

Credit risk and incomplete information: linear filtering and EM parameter estimation

Claudio Fontana and Wolfgang J. Runggaldier

Dipartimento di Matematica Pura ed Applicata, Università di Padova,
Via Trieste, 63, I-35121 Padova, Italy
e-mails : fontana@math.unipd.it, runggal@math.unipd.it

June 18, 2008

Abstract

We consider a reduced-form credit risk model where default intensity and interest rate are linear functions of a not fully observable Markovian factor process. 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 linear 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 intensity that affects the default probability 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, which is observed in real markets.

To this effect, we propose a model in an incomplete information framework, that is, we assume that both the default intensity and the interest rate are given as linear functions of a multivariate stochastic factor process and that at least some of the components of the latter are unobservable. Typically, the components of the factor process represent the evolution of the macroeconomic environment but some of them may be without specific economic interpretation and are introduced only to enhance the flexibility of the model. This modeling approach generalizes the *frailty*-based approach proposed e.g. by Schönbucher (2003b), Duffie et al. (2006) and Azizpour & Giesecke (2008) in the sense that the factors can be seen as dynamic frailty parameters and it allows one to introduce an *information-driven* default contagion effect among the defaults of several issuers. In addition, the formulation of a model under incomplete information has the advantage of avoiding a possibly inadequate specification of the factor model. Notice also that Duffie et al. (2006) provides strong empirical evidence for the assertion that an unobservable stochastic process driving the default intensities is needed on top of the observable covariates in order to explain clustering of defaults in historical data.

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 could be quite difficult under incomplete information and, moreover, lead to unstable parameter values. The main contribution of our work consists

in proposing a coherent and unified methodology that is based on linear filtering and avoids standard calibration techniques for dealing effectively with both these problems.

More precisely, we propose a linear filtering approach to dynamically update the information on the unobserved factors, based on observations of market data. As market data we shall take 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. In the present study the default indicator process is considered only indirectly in the observations via the credit spreads; an approach that is based exclusively on the latter type of observations 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. This approach allows us to obtain parameter estimates which are not just the result of a pure and possibly unstable calibration to market data, but are estimates that evolve according to the filter solution, thus allowing the model to continuously adapt to the actual market situation.

Possible extensions that we are not dealing with in the present paper consist in considering the model under the physical/historical probability measure, specifying the *risk premia*, which characterize the change of measure, as further unobserved stochastic processes (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*.

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 and present some simulation results.

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 $\mathcal{G}_T = \mathcal{G}$ 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 -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). It then follows that the \mathcal{G}_t -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, of which the values are described by an n -dimensional process Ψ_t that is \mathcal{F}_t -adapted and satisfies the following Vasicek-type dynamics

$$d\Psi_t = (C - A\Psi_t) dt + Bdw_t \quad (2)$$

where C, A, B are given vectors/matrices and w_t is a (multivariate) (\mathcal{F}_t, P) -Wiener process. We shall actually assume that $\mathcal{F}_t = \mathcal{F}_t^\Psi = \mathcal{F}_t^w$. The vector Ψ_t may include components with and without economic interpretation; in the next section we shall extend it to include also components that represent observation noises. The Gaussian nature of the process Ψ_t in (2) allows for easier tractability; on the other hand it makes also negative values for the factors possible. The mean reverting character keeps however the probability of negative values low. At the expense of more involved formulae one may also consider CIR-type dynamics for the factors and this guarantees for their positivity.

Next we specify the short rate of interest r_t and the default intensities λ_t^j as affine functions of the factors, namely

$$\begin{cases} r_t = a + b\Psi_t \\ \lambda_t^j = c^j + d^j\Psi_t ; \quad j = 1, \dots, m \end{cases} \quad (3)$$

where a, c^j are constant parameters and b, d^j are n -dimensional row vectors of parameters. This setup allows for correlation between interest rate and default intensity, which (see Schönbucher (2003a)) is a desirable property for a stochastic credit risk model.

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 Duffie & Garleanu (2001), Frey et al. (2006)) 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 $(\mathcal{G}_t)_{0 \leq t \leq T}$. More precisely, the price of a default-free 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)

$$\Pi_{DF}(t, T) = E \left\{ \exp \left[- \int_t^T r_s ds \right] \mid \mathcal{G}_t \right\}, \quad (4)$$

can be expressed as (see McNeil et al. (2005), section 9.5, or Lambertson & Lapeyre (1995), section 6.2.2)

$$\Pi_{DF}(t, T) = \exp [A(t, T) - B(t, T) \Psi_t] \quad (5)$$

where the scalar $A(\cdot)$ and vector $B(\cdot)$ satisfy each a well-known first order ordinary differential equation that for $B(\cdot)$ turns out to be of the Riccati type (the coefficients in these equations depend on the parameters in (2) and (3)). 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), Thm 9.23)

$$\Pi(t, T) = E \left\{ e^{-\int_t^T r_s ds} \mathbf{1}_{\{\tau > T\}} \mid \mathcal{G}_t \right\} = \mathbf{1}_{\{\tau > t\}} E \left\{ e^{-\int_t^T (r_s + \lambda_s) ds} \mid \mathcal{F}_t \right\} \quad (6)$$

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

$$\Pi(t, T) = \mathbf{1}_{\{\tau > t\}} \exp \left[\tilde{A}(t, T) - \tilde{B}(t, T) \Psi_t \right] \quad (7)$$

where, again, $\tilde{A}(\cdot), \tilde{B}(\cdot)$ satisfy each one a well-known first order ordinary differential equation that for $\tilde{B}(\cdot)$ is of the Riccati type and the coefficients in these equations depend again on the parameters in (2) and (3). To see better the impact of these parameters we report here the expressions for $\tilde{B}(\cdot)$ and $\tilde{A}(\cdot)$ for the simpler case of a scalar Ψ_t (the coefficients C, A, B in (2) are thus scalar and so is $\tilde{B}(\cdot)$) where, with only one defaultable firm, (3) is simplified to become

$$\begin{cases} r_t = \Psi_t \\ \lambda_t = \lambda \Psi_t \end{cases}$$

The expressions are

$$\begin{cases} \tilde{B}(t, T) = \frac{(1+\lambda)}{A} (1 - e^{-A(T-t)}) \\ \tilde{A}(t, T) = \frac{[\tilde{B}(t, T) - (1+\lambda)(T-t)][AC - \frac{1}{2}B^2(1+\lambda)]}{A^2} - \frac{B^2 \tilde{B}^2(t, T)}{4A} \end{cases} \quad (8)$$

From (5) and (7) we have that the yield (for $t < T$) of a 0-coupon default-free bond is given by

$$YL(t, T) := -\frac{1}{T-t} \log \Pi_{DF}(t, T) = -\frac{A(t, T)}{T-t} + \frac{B(t, T)}{T-t} \Psi_t \quad (9)$$

and results in an affine function of Ψ_t . Furthermore, 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 $t = \tau$) by

$$\begin{aligned} CS(t, T) &:= -\frac{1}{T-t} \log \left[\frac{\Pi(t, T)}{\Pi_{DF}(t, T)} \right] \\ &= -\frac{1}{T-t} \left(\tilde{A}(t, T) - A(t, T) \right) + \frac{1}{T-t} \left(\tilde{B}(t, T) - B(t, T) \right) \Psi_t; \quad t < \tau \wedge T \end{aligned} \quad (10)$$

again an affine function of Ψ_t .

We conclude this subsection by showing that more general and more liquid 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. We do not consider here other liquid products such as CDOs; notice however that for corporate bonds and CDS the link between their values and the default event of a given firm is much clearer than for CDOs. In view of our purpose here we first recall that a *recovery payment* consists of a non-negative process Z_t , adapted to (\mathcal{F}_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 (recall that we consider P to be a martingale measure)

$$R = E \left\{ \mathbf{1}_{\{\tau > t\}} e^{-\int_t^\tau r_s ds} Z_\tau \mathbf{1}_{\{\tau \leq T\}} \mid \mathcal{G}_t \right\} \quad (11)$$

We show now the following

Lemma 1. *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 bond's (unitary) face value, we have*

$$E \left\{ \mathbf{1}_{\{\tau > t\}} e^{-\int_t^\tau r_s ds} \delta \mathbf{1}_{\{\tau \leq T\}} \mid \mathcal{G}_t \right\} = \delta \left(\mathbf{1}_{\{\tau > t\}} - \Pi(t, T) - \int_t^T f(t, s) \Pi(t, s) ds \right) \quad (12)$$

where $f(t, s)$ denotes the forward rate (for deterministic r_t , the $f(t, s)$ is replaced by r_s).

Proof: Denoting by Q^T the forward measure for maturity T and by E^T the corresponding expectation, notice first that from (see (6))

$$\begin{aligned} \Pi(t, T) &= \mathbf{1}_{\{\tau > t\}} E \left\{ e^{-\int_t^T r_s ds} \mathbf{1}_{\{\tau > T\}} \mid \mathcal{G}_t \right\} = \mathbf{1}_{\{\tau > t\}} E \left\{ e^{-\int_t^T (r_s + \lambda_s) ds} \mid \mathcal{F}_t \right\} \\ &= \mathbf{1}_{\{\tau > t\}} \Pi_{DF}(t, T) E^T \left\{ e^{-\int_t^T \lambda_s ds} \mid \mathcal{F}_t \right\} \end{aligned} \quad (13)$$

we have

$$E^T \left\{ e^{-\int_t^T \lambda_s ds} \mid \mathcal{F}_t \right\} = \begin{cases} \frac{\Pi(t, T)}{\Pi_{DF}(t, T)} & \text{if } \tau > t, \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

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

$$E \left\{ \mathbf{1}_{\{\tau > t\}} e^{-\int_t^\tau r_s ds} \delta \mathbf{1}_{\{\tau \leq T\}} \mid \mathcal{G}_t \right\} = \mathbf{1}_{\{\tau > t\}} \delta E \left\{ \int_t^T \lambda_s e^{-\int_t^s (r_u + \lambda_u) du} ds \mid \mathcal{F}_t \right\} \quad (15)$$

and that, by the Gaussianity of Ψ_t and the fact that λ_t and r_t are affine in Ψ_t , we have

$$E \left\{ \int_t^T \left| \lambda_s e^{-\int_t^s (r_u + \lambda_u) du} \right| ds \mid \mathcal{F}_t \right\} < +\infty.$$

Applying then, in the following order: (15) and Fubini's theorem, a change of measure from P to Q^s , (14) and integration by parts, we obtain the result, namely

$$\begin{aligned} E \left\{ \mathbf{1}_{\{\tau > t\}} e^{-\int_t^\tau r_s ds} \delta \mathbf{1}_{\{\tau \leq T\}} \mid \mathcal{G}_t \right\} &= \mathbf{1}_{\{\tau > t\}} \delta \int_t^T E \left\{ \lambda_s e^{-\int_t^s (r_u + \lambda_u) du} \mid \mathcal{F}_t \right\} ds \\ &= \mathbf{1}_{\{\tau > t\}} \delta \int_t^T \Pi_{DF}(t, s) E^s \left\{ \frac{\lambda_s e^{-\int_t^s \lambda_u du}}{\Pi_{DF}(s, s)} \mid \mathcal{F}_t \right\} ds \\ &= -\mathbf{1}_{\{\tau > t\}} \delta \int_t^T \Pi_{DF}(t, s) \frac{\partial}{\partial s} E^s \left\{ e^{-\int_t^s \lambda_u du} \mid \mathcal{F}_t \right\} ds = -\delta \int_t^T \Pi_{DF}(t, s) \frac{\partial}{\partial s} \left(\frac{\Pi(t, s)}{\Pi_{DF}(t, s)} \right) ds \\ &= -\delta \left[\Pi_{DF}(t, s) \frac{\Pi(t, s)}{\Pi_{DF}(t, s)} \right] \Big|_t^T + \delta \int_t^T \frac{\partial \Pi_{DF}(t, s)}{\partial s} \frac{\Pi(t, s)}{\Pi_{DF}(t, s)} ds \\ &= \delta \left(\mathbf{1}_{\{\tau > t\}} - \Pi(t, T) \right) + \delta \int_t^T \frac{\partial \log \Pi_{DF}(t, s)}{\partial s} \Pi(t, s) ds \\ &= \delta \left(\mathbf{1}_{\{\tau > t\}} - \Pi(t, T) - \int_t^T f(t, s) \Pi(t, s) ds \right). \end{aligned} \quad (16)$$

■

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 times $t_1, \dots, t_n = T$, one has that its value is given by

$$\Pi^{\delta,c}(t, T) = \sum_{i=1}^n c(t_i - t_{i-1}) \Pi(t, t_i) + \Pi(t, T) + \delta \left(\mathbf{1}_{\{\tau > t\}} - \Pi(t, T) - \int_t^T f(t, s) \Pi(t, s) ds \right) \quad (17)$$

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

$$V_P = \sum_{k=1}^n S(t_k - t_{k-1}) \Pi(t, t_k) \quad (18)$$

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

$$\begin{aligned} V_D &= E \left\{ \mathbf{1}_{\{\tau > t\}} e^{-\int_t^\tau r_s ds} (1 - \delta) \mathbf{1}_{\{\tau \leq T\}} \mid \mathcal{G}_t \right\} \\ &= (1 - \delta) \left(\mathbf{1}_{\{\tau > t\}} - \Pi(t, T) - \int_t^T f(t, s) \Pi(t, s) ds \right) \end{aligned} \quad (19)$$

For the (fair) *CDS spread* it then follows

$$S = \frac{(1 - \delta) \left(1 - \Pi(t, T) - \int_t^T f(t, s) \Pi(t, s) ds \right)}{\sum_{k=1}^n (t_k - t_{k-1}) \Pi(t, t_k)} \quad (20)$$

Notice that the *accrued payment* between the last premium payment date 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)).

On one hand the representations (17) and (20) allow to reconstruct 0-coupon 0-recovery bond prices from the more liquid corporate bond prices and CDS spreads and use these reconstructed values to determine, via the relation (see (10)) $CS(t, T) = -\frac{1}{T-t} \log \left[\frac{\Pi(t, T)}{\Pi_{DF}(t, T)} \right]$, observed values for $CS(t, T)$ that, within our modeling framework (see again (10)) are considered as affine functions of the factors Ψ_t . 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) and (10) one can express $\Pi(t, T)$ as (nonlinear) function of the factors Ψ_t , (17) and (20) 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

Some of the factors with an economic interpretation may be directly observable, others may not. Also the factors without economic interpretation, which are essentially introduced only to enhance the flexibility of the model, 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. (2006), Azizpour & Giesecke (2008)). On the other hand this implies that we cannot directly use formulas (5) and (7), and thus neither (17) nor (20) 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 -measurable and this implies by (3) that one can observe a linear combination of the factors. We suppose furthermore that, in addition to the defaults, 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); notice that these latter observations take place only up to the (observed) default of the issuer.

Observe next that, by (9) and (10), in addition to r_t also yields and spreads are affine functions of Ψ_t . If then one has more observations than factors, namely if $1 + p + q > n$, 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 credit default swaps and, even if they could be observed directly, they are affected by bid-ask spreads and 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 $n + \ell > p + q + 1$. In fact, 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 too follow dynamics of the type (2). 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 Ψ_t^* the vector of dimension $(n + \ell)$ obtained by adding the ℓ noise factors to Ψ_t .

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

$$\begin{cases} r_t & = a + b\Psi_t \\ YL(t, T_i) & = \alpha^i(t) + \beta^i(t) \Psi_t^* & i = 1, \dots, p \\ CS(t, T_j) & = \gamma^j(t) + \delta^j(t) \Psi_t^* & j = 1, \dots, q \end{cases} \quad (21)$$

where $\alpha^i(t), \gamma^j(t)$ are deterministic functions of time, $\beta^i(t), \delta^j(t)$ are $(n + \ell)$ -dimensional row vectors composed of deterministic functions of time, and all depend on the parameters in (2) and (3) to be “calibrated” to the market. We can also write more specifically

$$\mathcal{Y}_t = \sigma\{r_s, YL(s, T_i), CS(s, T_j) : s \leq t, i = 1, \dots, p; j = 1, \dots, q\} \vee \mathcal{H}_t$$

and notice that q is reduced by 1 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 filter problem

Our purpose is to compute *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 at the end of 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 (5), (7)) as (exponentially) affine functions of 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 2. *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*

$$\hat{\Pi}(t, T) = E \{ \Pi(t, T; \Psi_t) \mid \mathcal{Y}_t \} \quad (22)$$

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

$$\begin{aligned} B^{-1}(t_1)\hat{\Pi}(t_1, T) &= E \{ B^{-1}(t_1)\Pi(t_1, T; \Psi_{t_1}) \mid \mathcal{Y}_{t_1} \} \\ &= E \{ E \{ B^{-1}(t_2)\Pi(t_2, T; \Psi_{t_2}) \mid \mathcal{G}_{t_1} \} \mid \mathcal{Y}_{t_1} \} \\ &= E \{ B^{-1}(t_2)\Pi(t_2, T; \Psi_{t_2}) \mid \mathcal{Y}_{t_1} \} = E \{ B^{-1}(t_2) E \{ \Pi(t_2, T; \Psi_{t_2}) \mid \mathcal{Y}_{t_2} \} \mid \mathcal{Y}_{t_1} \} \\ &= E \{ B^{-1}(t_2)\hat{\Pi}(t_2, T) \mid \mathcal{Y}_{t_1} \} \end{aligned} \quad (23)$$

■

In order to compute the expression in the right hand side in (22) we need the conditional distribution of Ψ_t , given \mathcal{Y}_t , which is what is called the *filter distribution* at time t and we discuss now how to determine it.

First notice that the observations in (21) are linear with respect to the unobserved factors and some of them may not be affected by any of the factors representing a noise term (or affected by identical noise terms). A direct formulation of the state-observation filtering system, that is, letting Ψ_t^* be the unobserved state, would cause the filtering problem to degenerate. It is however possible to reduce Ψ_t^* to an auxiliary unobserved state process X_t of dimension $(n + \ell) - (p + q + 1)$ and obtain a corresponding non-degenerate linear filtering problem that is as informative as the original problem with state Ψ_t . To this effect we follow Frey & Runggaldier (2008), section 3.2 (see also Fontana (2007)), letting $Y_t := [r_t, Y_t^1, \dots, Y_t^{p+q}]$ where Y_t^i represent the observations $YL(t, T_i)$ and $CS(t, T_j)$. With the so defined process Y_t one can rewrite system (21) as

$$Y_t = \mu_t + M_t \Psi_t^* \quad (24)$$

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, considering only nonredundant observations in (21), can be assumed to be of full rank. We can now prove the following

Proposition 3. *One can choose a time-varying $(n + \ell - p - q - 1, n + \ell)$ 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. Defining then*

$$X_t := L_t \Psi_t^*, \quad (25)$$

the pair (X_t, Y_t) satisfies

$$\begin{cases} dX_t &= (R_t + F_t X_t + G_t Y_t) dt + D_t dw_t \\ dY_t &= (V_t + C_t X_t + K_t Y_t) dt + E_t dw_t \end{cases} \quad (26)$$

for a $(n + \ell - p - q - 1)$ -vector R_t , a $(p + q + 1)$ -vector V_t and time-varying matrices $F_t, G_t, D_t, C_t, K_t, E_t$ of appropriate dimensions that all depend on the parameters in (2) and (3).

Furthermore, there exist time varying matrices Γ_t, Δ_t and a vector ξ_t such that

$$\Psi_t = \Gamma_t X_t + \Delta_t Y_t + \xi_t \quad (27)$$

Proof: 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 , for appropriate matrices $\hat{\Gamma}_t, \hat{\Delta}_t$ it then holds that

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

Denoting by Γ_t, Δ_t the submatrices of $\hat{\Gamma}_t, \hat{\Delta}_t$ formed by their first n rows (that correspond to Ψ_t as subvector of Ψ_t^*), we obtain (27). Using Ito's formula as well as (2), (21) and (28), one then obtains also (26). \blacksquare

Notice that in system (26) X_t is now the unobservable state vector of dimension $(n + \ell - p - q - 1)$ and Y_t is the $(p + q + 1)$ -dimensional vector of the observations. Furthermore, system (26) is of the type of a non-degenerate, linear conditionally Gaussian state-observation system so that the conditional (filter) distribution $p_{X_t | \mathcal{F}_t^Y}$, where $\mathcal{F}_t^Y = \sigma\{Y_s; s \leq t\}$, is Gaussian, i.e.

$$p_{X_t | \mathcal{F}_t^Y} = \mathcal{N}(X_t; m_t, P_t)$$

where the conditional mean m_t and covariance P_t can be computed by the Kalman filter. Notice that, given the representation of Ψ_t in (27), the right-hand side in (22) can be computed as

$$\hat{\Pi}(t, T) = E \{ \Pi(t, T; \Gamma_t X_t + \Delta_t Y_t + \xi_t \mid \mathcal{Y}_t \} \quad (29)$$

Since \mathcal{Y}_t differs from \mathcal{F}_t^Y only at a default time, where the only effect is that q and possibly also ℓ are reduced by 1, the expression (29) can be continued as

$$\begin{aligned} \hat{\Pi}(t, T) &= E \{ \Pi(t, T; \Gamma_t X_t + \Delta_t Y_t + \xi_t \mid \mathcal{F}_t^Y \} \\ &= \int \Pi(t, T; \Gamma_t x + \Delta_t Y_t + \xi_t) p_{X_t | \mathcal{F}_t^Y}(dx) \end{aligned} \quad (30)$$

Notice also that, in the case of default-free and defaultable bonds, $\Pi(t, T; \Psi_t)$ as a function of Ψ_t is exponentially affine so that (30) leads to the conditional moment generating function that, having obtained m_t and P_t from the Kalman filter, gives an explicit expression for $\hat{\Pi}(t, T)$. In the case of a defaultable bond, according to (7) and using the representation (27), one then has for instance

$$\hat{\Pi}(t, T) = \mathbf{1}_{\{\tau > t\}} K(t, T; Y_t, \xi_t) \exp \left[-\hat{B}(t, T) m_t + \frac{1}{2} \hat{B}(t, T) P_t \hat{B}'(t, T) \right] \quad (31)$$

for a suitable $K(\cdot)$ that has an explicit expression as function of its arguments and where $\hat{B}(t, T) = \tilde{B}(t, T) \Gamma_t$. Extensions of the filtering problem are possible to the case of CIR dynamics of the factors and parsimonious parametrizations of the model to induce correlations among the factors (see Fontana (2007)).

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 (3). 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. Elliott et al. (1995)) that we are going to discuss next.

Let θ be the vector composed of all the parameters characterizing the model. Notice that, even if we assume our parameters to be constant with respect to time t , through their continuous updating via the EM algorithm below, they acquire a form of time dependence thereby allowing to track the actual market situation. 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

$$Q(\theta, \theta') := E_{\theta'} \left\{ \log \frac{dP^\theta}{dP^{\theta'}} \mid \mathcal{F}_t^Y \right\} \quad (32)$$

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 (32) it can be shown (see e.g. Campillo & Le Gland (1989)) that, with every iteration, the values of θ obtained from the EM algorithm generate non-decreasing values of the (conditional) log-likelihood function

$$l(\theta) = \log E_{\bar{P}} \left\{ \frac{dP^\theta}{d\bar{P}} \mid \mathcal{F}_t^Y \right\} \quad (33)$$

and thus also of the likelihood function $L(\theta) = \exp\{l(\theta)\}$. In fact, from its definition and using Jensen's inequality, one first has

$$\begin{aligned} l(\theta) - l(\theta') &= \log \frac{E_{\bar{P}} \left\{ \frac{dP^\theta}{dP^{\theta'}} \cdot \frac{dP^{\theta'}}{d\bar{P}} \mid \mathcal{F}_t^Y \right\}}{E_{\bar{P}} \left\{ \frac{dP^{\theta'}}{d\bar{P}} \mid \mathcal{F}_t^Y \right\}} = \log E_{\theta'} \left\{ \frac{dP^\theta}{dP^{\theta'}} \mid \mathcal{F}_t^Y \right\} \\ &\geq E_{\theta'} \left\{ \log \frac{dP_\theta}{dP_{\theta'}} \mid \mathcal{F}_t^Y \right\} = Q(\theta, \theta') \end{aligned} \quad (34)$$

Denote next by $\hat{\theta}_p$ the maximizing value of θ in $Q(\theta, \theta')$ at the generic step p so that $Q(\theta, \hat{\theta}_p)$ as a function of θ is maximized by $\theta = \hat{\theta}_{p+1}$ while, by its very definition, $Q(\hat{\theta}_p, \hat{\theta}_p) = 0$. From

(34) it then follows that

$$l(\hat{\theta}_{p+1}) \geq l(\hat{\theta}_p) + Q(\hat{\theta}_{p+1}, \hat{\theta}_p) \geq l(\hat{\theta}_p) + Q(\hat{\theta}_p, \hat{\theta}_p) = l(\hat{\theta}_p) \quad (35)$$

Under appropriate regularity conditions (see Wu (1983), Thm.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 (see e.g. Cont & Tankov (2004)).

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. In general, this difficulty 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. Our filtering problem concerns the state-observation system (26) but a direct application of the EM algorithm to the estimation of the parameters in (26) leads to the following problems:

- i) The diffusion coefficients, namely the matrices D_t and E_t , 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 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)).
- ii) Some of the coefficients in the system of equations to be solved during the maximization step can only be obtained by solving also the so-called *smoothing problem*, namely computing $E\{X_s | \mathcal{F}_t^Y\}$ for $s < t$, and this is computationally more demanding than just the filtering problem.

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 state-observation system (26). 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 alternating between several parallel state-observation systems, as described below for the example, and using results from Elliott & Krishnamurthy (1999) for discrete time linear Gaussian models, the parameter estimates can be computed directly as functions of the filter solution.

4.2 Example

We illustrate here the discrete time alternating algorithm for the case of a model as described above (Sections 2 and 3) where $n = 3, p = q = 1, \ell = 1$.

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

$$d\Psi_t^i = (\alpha_i - \beta_i \Psi_t^i) dt + \sigma_i dw_t^i \quad (36)$$

and the additional component Ψ_t^4 in Ψ_t^* is

$$d\Psi_t^4 = dw_t^4 \quad (37)$$

i.e. $\alpha_4 = \beta_4 = 0, \sigma_4 = 1$ and w_t^i ($i = 1, \dots, 4$) are independent (\mathcal{F}_t, P) -Wiener processes. Furthermore, we let (3) take the specific form

$$\begin{cases} r_t &= \Psi_t^1 + \Psi_t^2 \\ \lambda_t &= \lambda (\Psi_t^2 + \Psi_t^3) \end{cases} \quad (38)$$

where we have only one intensity since, having assumed $q = 1$, we observe the spread of only one defaultable bond that corresponds to just one issuer that may default. From (36) we have, for $i = 1, 2, 3$, that $\Psi_t^i \sim \mathcal{N}(m_t^i, V_t^i)$ are independent Gaussian with unconditional mean m_t^i and variance V_t^i that depend on the parameters $\alpha_i, \beta_i, \sigma_i$ in (36). Define

$$u_t^i := \Psi_t^i - m_t^i \sim \mathcal{N}(0, V_t^i), \quad i = 1, 2, 3 \quad (39)$$

According to Proposition 3 there exists a 1-dimensional process X_t that summarizes the effect of the entire vector Ψ_t of the factors. We may take as X_t any of the components of Ψ_t which leads to three possible state-observation systems (see (42)-(44) below), but not all the observations depend necessarily on each of the components of Ψ_t .

Introducing now a time discretization with step Δ and using the Euler scheme, for the generic Ψ^i , $i = 1, 2, 3$ we obtain the discrete time dynamics

$$\Psi_{k+1}^i = \alpha_i \Delta + (1 - \beta_i \Delta) \Psi_k^i + \sigma_i \sqrt{\Delta} \varepsilon_{k+1}^i \quad (40)$$

where k denotes the generic time step and, for each $i = 1, 2, 3$, the ε_k^i form sequences of i.i.d. standard Gaussian random variables. Since $p = q = 1$ and we have just one issuer, we denote by T the maturity for the yield and the spread and put $Y_k^1 = YL(k\Delta, T)$, $Y_k^2 = CS(k\Delta, T)$ so that, at the generic time step k , we have the system

$$\left\{ \begin{array}{l} \Psi_{k+1}^i = \alpha_i \Delta + (1 - \beta_i \Delta) \Psi_k^i + \sigma_i \sqrt{\Delta} \varepsilon_{k+1}^i, \quad i = 1, 2, 3 \\ r_k = \Psi_k^1 + \Psi_k^2 \\ Y_k^1 = a_k + b_k^1 \Psi_k^1 + b_k^2 \Psi_k^2 + \gamma \Psi_k^4 \\ Y_k^2 = \tilde{a}_k + \tilde{b}_k^2 \Psi_k^2 + \tilde{b}_k^3 \Psi_k^3 + \rho \Psi_k^4 \end{array} \right. \quad (41)$$

The coefficients in the equations for Y_k^1 , Y_k^2 depend, see (9) and (10), on the functions $A(t, T)$, $B(t, T)$, $\tilde{A}(t, T)$, $\tilde{B}(t, T)$ that, through the system of ordinary differential equations that they have to satisfy, depend in turn on the parameters in (2) and (3) (see e.g. the case illustrated in (8)). In the present example the relations (2) and (3) correspond to (36) and (38) and the parameters are $\alpha_i, \beta_i, \sigma_i$; $i = 1, 2, 3$ as well as λ in addition to γ and ρ . Notice that, even though the spread derives from both the default-free and defaultable bonds, for our model and in particular by (38), the observations Y_k^2 do not depend on Ψ_k^1 (in fact, Ψ_k^1 affects only the short rate).

From system (41) we now deduce three parallel systems obtained by choosing as X_t the components $\Psi_t^1, \Psi_t^2, \Psi_t^3$ respectively and keeping only those observations that, after using also (39), depend only on the chosen factor Ψ_t^i . The systems are

$$\left\{ \begin{array}{l} \Psi_{k+1}^1 = \alpha_1 \Delta + (1 - \beta_1 \Delta) \Psi_k^1 + \sigma_1 \sqrt{\Delta} \varepsilon_{k+1}^1 \\ r_k = \Psi_k^1 + m_k^2 + u_k^2 \\ Y_k^1 = a_k + b_k^1 \Psi_k^1 + b_k^2 m_k^2 + b_k^2 u_k^2 + \gamma \Psi_k^4 \end{array} \right. \quad (42)$$

$$\left\{ \begin{array}{l} \Psi_{k+1}^2 = \alpha_2 \Delta + (1 - \beta_2 \Delta) \Psi_k^2 + \sigma_2 \sqrt{\Delta} \varepsilon_{k+1}^2 \\ r_k = \Psi_k^2 + m_k^1 + u_k^1 \\ Y_k^1 = a_k + b_k^1 m_k^1 + b_k^2 \Psi_k^2 + b_k^1 u_k^1 + \gamma \Psi_k^4 \\ Y_k^2 = \tilde{a}_k + \tilde{b}_k^2 \Psi_k^2 + \tilde{b}_k^3 m_k^3 + \tilde{b}_k^3 u_k^3 + \rho \Psi_k^4 \end{array} \right. \quad (43)$$

$$\left\{ \begin{array}{l} \Psi_{k+1}^3 = \alpha_3 \Delta + (1 - \beta_3 \Delta) \Psi_k^3 + \sigma_3 \sqrt{\Delta} \varepsilon_{k+1}^3 \\ Y_k^2 = \tilde{a}_k + \tilde{b}_k^2 m_k^2 + \tilde{b}_k^3 \Psi_k^3 + \tilde{b}_k^2 u_k^2 + \rho \Psi_k^4 \end{array} \right. \quad (44)$$

In addition, we consider also the following two systems, always derived from (41), but with the purpose of focusing on the parameters (γ, ρ, λ) .

$$\left\{ \begin{array}{l} \Psi_{k+1}^1 = \alpha_1 \Delta + (1 - \beta_1 \Delta) \Psi_k^1 + \sigma_1 \sqrt{\Delta} \varepsilon_{k+1}^1 \\ Y_k^1 = a_k + b_k^2 r_k + (b_k^1 - b_k^2) \Psi_k^1 + \gamma \Psi_k^4 \end{array} \right. \quad (45)$$

$$\left\{ \begin{array}{l} \Psi_{k+1}^2 = \alpha_2 \Delta + (1 - \beta_2 \Delta) \Psi_k^2 + \sigma_2 \sqrt{\Delta} \varepsilon_{k+1}^2 \\ \Psi_{k+1}^3 = \alpha_3 \Delta + (1 - \beta_3 \Delta) \Psi_k^3 + \sigma_3 \sqrt{\Delta} \varepsilon_{k+1}^3 \\ Y_k^2 = \tilde{a}_k + \tilde{b}_k^2 \Psi_k^2 + \tilde{b}_k^3 \Psi_k^3 + \rho \Psi_k^4 \end{array} \right. \quad (46)$$

4.2.1 The EM algorithm itself

Notice that each of the systems (42)-(46) is of the discrete-time linear-Gaussian type so that the Kalman filter can be applied to determine the filter distribution that is Gaussian and thus determined 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.

First notice that each of the systems (42)-(46) is of the form

$$\left\{ \begin{array}{l} X_{k+1} = R_k(\theta) + F_k(\theta) X_k + D_{k+1}(\theta) \varepsilon_{k+1} \\ Y_k = V_k(\theta) + C_k(\theta) X_k + E_k(\theta) \eta_k \end{array} \right. \quad (47)$$

where X_k is the unobservable component of the state-observation system (corresponds to one or more components of Ψ_k) and Y_k is a subvector of (r_k, Y_k^1, Y_k^2) . In (47) we have made explicit the dependence of the coefficients on the parameter vector θ and notice that ε_k and η_k are independent

sequences of independent Gaussian variables/vectors. Notice next that the fundamental quantity underlying the EM algorithm is $Q(\theta, \theta')$ in (32), where we may write

$$\frac{dP^\theta}{dP^{\theta'}} = \frac{dP^\theta}{d\Lambda} \cdot \frac{d\Lambda}{dP^{\theta'}} \quad (48)$$

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

$$\begin{aligned} Q(\theta, \theta') &= -\sum_{l=1}^k \log |D_l(\theta)| - \sum_{l=0}^k \log |E_l(\theta)| \\ &- \frac{1}{2} E_{\theta'} \left\{ \sum_{l=0}^{k-1} (X_{l+1} - F_l(\theta) X_l - R_l(\theta))' (D_{l+1}(\theta) D_{l+1}'(\theta))^{-1} (X_{l+1} - F_l(\theta) X_l - R_l(\theta)) \mid \mathcal{F}_k^Y \right\} \\ &- \frac{1}{2} E_{\theta'} \left\{ \sum_{l=0}^k (Y_l - C_l(\theta) X_l - V_l(\theta))' (E_l(\theta) E_l'(\theta))^{-1} (Y_l - C_l(\theta) X_l - V_l(\theta)) \mid \mathcal{F}_k^Y \right\} \\ &+ E_{\theta'} \{ Z_k(\theta') \mid \mathcal{F}_k^Y \} \end{aligned} \quad (49)$$

where $Z_k(\theta')$ 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 (49) 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

$$\begin{aligned} E_{\theta'} \left\{ \sum_{l=0}^k X_l' f_l(\theta) \mid \mathcal{F}_k^Y \right\} \quad , \quad E_{\theta'} \left\{ \sum_{l=0}^k X_l' \bar{f}_l(\theta) X_l \mid \mathcal{F}_k^Y \right\} \\ E_{\theta'} \left\{ \sum_{l=0}^{k-1} X_{l+1}' g_l(\theta) X_{l+1} \mid \mathcal{F}_k^Y \right\} \quad , \quad E_{\theta'} \left\{ \sum_{l=0}^{k-1} X_{l+1}' \bar{g}_l(\theta) X_l \mid \mathcal{F}_k^Y \right\} \\ E_{\theta'} \left\{ \sum_{l=0}^k X_l' h_l(\theta) Y_l \mid \mathcal{F}_k^Y \right\} \end{aligned} \quad (50)$$

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 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 (47) by setting $\theta = \theta'$. We show here the expressions for the first two quantities in (50), for the remaining ones the expressions are analogous and can be found in Fontana (2007).

For the first quantity we have

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

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

$$\begin{cases} \Sigma_k &= [D_k(\theta)D'_k(\theta)]^{-1} F'_k(\theta)\sigma_k^{-1} \\ S_k &= \sigma_{k+1}^{-1} (P_{k|k})^{-1} \hat{X}_{k|k} \end{cases}$$

For the second quantity one has instead

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

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

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

4.2.2 The alternating iterative algorithm

While step 1 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 2, namely the maximization of $Q(\theta, \theta')$ with respect to θ or, equivalently, the solution of the system $\frac{\partial Q(\theta, \theta')}{\partial \theta} = 0$. For the model in the present example, which is synthesized in (41) and leads to the parallel state-observation systems (42)-(46), the parameters are $(\alpha_i, \beta_i, \sigma_i)$, $i = 1, 2, 3$ and (γ, ρ, λ) . The latter enter only the coefficients in the observation equations and here the dependence on λ is nonlinear (see (6), (7), (8), (10)), while the dependence of the coefficients in the state dynamics on $(\alpha_i, \beta_i, \sigma_i)$ is linear. For what concerns the maximization with respect to $(\alpha_i, \beta_i, \sigma_i)$, an advantage of having the three parallel systems (42)-(44) is that in each of these systems the state dynamics depend on $(\alpha_i, \beta_i, \sigma_i)$ for just one of the indices $i = 1, 2, 3$ and this 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 five systems with all twelve parameters simultaneously until a stopping criterion is met. One can however alternate the iterations between the individual systems (42)-(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 (42)-(46) is then the following.

- For system (42): $\theta = (\alpha_1, \beta_1, \sigma_1)$; $\theta' = (\alpha_i, \beta_i, \sigma_i, \gamma, \rho, \lambda)_{i=2,3}$;
- for system (43): $\theta = (\alpha_2, \beta_2, \sigma_2)$; $\theta' = (\alpha_i, \beta_i, \sigma_i, \gamma, \rho, \lambda)_{i=1,3}$;
- for system (44): $\theta = (\alpha_3, \beta_3, \sigma_3)$; $\theta' = (\alpha_i, \beta_i, \sigma_i, \gamma, \rho, \lambda)_{i=1,2}$;
- for system (45): $\theta = \gamma$; $\theta' = (\alpha_i, \beta_i, \sigma_i)_{i=1,2,3}$;
- for system (46): $\theta = (\rho, \lambda)$; $\theta' = (\alpha_i, \beta_i, \sigma_i)_{i=1,2,3}$.

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 (42),(43),(46)

and that formed by (44), (45). More precisely, considering that the global vector of parameters is given by $\theta = (\alpha_1, \beta_1, \sigma_1, \dots, \alpha_3, \beta_3, \sigma_3, \gamma, \rho, \lambda)$, we have the following

Algorithm:

0. Initialize the algorithm with a guess θ_0 for the entire vector θ of the parameters.
1. Apply in parallel on each of the systems (42),(43),(46) the EM algorithm to estimate $(\alpha_i, \beta_i, \sigma_i)_{i=1,2}$ and (ρ, λ) while keeping $(\alpha_3, \beta_3, \sigma_3, \gamma)$ fixed at their previous levels. The algorithm iterates through the two EM steps (expectation and maximization) until a stopping criterion is met, thereby producing estimates $(\hat{\alpha}_i, \hat{\beta}_i, \hat{\sigma}_i)_{i=1,2}, \hat{\rho}, \hat{\lambda}$
2. Apply in parallel on each of the systems (44),(45) the EM algorithm to estimate $(\alpha_3, \beta_3, \sigma_3, \gamma)$ keeping $(\hat{\alpha}_i, \hat{\beta}_i, \hat{\sigma}_i)_{i=1,2}, \hat{\rho}, \hat{\lambda}$ fixed at their previously estimated values. The algorithm iterates through the two EM steps until a stopping criterion is met, thereby producing estimates $(\hat{\alpha}_3, \hat{\beta}_3, \hat{\sigma}_3, \hat{\gamma})$.
3. Reset θ at the most recently estimated levels and return to step 1. Terminate the entire algorithm as soon as a global stopping criterion is met.

For the initial guess θ_0 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 parameter λ one may consider the approximation for the intensity, frequently used in practice, whereby one puts $\lambda_t \approx \frac{S_t}{LGD}$ with S_t a CDS spread and LGD the loss given default.

The advantage of the proposed procedure is that the EM algorithm is applied in the simplest possible context, namely when the system is of the discrete time linear-Gaussian type with uncorrelated noises in the state and the observations. No smoothing procedure is needed.

4.2.3 Simulation results

In order to verify the convergence of the estimated parameter values, the just described algorithm has been applied to our example taking as observations simulated data. For given values of the parameters in θ (in the tables below referred to as “true values”) a sequence of tuples $(\Psi_k^1, \Psi_k^2, \Psi_k^3, r_k, Y_k^1, Y_k^2)$ was generated according to (41) for $k = 0, \dots, 100$ and the values of (r_k, Y_k^1, Y_k^2) recorded.

Since for our simulation purposes there is no reason to prefer certain parameter values to others, the initial value θ_0 for θ was randomly generated. The alternating iterative 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 1000 iterations. The results of 15 alternating iterative steps are reported in the following two tables. The first table shows the evolution over the 15 alternating steps (simply called “Iteration”) of the estimated values of $(\alpha_i, \beta_i, \sigma_i)_{i=1,2}$ and (ρ, λ) . The second table shows the evolution of the estimated values of $(\alpha_3, \beta_3, \sigma_3, \gamma)$. The “alternation” is to be interpreted in the sense that, after the first “iteration” in the first table one moves to the first “iteration” in the second table, from there to the second “iteration” in the first table and so on for 15 times.

From these simulation results one can see that most of the parameters are estimated with sufficient accuracy after 15 iterations and that, for the crucial parameter λ , the sequence of estimated values approaches steadily the true value as the iterations proceed.

Parameter	α_1	β_1	σ_1	α_2	β_2	σ_2	λ	ρ
True value	0.5	0.3	0.01	0.5	0.3	0.01	0.3	0.05
Initial condition	0.5965	0.9333	0.1440	0.2802	0.0382	0.0148	0.1538	0.1755
Iteration 1	0.4744	0.2951	0.0131	0.5084	0.2887	0.0131	0.3427	0.0498
Iteration 2	0.4819	0.2953	0.0131	0.5008	0.2881	0.0130	0.3414	0.0495
Iteration 3	0.4818	0.2954	0.0131	0.5007	0.2880	0.0130	0.3401	0.0496
Iteration 5	0.4816	0.2956	0.0131	0.5010	0.2879	0.0130	0.3309	0.0496
Iteration 10	0.4813	0.2958	0.0131	0.5013	0.2877	0.0130	0.3227	0.0497
Iteration 15	0.4813	0.2959	0.0131	0.5014	0.2876	0.0130	0.3191	0.0498

Table 1:

Parameter	α_3	β_3	σ_3	γ
True value	0.5	0.3	0.01	0.05
Initial condition	0.4773	0.4130	0.0764	0.1458
Iteration 1	0.4814	0.4022	0.0070	0.0498
Iteration 2	0.4928	0.4094	0.0065	0.0498
Iteration 3	0.5096	0.4145	0.0062	0.0498
Iteration 5	0.5161	0.4208	0.0062	0.0498
Iteration 10	0.5163	0.4222	0.0061	0.0498
Iteration 15	0.5164	0.4223	0.0061	0.0498

Table 2:

References

- AZIZPOUR, S. & GIESECKE, K. (2008), Self-exciting corporate defaults. Working Paper, Stanford University.
- BIELECKI, T.R., JEANBLANC, M. & RUTKOWSKI, M. (2004), 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 Verlag, Berlin-Heidelberg.
- CAMPILLO, F. & LE GLAND, F. (1989), MLE for partially observed diffusions: direct maximization vs. the EM Algorithm, *Stochastic Processes and Their Applications*, 57(1):41–59.
- CONT, R. & TANKOV, P. (2004), Non-Parametric Calibration of Jump-Diffusion Option Pricing Models, *Journal of Computational Finance*, 7(3): 1–49.
- DUFFIE, D. & GARLEANU, N. (2001), Risk and valuation of collateralized debt obligations, *Financial Analysts Journal*, 33: 245–274.
- DUFFIE, D., ECKNER, A., HOREL, G. & SAITA, L. (2006), Frailty correlated default. Working Paper, Stanford University.
- ELLIOTT, R.J., AGGOUN, L., & MOORE, J.B. (1995), *Hidden Markov Models: Estimation and Control*, Springer Verlag, New York.

- ELLIOTT, R.J. & KRISHNAMURTHY, V. (1999), New Finite Dimensional Filters for Parameter Estimation of Discrete-Time Linear Gaussian Models, *IEEE Transactions on Automatic Control*, 44(5): 938–951.
- FONTANA, C. (2007), Affine Multi-factor Credit Risk Models Under Incomplete Information: Filtering and Parameter Estimation. Thesis, University of Padova.
- FREY, R., PROSDOCIMI, C. & RUNGALDIER, W.J. (2006), 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.
- FREY, R. & RUNGALDIER, W.J. (2008), Nonlinear Filtering in Models for Interest-Rate and Credit Risk. Submitted to: *Handbook on Nonlinear Filtering (D.Crisan, B.Rozovski, eds.)*, Oxford University Press.
- GOMBANI, A., JASCHKE, S. R. & RUNGALDIER, W.J. (2005), A Filtered No Arbitrage Model for Term Structures from Noisy Data, *Stochastic Processes and their Applications*, 115: 381–400.
- LAMBERTON, D. & LAPEYRE, B. (1995), *An Introduction to Stochastic Calculus Applied to Finance*, Chapman and Hall, London.
- MCNEIL, A.J., FREY, R. & EMBRECHTS, P. (2005), *Quantitative Risk Management*, Princeton University Press, Princeton.
- RUNGALDIER, W.J. (2004), Estimation via Stochastic Filtering in Financial Market Models, *Contemporary Mathematics*, 351: 309–318.
- SCHÖNBUCHER, P.J. (2003A), *Credit Derivatives Pricing Models: Models, Pricing and Implementation*, Wiley, Chichester.
- SCHÖNBUCHER, P.J. (2003B), Information-Driven Default Contagion. Working paper, Department of Mathematics, ETH Zürich.
- WU, C.F.J., (1983), On the Convergence Properties of the EM Algorithm, *The Annals of Statistics*, 11(1): 95–103.