ with respect to θ and solving the system that results from putting $\frac{\partial Q(\theta, \theta')}{\partial \theta} = 0$. From (55) one can now see that the coefficients in these equations can be determined by computing (this corresponds to step 1 of the EM algorithm), for a given value of θ' and a generic value of θ , expressions of the form

$$\begin{aligned} E_{\theta'} \left[\sum_{l=0}^k X_l' f_l(\theta) \mid \mathcal{F}_k^Y \right] \quad , \quad E_{\theta'} \left[\sum_{l=0}^k X_l' \bar{f}_l(\theta) X_l \mid \mathcal{F}_k^Y \right] \\ E_{\theta'} \left[\sum_{l=0}^{k-1} X_{l+1}' g_l(\theta) X_{l+1} \mid \mathcal{F}_k^Y \right] \quad , \quad E_{\theta'} \left[\sum_{l=0}^{k-1} X_{l+1}' \bar{g}_l(\theta) X_l \mid \mathcal{F}_k^Y \right] \end{aligned} \quad (56)$$

Strictly speaking, since the previous expressions require the conditional expectations of, say, X_l given \mathcal{F}_k^Y where $l < k$, a smoothing procedure would be needed. At this point one can however make use of the results in Elliott & Krishnamurthy (1999) that allow these quantities to be computed as functions of the (conditional) mean and variance under $P^{\theta'}$ of the filter distribution of X_k given \mathcal{F}_k^Y . This mean, denoted at the generic time step by $\hat{X}_{k|k}$, and variance, denoted by $P_{k|k}$, can be computed using the

Kalman filter applied to the generic system (53) by setting $\theta = \theta'$. We show here the expressions for the first two quantities in (56), for the remaining ones the expressions are analogous and can be found in Elliott & Hyndman (2006) and Fontana (2007).

For the first quantity we have

$$E_{\theta'} \left[\sum_{l=0}^k X_l' f_l(\theta) \middle| \mathcal{F}_k^Y \right] = \phi_k + \psi_k' \hat{X}_{k|k}$$

with ϕ_k and ψ_k satisfying the recursions

$$\begin{cases} \phi_{k+1} = \phi_k + \psi_k' S_k & ; \quad \phi_0 = 0 \\ \psi_{k+1} = \Sigma_{k+1} \psi_k + f_{k+1}(\theta) & ; \quad \psi_0 = f_0(\theta) \end{cases}$$

where, putting

$$\sigma_k := F_k(\theta) [D_k(\theta) D_k'(\theta)]^{-1} F_k'(\theta) + (P_{k-1|k-1})^{-1}$$

one has

$$\begin{cases} \Sigma_k := [D_k(\theta) D_k'(\theta)]^{-1} F_k'(\theta) \sigma_k^{-1} \\ S_k := \sigma_{k+1}^{-1} \left((P_{k|k})^{-1} \hat{X}_{k|k} - F_k(\theta) [D_k(\theta) D_k'(\theta)]^{-1} V_k(\theta) \right) \end{cases}$$

For the second quantity one has instead

$$E_{\theta'} \left[\sum_{l=0}^k X_l' \bar{f}_l(\theta) X_l \middle| \mathcal{F}_k^Y \right] = \phi_k + \psi_k' \hat{X}_{k|k} + \text{tr}(\Lambda_k P_{k|k}) + \hat{X}_{k|k}' \Lambda_k \hat{X}_{k|k}$$

where, with σ_k, Σ_k, S_k as before, one has the recursions

$$\begin{cases} \phi_{k+1} = \phi_k + \psi_k' S_k + \text{tr}(\Lambda_k \sigma_{k+1}^{-1}) + S_k' \Lambda_k S_k & ; \quad \phi_0 = 0 \\ \psi_{k+1} = \Sigma_{k+1} (\psi_k + 2\Lambda_k S_k) & ; \quad \psi_0 = 0 \\ \Lambda_{k+1} = \Sigma_{k+1}' \Lambda_k \Sigma_{k+1} + \bar{f}_l(\theta) & ; \quad \Lambda_0 = \bar{f}_0(\theta) \end{cases}$$

4.2.2 The alternating algorithm

While step i) of the EM algorithm, namely computing the conditional expectations, has been discussed at the end of the previous subsection, in this subsection we deal with step ii), namely the maximization of $Q(\theta, \theta')$ with respect to θ which involves the solution of the system $\frac{\partial Q(\theta, \theta')}{\partial \theta} = 0$. For the model in the present Example, which leads to the parallel state-observation systems (41)-(46) (in the form (52) after discretizing in time and linearizing the dynamics), the parameters included in θ are (a^i, b^i) , $i = 1, 2, 3$, and $(\lambda^A, \lambda^B, \nu, \rho^A, \rho^B)$. Notice that the latter enter only the coefficients in the observation equations. For what concerns the maximization with respect to (a^i, b^i) , an advantage of having the three parallel systems (41)-(43) is that in each of these systems the state dynamics depend on (a^i, b^i) for just one of the indices $i = 1, 2, 3$ and this already reduces the number of parameters to be estimated in each individual step of the EM iterations.

A direct application of the EM algorithm would now consist in iterating between the two EM steps, namely the expectation and the maximization steps, globally for all eleven parameters in the six systems simultaneously until a stopping criterion is met. One can however alternate the iterations between the individual systems (41)-(46) or subgroups of these systems. To this effect let, with some abuse of notation, θ be given by the subset of the parameters with respect to which one performs the maximization corresponding to the generic iteration step and θ' be the subset of the parameters that for this maximization step are kept fixed at their previously estimated values. The situation corresponding to the various systems (41)-(46) is then the following.

- For system (41): $\theta = (a^1, b^1)$; $\theta' = (a^i, b^i, \lambda^j, \nu, \rho^j)_{i=2,3; j=A,B}$;
- for system (42): $\theta = (a^2, b^2)$; $\theta' = (a^i, b^i, \lambda^j, \nu, \rho^j)_{i=1,3; j=A,B}$;
- for system (43): $\theta = (a^3, b^3)$; $\theta' = (a^i, b^i, \lambda^j, \nu, \rho^j)_{i=1,2; j=A,B}$;
- for system (44): $\theta = \nu$; $\theta' = (a^i, b^i, \lambda^j, \rho^j)_{i=1,2,3; j=A,B}$;
- for system (45): $\theta = (\lambda^A, \rho^A)$; $\theta' = (a^i, b^i, \lambda^B, \nu, \rho^B)_{i=1,2,3}$;
- for system (46): $\theta = (\lambda^B, \rho^B)$; $\theta' = (a^i, b^i, \lambda^A, \nu, \rho^A)_{i=1,2,3}$.

From our numerical simulations it turned out that the most efficient alternating iterative procedure consists in alternating the EM iterations between the group formed by (41),(46), that formed by (42),(45) and that formed by (43),(44). More precisely, considering that the global vector of parameters is given by $\theta = (a^1, b^1, \dots, a^3, b^3, \lambda^A, \lambda^B, \nu, \rho^A, \rho^B)$, we have the following iterative algorithm, where by θ_j we denote the vector of estimates of θ at the j -th step of the iterations.

Algorithm 10.

0. Initialize the algorithm with a guess $\hat{\theta}$ for the entire vector θ and, setting $j = 0$, put $\theta_j = \hat{\theta}$;
1. Apply in parallel on each of the systems (41),(46) the EM algorithm to estimate (a^1, b^1) and (λ^B, ρ^B) while keeping the other parameters fixed at their previously estimated values $(a_j^2, b_j^2, a_j^3, b_j^3, \lambda_j^A, \nu_j, \rho_j^A)$. The algorithm iterates through the two EM steps (expectation and maximization) until a stopping criterion is met, thereby producing estimates $(a_{j+1}^1, b_{j+1}^1, \lambda_{j+1}^B, \rho_{j+1}^B)$;
2. Apply in parallel on each of the systems (42),(45) the EM algorithm to estimate (a^2, b^2) and (λ^A, ρ^A) while keeping the other parameters fixed at their previously estimated values $(a_{j+1}^1, b_{j+1}^1, a_j^3, b_j^3, \lambda_{j+1}^B, \nu_j, \rho_{j+1}^B)$. The algorithm iterates through the two EM steps until a stopping criterion is met, thereby producing estimates $(a_{j+1}^2, b_{j+1}^2, \lambda_{j+1}^A, \rho_{j+1}^A)$;
3. Apply in parallel on each of the systems (43),(44) the EM algorithm to estimate (a^3, b^3, ν) keeping all others parameters fixed at their previously estimated values. The algorithm iterates through the two EM steps until a stopping criterion is met, thereby producing estimates $(a_{j+1}^3, b_{j+1}^3, \nu_{j+1})$;
4. Put $\theta_{j+1} = (a_{j+1}^1, b_{j+1}^1, \dots, a_{j+1}^3, b_{j+1}^3, \lambda_{j+1}^A, \lambda_{j+1}^B, \nu_{j+1}, \rho_{j+1}^A, \rho_{j+1}^B)$ and, setting $j = j + 1$, return to step 1. Terminate the entire algorithm as soon as a global stopping criterion is met.

For the initial guess $\hat{\theta}$ of the parameter vector θ one may follow some guidelines as for instance: for the diffusion coefficients in the observations one may use the bid-ask spread observed on the market; for the parameters λ^j ($j = A, B$) one may consider the approximation for the intensity, frequently used in practice, whereby one puts $\lambda_t^j \approx \frac{S_t^j}{LGD^j}$ with S_t^j a CDS spread corresponding to firm j and LGD^j the loss given default for that same firm.

Besides the already mentioned facts, one additional advantage of the proposed procedure is that the EM algorithm is applied in a simple context, namely when the systems are of the discrete time linear conditionally Gaussian type with uncorrelated noises in the state and the observations. Thus, no smoothing procedure is needed.

5 Simulation results

This section is meant to illustrate with some numerical simulations the main features of the model and the performance of the just described estimation algorithm. The model considered for the numerical simulations below is formulated according to the Example of Section 4.2, with the parameters in (37), (39) and (40) chosen as shown in Table 1.

Parameter	Value	Parameter	Value
a^1	-0.15	b^1	0.60
a^2	-0.20	b^2	0.70
a^3	-0.25	b^3	0.80
λ^A	0.10	λ^B	0.30
ρ^A	0.01	ρ^B	0.02
ν	0.005		

Table 1: Parameter values for simulation study

For these parameter values, we first generate a sequence of tuples $(\Psi_k^1, \Psi_k^2, \Psi_k^3)$ on the basis of the dynamics (37). Then, given the realizations of the unobserved factors and the values of the intensity parameters λ^A and λ^B , we randomly draw the two default times τ^A and τ^B , following the procedure described in McNeil et al. (2005), Algorithm 9.34. Finally, a sequence of corresponding observations (given, as in Section 4.2, by the interest rate, one yield and two credit spreads) is generated according to (20), with the credit spreads of firm j being observed only up to τ^j (for $j = A, B$). Furthermore, we assume that yield and credit spreads are relative to the same maturity T , chosen as $T = 10$, and we adopt a constant step size in the Euler-Maruyama time discretization equal to $\Delta = 0.02$, that is, we assume to have access to (approximately) weekly observations of $(r_k, yl_k, cs_k^A, cs_k^B)$, starting from $k = 0$ and up to $k = T/\Delta$ (up to $k = \tau^j/\Delta$ if $\tau^j < T$ for cs_k^j , $j = A, B$). This results in a total number of 500 observations for the interest rate (r_k) and the yield (yl_k) and in a maximum of 500 observations for the two credit spreads (cs_k^A and cs_k^B).

As mentioned in Section 3.2, the only approximation introduced in our study consists in the solution of the filtering problem via the Extended Kalman Filter. For this reason, we have performed a Monte Carlo analysis in order to assess the magnitude of the approximation error. We simulated the whole series of observations $(r_k, yl_k, cs_k^A, cs_k^B)_{k=0, \dots, T/\Delta}$ and then, using systems (41)-(43) in the form (52) we computed the filtered estimate of the unobserved state process by applying to each system the EKF. The filtering recursions (see Anderson & Moore (1979), Section 8.2) were initialized at the (unconditional) mean and variance of the stationary distribution of Ψ_t^i , which can be explicitly computed since it is known that, due to the dynamics (37), $e^{\Psi_t^i} = \Phi_t^i$ follows a (non-central) chi-square distribution. We repeated this procedure for 1000 independent samples and in Table 2 we summarize the results by showing the mean and the root mean square error (RMSE) of $\Psi_t^i - \hat{\Psi}_{t|t}^i$, $i = 1, 2, 3$.

The same analysis was performed with systems (44)-(46) and the results are comparable to those reported in Table 2, thus showing that the EKF does not introduce a severe bias in the filtered estimates of the unobserved factors. This is consistent with most of the literature dealing with state-space estimation of partially observed term structures models (see e.g. Chen & Scott (2003)).

Having shown that the EKF leads to satisfactory results, we now test the performance of the estimation algorithm described in Section 4.2.2 in order to show the convergence of the estimated parameter values. Since for our simulation purposes there is no reason to prefer certain parameter values to others, the starting vector $\hat{\theta}$ for the parameter values was randomly generated within a suitable interval around the “true” values given in Table 1. However, notice that in real applications one can make a reasonable guess for $\hat{\theta}$ on the basis of the available financial/historical information, as mentioned at the end of the

System (41)	
Mean of $\Psi_t^1 - \hat{\Psi}_{t t}^1$	RMSE of $\Psi_t^1 - \hat{\Psi}_{t t}^1$
-8.7891×10^{-5}	0.00215
System (42)	
Mean of $\Psi_t^2 - \hat{\Psi}_{t t}^2$	RMSE of $\Psi_t^2 - \hat{\Psi}_{t t}^2$
-8.9599×10^{-5}	0.00214
System (43)	
Mean of $\Psi_t^3 - \hat{\Psi}_{t t}^3$	RMSE of $\Psi_t^3 - \hat{\Psi}_{t t}^3$
-0.00039	0.01776

Table 2: Results of the EKF algorithm over 1000 independent samples

last subsection. The alternating algorithm was applied stopping the individual EM iterations as soon as two successive values of all the parameters differed by less than 10^{-5} up to a maximum of 500 iterations. While the convergence can sometimes be slow in the individual EM iterations (those within steps 1., 2. and 3. of Algorithm 10), from our numerical tests we found that the number of “alternations” (i.e. iterations between the steps 1., 2. and 3.) does not need to be large (in most cases up to 10 “alternations” suffice to reach convergence).

In Table 3 we report, along with the “true” values of the parameters, the means of the estimated values of the individual parameters and their associated standard deviations, computed from 50 independent runs of the entire estimation procedure, where each run comprises at most 10 “alternations”. From these simulation results one can see that most of the parameters, including the crucial intensity parameters, namely λ^A and λ^B , are estimated with satisfactory accuracy.

Parameter	True value	Estimate	Std. dev.
a^1	-0.15	-0.15092	0.02980
b^1	0.60	0.60141	0.05163
a^2	-0.20	-0.19960	0.02529
b^2	0.70	0.69943	0.04205
a^3	-0.25	-0.24826	0.06377
b^3	0.80	0.80812	0.09398
λ^A	0.10	0.09981	0.02204
λ^B	0.30	0.30996	0.02533
ν	0.005	0.00509	0.00053
ρ^A	0.01	0.01001	0.00055
ρ^B	0.02	0.01978	0.00106

Table 3: Mean and standard deviation of the estimated parameter values (from a sample of 50 independent runs of the alternating algorithm).

A recurrent observation arising from our simulations is that the estimates of the noise coefficients ν, ρ^A, ρ^B converge to their true values very quickly, in most cases after the first few iterations, while the parameters a^i, b^i ($i = 1, 2, 3$) which determine the mean of the factors Ψ^i need more iterations to converge steadily to their true values. This corresponds to the well-known fact that it is generally more difficult to obtain reliable estimates of parameters determining the mean in comparison to parameters for the volatility or the noise (see Duffee & Stanton (2004)).

Having shown the performance of both the EKF and our alternating algorithm, we now show how

even with the simple formulation of the model considered in Section 4.2 we can capture the phenomenon of the *information-driven default contagion*. More precisely, we simulate the trajectories of the unobserved factors Ψ_t^i ($i = 1, 2, 3$) and the default times τ^A, τ^B and then we run the EKF up to the first default time $\tau^* := \tau^A \wedge \tau^B$. At the first default time we update, according to Proposition 8, the incoming filter distribution for the factors provided by the EKF and then we compute the conditional distribution (given the observations up to the first default time) of the default intensity of the surviving firm. We can then compare the latter distribution with the distribution of the intensity prior to the default time, as shown in Figure 1.

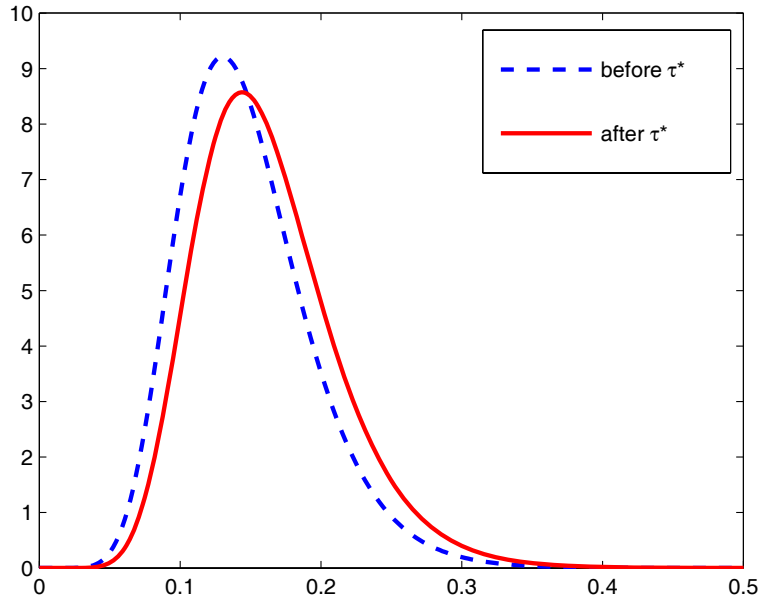


Figure 1: Effect of the update of the filter distribution of the default intensity of the surviving firm at the first default time

From Figure 1 we can see clearly that, as pointed out in Remark 9.ii), at a default time the mean of the filter distribution of the intensity of the surviving firm (i.e. the “filtered default intensity”) exhibits a jump, thus providing evidence of the phenomenon of *information induced default contagion*.

6 Conclusions and further developments

We have proposed a general affine intensity-based credit risk model in which the interest rate and the default intensities depend on several (partially) unobservable stochastic factors, which represent both common and idiosyncratic risk factors. In this way we can capture, besides default clustering, also the significant phenomenon of *information driven default contagion*. Within this framework we have developed a *filter-based pricing model*. More precisely, we have shown how via a filtering approach we can solve both the problem of pricing *OTC* products and the problem of estimating the parameters characterizing the model. Moreover, we have shown how within our framework these problems are intimately linked to one another, since the dynamic updating of the filter solution allows both prices and parameter estimates to be tuned to the actual market situation.

Possible extensions that we have not been dealing with in the present paper consist in considering the model under the physical/historical probability measure, specifying as further unobserved stochastic

processes the *risk premia*, which characterize the change of measure (in this context see e.g. Runggaldier (2004)). One can then consider also information coming from sources outside the financial market, such as the *rating score*. From the risk management perspective, an interesting application in this latter context would then be the estimation of default probabilities, on the basis of the information deriving from both the financial market and the *rating score*.

7 Appendix

Proof of Lemma 4

Proof. We start by introducing the T -forward survival measure \tilde{P}^T (also named *restricted survival measure* in Bielecki & Rutkowski (2002)), defined by its density process $(\tilde{Z}_t^T)_{0 \leq t \leq T}$ wrt. P :

$$\tilde{Z}_t^T := \frac{d\tilde{P}^T}{dP} \Big|_{\mathcal{F}_t} = e^{-\int_0^t (r_u + \lambda_u) du} \frac{E^P \left[e^{-\int_t^T (r_u + \lambda_u) du} \mid \mathcal{F}_t \right]}{E^P \left[e^{-\int_0^T (r_u + \lambda_u) du} \right]} = e^{-\int_0^t \lambda_u du} \frac{E^{P^T} \left[e^{-\int_t^T \lambda_u du} \mid \mathcal{F}_t \right]}{E^{P^T} \left[e^{-\int_0^T \lambda_u du} \right]} \frac{dP^T}{dP} \Big|_{\mathcal{F}_t} \quad (57)$$

Then, for any integrable and \mathcal{F}_T -measurable random variable X we have:

$$\begin{aligned} E^P \left[e^{-\int_t^T r_u du} X \mathbf{1}_{\{\tau > T\}} \mid \mathcal{G}_t \right] &= \mathbf{1}_{\{\tau > t\}} E^P \left[e^{-\int_t^T (r_u + \lambda_u) du} X \mid \mathcal{F}_t \right] \\ &= \mathbf{1}_{\{\tau > t\}} E^P \left[e^{-\int_t^T (r_u + \lambda_u) du} \mid \mathcal{F}_t \right] E^{\tilde{P}^T} [X \mid \mathcal{F}_t] = \Pi(t, T) E^{\tilde{P}^T} [X \mid \mathcal{F}_t] \end{aligned} \quad (58)$$

where the last equality follows from (8). Moreover, on the set $\{\tau > t\}$, using (57) and Corollary 9.10 of McNeil et al. (2005) we have:

$$\begin{aligned} E^{\tilde{P}^T} [X \mid \mathcal{F}_t] &= \frac{E^P \left[e^{-\int_t^T (r_u + \lambda_u) du} X \mid \mathcal{F}_t \right]}{E^P \left[e^{-\int_t^T (r_u + \lambda_u) du} \mid \mathcal{F}_t \right]} = \frac{E^{P^T} \left[e^{-\int_t^T \lambda_u du} X \mid \mathcal{F}_t \right]}{E^{P^T} \left[e^{-\int_t^T \lambda_u du} \mid \mathcal{F}_t \right]} = \frac{E^{P^T} [X \mathbf{1}_{\{\tau > T\}} \mid \mathcal{F}_t]}{E^{P^T} [\mathbf{1}_{\{\tau > T\}} \mid \mathcal{F}_t]} \\ &= \frac{E^{P^T} [X \mathbf{1}_{\{\tau > T\}} \mid \mathcal{G}_t]}{E^{P^T} [\mathbf{1}_{\{\tau > T\}} \mid \mathcal{G}_t]} \end{aligned} \quad (59)$$

Recall, furthermore, that (see McNeil et al. (2005), Theorem 9.23):

$$E^P \left[\mathbf{1}_{\{\tau > t\}} e^{-\int_t^\tau r_u du} \mathbf{1}_{\{\tau \leq T\}} \mid \mathcal{G}_t \right] = \mathbf{1}_{\{\tau > t\}} \int_t^T E^P \left[\lambda_s e^{-\int_t^s (r_u + \lambda_u) du} \mid \mathcal{F}_t \right] ds \quad (60)$$

The integrand in (60) can now be written as follows. We first use differentiation inside the expectation of the expression for $\Pi(t, T)$ in (8), which can be justified here via Dominated Convergence that is applicable thanks to the fact that r_t and λ_t take positive values. Combining this with (58) then leads to

$$\begin{aligned} \mathbf{1}_{\{\tau > t\}} E^P \left[\lambda_s e^{-\int_t^s (r_u + \lambda_u) du} \mid \mathcal{F}_t \right] &= -\frac{\partial}{\partial s} \Pi(t, s) - \mathbf{1}_{\{\tau > t\}} E^P \left[r_s e^{-\int_t^s (r_u + \lambda_u) du} \mid \mathcal{F}_t \right] \\ &= -\frac{\partial}{\partial s} \Pi(t, s) - \Pi(t, s) E^{\tilde{P}^s} [r_s \mid \mathcal{F}_t] \end{aligned} \quad (61)$$

Straightforward computations, using again differentiation inside the expectation of the expression for $\Pi(t, T)$ as well as the definition of the density process \tilde{Z}_t^T for the measure \tilde{P}^T , allow to show that

$\tilde{f}(t, T) = \mathbb{E}^{\tilde{P}^T} [r_T + \lambda_T | \mathcal{F}_t]$ on the set $\{\tau > t\}$ and for any $T > t$. Therefore, using (59):

$$\begin{aligned} E^{\tilde{P}^T} [r_T | \mathcal{F}_t] &= E^{\tilde{P}^T} [r_T + \lambda_T | \mathcal{F}_t] - E^{\tilde{P}^T} [\lambda_T | \mathcal{F}_t] = \tilde{f}(t, T) - E^{\tilde{P}^T} [\lambda_T | \mathcal{F}_t] \\ &= \tilde{f}(t, T) - \frac{E^{P^T} [\lambda_T \mathbf{1}_{\{\tau > T\}} | \mathcal{F}_t]}{E^{P^T} [\mathbf{1}_{\{\tau > T\}} | \mathcal{F}_t]} = \tilde{f}(t, T) - \lim_{\Delta \searrow 0} \frac{P_t^T (T < \tau \leq T + \Delta | \tau > T)}{\Delta} \\ &=: \tilde{f}(t, T) - \tilde{h}(t, T) \end{aligned} \tag{62}$$

Putting together (60), (61) and (62) we immediately obtain (13). To prove (14) notice that if r_t and λ_t follow independent stochastic processes (conditionally on \mathcal{F}_t), by (59):

$$E^{\tilde{P}^T} [r_T | \mathcal{F}_t] = \frac{E^{P^T} [r_T \mathbf{1}_{\{\tau > T\}} | \mathcal{F}_t]}{E^{P^T} [\mathbf{1}_{\{\tau > T\}} | \mathcal{F}_t]} = E^{P^T} [r_T | \mathcal{F}_t] = f(t, T) \quad \text{for any } T > t \tag{63}$$

which, together with (60) and (61) shows (14). \square

Remark 11. We point out that when r_t and λ_t are not independent, then $\tilde{f}(t, T) - \tilde{h}(t, T) \neq f(t, T)$ in general. However, due to (15) and after some simple (but rather lengthy) computations we can evaluate the error introduced by eventually approximating $\tilde{f}(t, T) - \tilde{h}(t, T)$ with $f(t, T)$ as:

$$|f(t, T) - \tilde{f}(t, T) + \tilde{h}(t, T)| = \frac{1}{P_t^T(\tau > T)} \lim_{\Delta \searrow 0} \frac{|Cov_t^T(\mathbf{1}_{\{\tau > T + \Delta\}}, e^{-\int_T^{T+\Delta} r_u du})|}{\Delta} \tag{64}$$

Acknowledgment: The authors acknowledge useful comments made by Thorsten Schmidt on an earlier version of the paper. Part of the contribution by the second author was obtained while he was visiting professor 2009 for the chair Quantitative Finance and Insurance at the LMU University in Munich funded by LMU Excellent. Hospitality and financial support are gratefully acknowledged.

References

- Anderson, B.D.O. and Moore, J.B., *Optimal Filtering*. Prentice-Hall, Englewoods Cliffs, 1979.
- Azizpour, S. and Giesecke, K., Self-exciting Corporate Defaults: Contagion vs. Frailty. Working Paper, Stanford University, 2008.
- Baadsgaard, M., Nielsen, J.N. and Madsen, H., Estimating Multivariate Exponential-Affine Term Structure Models from Coupon Bond Prices using Nonlinear Filtering. *Econometric Journal*, 2000, 3, 1–20.
- Bielecki, T.R., Jeanblanc, M. and Rutkowski, M., Modeling and Valuation of Credit Risk, in: Frittelli, M. & Runggaldier, W. (editors), *Stochastic Methods in Finance - Lectures given at the CIME-EMS Summer School in Bressanone/Brixen, July 2003*. Springer, Berlin-Heidelberg-New York, 2004.
- Bielecki, T.R. and Rutkowski, M., *Credit Risk: Modeling, Valuation, and Hedging*. Springer, Berlin-Heidelberg-New York, 2002.

- Bhar, R. and Handzic, N., A Multifactor Model of Credit Spreads. Working paper, University of New South Wales, Sydney, 2008.
- Chen, R.R. and Scott, L., Maximum Likelihood Estimation for a Multifactor Equilibrium Model of the Term Structure of Interest Rates. *Journal of Fixed Income*, 1993, 3(3), 14–32.
- Chen, R.R. and Scott, L., Multi-Factor Cox-Ingersoll-Ross Models of the Term Structure: Estimates and Tests from a Kalman Filter Model. *Journal of Real Estate Finance and Economics*, 2003, 27(2), 143–172.
- Cont, R. and Tankov, P., Non-Parametric Calibration of Jump-Diffusion Option Pricing Models. *Journal of Computational Finance*, 2004, 7(3), 1–49.
- Dai, Q. and Singleton, K.J., Specification Analysis of Affine Terms Structure Models. *Journal of Finance*, 2000, LV(5), 1943–1978.
- Das, S.R., Duffie, D., Kapadia, N. and Saita, L., Common Failings: How Corporate Defaults are Correlated. *Journal of Finance*, 2007, LXII(1), 93–117.
- Duan, J.C. and Simonato, J.G., Estimating and Testing Exponential-Affine Term Structure Models by Kalman Filter. *Review of Quantitative Finance and Accounting*, 1999, 13(2), 111–135.
- Duffee, G.R. and Stanton, R.H., Estimation of Dynamic Term Structure Models. Working paper, Haas School of Business, Berkeley, 2004.
- Duffie, D., Eckner, A., Horel, G. and Saita, L., Frailty Correlated Default. *Journal of Finance*, 2009, LXIV(5), 2089–2123.
- Duffie, D. and Kan, R., A Yield-factor Model of Interest Rates. *Mathematical Finance*, 1996, 6(4), 379–406.
- Duffie, D. and Singleton, K.J., An Econometric Model of the Term Structures of Interest-Rate Swap Yields. *Journal of Finance*, 1997, LII(4), 1287–1321.
- Duffie, D. and Singleton, K.J., Modeling Term Structures of Defaultable Bonds. *Review of Financial Studies*, 1999, 12(4), 687–720.
- Elliott, R.J. and Hyndman, C.B., Parameter Estimation in Commodity Markets: A Filtering Approach. *Journal of Economic Dynamics and Control*, 2006, 31(7), 2350–2373.
- Elliott, R.J. and Krishnamurthy, V., Exact Finite-dimensional Filters for Maximum Likelihood Parameter Estimation of Continuous-time Linear Gaussian Systems. *SIAM Journal of Control and Optimization*, 1997, 35(6), 1908–1923.
- Elliott, R.J. and Krishnamurthy, V., New Finite Dimensional Filters for Parameter Estimation of Discrete-Time Linear Gaussian Models. *IEEE Transactions on Automatic Control*, 1999, 44(5), 938–951.
- Feldhütter, P. and Lando, D., Decomposing Swap Spreads. *Journal of Financial Economics*, 2008, 88, 375–405.
- Fontana, C., Affine Multi-factor Credit Risk Models under Incomplete Information: Filtering and Parameter Estimation. Thesis, University of Padova, 2007.

- Frey, R., Prosdocimi, C. and Runggaldier, W.J., Affine Credit Risk Models under Incomplete Information, in: Akahori, J., Ogawa, S. & Watanabe, S. (editors), *Stochastic Processes and Applications to Mathematical Finance - Proceedings of the 6th Ritsumeikan International Symposium*. World Scientific, Singapore, 2006.
- Frey, R. and Runggaldier, W.J., Nonlinear Filtering in Models for Interest-Rate and Credit Risk. Submitted to: Crisan, D. & Rozovski, B. (editors), *Handbook on Nonlinear Filtering*, Oxford University Press, 2009.
- Geyer, A.L.J. and Pichler, S., A State-space Approach to Estimate and Test Multifactor Cox-Ingersoll-Ross Models of the Term Structure. *Journal of Financial Research*, 1999, 22(1), 107–130.
- Gombani, A., Jaschke, S.R. and Runggaldier, W.J., A Filtered No Arbitrage Model for Term Structures from Noisy Data. *Stochastic Processes and their Applications*, 2005, 115, 381–400.
- Jacobs, K. and Li, X., Modeling the Dynamics of Credit Spreads with Stochastic Volatility. *Management Science*, 2008, 54(6), 1176–1188.
- Karatzas, I. and Shreve, S.E., *Brownian Motion and Stochastic Calculus*. Springer, Berlin-Heidelberg-New York, 1991.
- Lamberton, D. and Lapeyre, B., *An Introduction to Stochastic Calculus Applied to Finance*. Chapman and Hall, London, 1995.
- Lando, D., On Cox Processes and Credit Risky Securities. *Review of Derivatives Research*, 1998, 2, 99–120.
- Lund, J., Econometric Analysis of Continuous-Time Arbitrage-Free Models of the Term Structure of Interest Rates. Working paper, Department of Finance, Aarhus School of Business, 1997.
- McLachlan, G.J. and Krishnan, T., *The EM Algorithm and Extensions*. John Wiley, New York, 1997.
- McNeil, A.J., Frey, R. and Embrechts, P., *Quantitative Risk Management*. Princeton University Press, Princeton, 2005.
- Pearson, N.D. and Sun, T.S., Exploiting the Conditional Density in Estimating the Term Structure: An Application to the Cox, Ingersoll, Ross Model. *Journal of Finance*, 1994, XLIX(4), 1279–1304.
- Runggaldier, W.J., Estimation via Stochastic Filtering in Financial Market Models. *Contemporary Mathematics*, 2004, 351, 309–318, American Mathematical Society.
- Schönbucher, P.J., *Credit Derivatives Pricing Models: Models, Pricing and Implementation*. Wiley, Chichester, 2003A.
- Schönbucher, P.J., Information-Driven Default Contagion. Working paper, Department of Mathematics, ETH Zürich, 2003B.
- Wu, C.F.J., On the Convergence Properties of the EM Algorithm. *The Annals of Statistics*, 1983, 11(1), 95–103.
- Yu, F., Default Correlation in Reduced-Form Models. *Journal of Investment Management*, 2005, 3(4), 33–42.