IMI Bank – PDG Internal Report 3/98

On three filtering problems arising in mathematical finance *

Damiano Brigo
Product Development Group
IMI Bank, San Paolo IMI Group
Corso Matteotti 6
20121 Milano, Italy
Fax: 39 02 7601 9324
E-mail: brigo@bimimi.it

Bernard Hanzon
Dept. Econometrics
Free University Amsterdam
De Boelelaan 1105, 1081 HV
Amsterdam, The Netherlands
Tel: +31-20-4446017
E-mail: bhanzon@econ.vu.nl

Abstract

Three situations in which filtering theory is used in mathematical finance are illustrated at different levels of detail. The three problems originate from the following different works:

1) On estimating the stochastic volatility model from observed bilateral exchange rate news, by R. Mahieu, and P. Schotman;

2) A state space approach to estimate multi-factors CIR models of the term structure of interest rates, by A.L.J. Geyer, and S. Pichler;

3) Risk–minimizing hedging strategies under partial observation in pricing financial derivatives, by P. Fischer, E. Platen, and W. J. Runggaldier;

In the first problem we propose to use a recent nonlinear filtering technique based on geometry to estimate the volatility time series from observed bilateral exchange rates. The model used here is the stochastic volatility model. The filters that we propose are known as projection filters, and a brief derivation of such filters is given. The second problem is introduced in detail, and a possible use of different filtering techniques is hinted at. In fact the filters used for this problem in 2) and part of the literature can be interpreted as projection filters and we will make some remarks on how more general and possibly more suitable projection filters can be constructed. The third problem is only presented shortly.

*This work was developed while the first named author was working at the Risk Management department of Cariplo Bank. A related paper appeared later on in: *Insurance. Mathematics and Economics*, 22(1) (1998) pp. 53-64.

Key words: Mathematical Finance, Stochastic Volatility Model, Filtering Theory, Projection Filter, Interest Rates, Cox–Ingersoll–Ross Model, Quasi Maximum Likelihood, Risk Minimizing Hedging Strategies, Partial Observation.

1 Introduction

The filtering problem consists of estimating a stochastic process X_t representing an unobserved signal, on the basis of the past and present observations $\{Y_s : \ 0 \le s \le t\}$ of a related measurement process Y. The information given by the measurement process up to time t is represented by the σ-algebra $\mathcal{Y}_t$ generated by $\{Y_s : \ 0 \le s \le t\}$. For a quick introduction to the filtering problem see Davis and Marcus (1981) [10]. For a more complete treatment see Liptser and Shiryayev (1978) [21] from a mathematical point of view or Jazwinski (1970) [19] for a more applied perspective. The solution of the filtering problem is the conditional density $p_{X_t|\mathcal{Y}_t}$ of the signal X_t given the observations $\mathcal{Y}_t$. Such a solution in general takes its values in an infinite dimensional function space in an essential way, as proven in Chaleyat-Maurel and Michel (1984) [9]. As a consequence, in general the filter cannot be implemented by an algorithm which updates only a finite number of parameters. This means that there can be no finite-memory computer implementation. An important exception is the linear-Gaussian case, where the solution $p_{X_t|\mathcal{Y}_t}$ is Gaussian at all time instants, and as such can be parameterized by mean and variance. This is the well known Kalman filter.

In the present paper we investigate three possible roles of filtering theory in mathematical finance.

The first problem concerns the stochastic volatility models. In recent applications, time varying volatility of financial time series has been modelled according to the stochastic volatility model, where the variance is considered to be a stochastic process representing an unobserved component. There are several reasons for which such a model represents a convenient choice: among them, the fact that such models are related to the type of diffusion processes one encounters in finance (asset pricing theory, see Melino and Turnbull (1990) [23]). Once the type of model is chosen, there are two problems to be solved:

i) estimate the model parameters on the basis of the observed bilateral exchange rates;

ii) estimate the volatility time series on the basis of the observed bilateral exchange rates.

We develop point ii) by suggesting a different approach based on the projection filter of Brigo, Hanzon and Le Gland (1995) [7], (1997) [8].

We continue by considering as a second problem the state space approach of Geyer and Pichler (1996) [15]. Such an approach is used to estimate and test multi-factors Cox-Ingersoll-Ross (CIR) models of the term structure of interest rates. We concentrate on the estimation procedure. We report the quasi-maximum-likelihood approach combined with a Kalman filter as suggested by Geyer and Pichler, and we also hint at a possible completely Bayesian approach which is sometimes used in system identification.

This state–space approach is convenient for several reasons. The model is estimated, as in the classical cross–section approach, from observations of yields. However, in the state–space approach yields are modelled by taking in account some noise. In this way, market imperfections and deviations from the true model are taken in account. Other advantages are listed in the section of the paper devoted to this approach, and are presented in larger detail in Geyer and Pichler (1996) [15].

The third problem presented concerns risk–minimizing hedging strategies under partial observation in pricing financial derivatives, and is reported as from Fischer, Platen and Runggaldier (1996) [13]. This result is reported and commented in a concise fashion, since it has been thoroughly developed by the authors. It is an excellent example of how filtering theory can fit nicely the mathematical-finance setup, and such examples are rare in the literature.

2 On estimating the stochastic volatility from observed bilateral exchange rate news

2.1 Introduction

The main problem econometricians face when dealing with a stochastic volatility model is the intractability of the likelihood function. In fact, the function turns out to involve a multiple integration, due to the unobserved stochastic variance. One can try to remedy this situation by using a quasi maximum likelihood (QML) method. Another possible remedy is the method of moments estimation (MME). Unfortunately, it has become clear that both methods are not always reliable (see Jacquier, Polson and Rossi (1994) [18] and Andersen (1994) [2]). In Mahieu and Schotman (1997) [22] a study of several possible estimation techniques is presented, and once the model has been estimated a Kalman smoother is applied to estimate the volatility time series. In order to do this, the model is transformed into a linear one and approximations are made to express the new additive noise, whose exact distribution is a log chi-squared. Some possibilities include the approximation of such new noise by a Gaussian of mean -1.27 and variance $\pi^2/2$ (QML). Another possible choice is to approximate the new noise via a mixture of Gaussian densities which should approximate the log chi-squared distribution and other possible noise-distributions in a rather satisfactory way. In Mahieu and Schotman (1997) [22] an application of all the mentioned techniques to financial data is considered, and conclusions are drawn. In the following we suggest a different possible approach to the estimation of the volatility time series from observed bilateral exchange rates. Once the model has been estimated, instead of transforming the original (nonlinear) stochastic volatility model into a linear one and approximating the log chi–squared noise, we keep the original *nonlinear* system with Gaussian white noise and we propose to adopt nonlinear filtering techniques in order to estimate the volatility. The nonlinear filters we use are the projection filters, which were defined and investigated in continuous time in Hanzon (1987) [16], Hanzon and Hut (1991) [17], Brigo (1995) [4], (1996) [5], [6], and Brigo, Hanzon and Le Gland (1995) [7],

(1997) [8]. In this paper we give a short derivation of the projection filter in discrete time, and we apply the theory for discrete time projection filters to the stochastic volatility model.

In general, our method features the advantage of fully taking in account the nonlinear nature of the model adopted. We do not transform the model, so that, once it has been estimated, the only approximation involved in the estimation of the volatility time series is in the filtering technique adopted. In a near future, we plan to analyze the quality of such approximation by means of auxiliary quantities associated to the projection filter.

2.2 Finite dimensional approximation via minimization of the Kullback–Leibler information

In this section we introduce briefly the Kullback-Leibler information and we explain its importance for our problem. Suppose we are given the space H of all the densities of probability measures on the real line equipped with its Borel field, which are absolutely continuous w.r.t. the Lebesgue measure. Then define

$$D(p_1, p_2) := E_{p_1}\{\log p_1 - \log p_2\} \geq 0, \quad p_1, p_2 \in H, \tag{1}$$

where in general

$$E_p\{\phi\} = \int \phi(x) p(x) dx, \quad p \in H.$$

The above quantity is the well-known Kullback-Leibler information (KLI). Its non-negativity follows from the Jensen inequality. It gives a measure of how much the density p_2 is displaced w.r.t. the density p_1. We remark the important fact that D is not a distance: in order to be a metric, it should be symmetric and satisfy the triangular inequality, which is not the case. However, the KLI features many properties of a distance in a generalized geometric setting (see for instance Amari (1985) [1]). For example, it is well-known that the KLI is infinitesimally equivalent to the Fisher information metric around every point of a finite–dimensional manifold of densities such as $EM(c)$ defined below. Consider a finite dimensional manifold of exponential probability densities such as

$$\begin{aligned} EM(c) &= \{p(\cdot, \theta) : \theta \in \Theta \subset I\!R^m\}, \quad \Theta \text{ open in } I\!R^m, \\ p(\cdot, \theta) &= \exp[\theta_1 c_1(\cdot) + \ldots + \theta_m c_m(\cdot) - \psi(\theta)], \end{aligned} \tag{2}$$

expressed w.r.t the expectation parameters η defined by

$$\eta_i(\theta) = E_{p(\cdot,\theta)}\{c_i\} = \partial_{\theta_i}\psi(\theta), \quad i = 1, .., m \tag{3}$$

(see for example Brigo, Hanzon and Le Gland (1997) [7] for more details). We define $p(x; \eta(\theta)) := p(x, \theta)$ (the semicolon identifies the parameterization). Now suppose we are given a density $p \in H$, and we want to approximate it by a density of the finite dimensional manifold $EM(c)$. It seems then reasonable to find a density $p(\cdot, \theta)$ in $EM(c)$

which minimizes the Kullback Leibler information $D(p,.)$. Compute

$$\begin{aligned}
\min_\theta D(p,p(\cdot,\theta)) &= \min_\theta\{E_p[\log p-\log p(\cdot,\theta)]\} \\
&= E_p\log p-\max_\theta\{\theta_1 E_p c_1+...+\theta_m E_p c_m-\psi(\theta)\} \\
&= E_p\log p-\max_\theta V(\theta), \\
V(\theta) &:= \theta_1 E_p c_1+...+\theta_m E_p c_m-\psi(\theta).
\end{aligned}$$

It follows immediately that a necessary condition for the minimum to be attained at θ^* is $\partial_{\theta_i}V(\theta^*)=0,\quad i=1,...,m$ which yields

$$E_p c_i-\partial_{\theta_i}\psi(\theta^*)=E_p c_i-E_{p(\theta^*)}c_i=0,\qquad i=1,..,m$$

i.e. $E_p c_i=\eta_i(\theta^*),\quad i=1,..,m$. This last result indicates that according to the Kullback Leibler information, the best approximation of p in the manifold $EM(c)$ is given by the density of $EM(c)$ which shares the same c_i expectations (c_i-moments) as the given density p. This means that in order to approximate p we only need its c_i moments, $i=1,2,..,m$.

One can look at the problem from the opposite point of view. Suppose we decide to approximate the density p by taking in account only its m c_i–moments. It can be proved (see Kagan, Linnik, and Rao (1973) [20], Theorem 13.2.1) that the maximum entropy distribution which shares the c–moments with the given p belongs to the family $EM(c)$.

Summarizing: If we decide to approximate by using c–moments, then entropy analysis supplies arguments to use the family $EM(c)$; and if we decide to use the approximating family $EM(c)$, Kullback–Leibler says that the "closest" approximating density in $EM(c)$ shares the c–moments with the given density.

2.3 The stochastic volatility model

Let $\{S_t,\ t\in T\},\ \ T=\{0,1,2,3,...\}$ be a stochastic sequence describing bilateral exchange rates in time, and define $Y_t:=\log S_{t+1}-\log S_t,\quad t\in T$. Assuming that the change Y_t of $\log S_t$ is unpredictable, the standard stochastic (logarithmic autoregressive) volatility model (SVM) is given by

$$\begin{aligned}
X_{t+1} &= \rho X_t+\sigma W_{t+1}, \\
Y_t &= \exp(\frac{X_t+\gamma}{2})V_t,
\end{aligned}\tag{4}$$

where $\{W_s,\quad s\in T\}$ and $\{V_s,\quad s\in T\}$ are independent standard Gaussian white noise processes and ρ,σ,γ are real constants. Usually the initial condition X_0 features a non informative density p_{X_0}. In such models the exchange rate features a fat tailed distribution due to the mixing of V_t and $\exp[(X_t+\gamma)/2]$. Consider the following nonlinear filtering problem:

Estimate the stochastic volatility time series $\exp[(X_t+\gamma)/2]$ *at time t from the following observations*

$$Y_0^t:=\{Y_s,s\in T,s\le t\}\tag{5}$$

of the changes in the logarithms of the bilateral exchange rates up to time t.

The general solution of such a problem consists of the conditional probability density $p_{X_t|Y_0^t}$, whose knowledge allows one to compute, among other estimates, the minimum mean square error estimate $E\{\exp[(X_t+\gamma)/2]|Y_0^t\}$ of the stochastic volatility. Such conditional densities obey the following Bayes formula:

$$p_{X_{t+1}|Y_0^{t+1}}(x) \quad = \quad \frac{p_{Y_{t+1}|X_{t+1}}(Y_{T+1};x)\int_{-\infty}^{+\infty} p_{X_{t+1}|X_t}(x;u)p_{X_t|Y_0^t}(u) \quad du}{N(Y_{t+1})}, \tag{6}$$

$$N(y) \quad := \quad \int_{-\infty}^{+\infty} p_{Y_{t+1}|X_{t+1}}(y;\xi)\int_{-\infty}^{+\infty} p_{X_{t+1}|X_t}(\xi;u)p_{X_t|Y_0^t}(u) \quad du \quad d\xi. \tag{7}$$

From the structure of the processes X_t and Y_t and from the assumptions on the noises V_t and W_t it follows immediately that $p_{Y_t|X_t}(y;x) = p_{\mathcal{N}(0,\exp(x+\gamma))}(y)$ and $p_{X_{t+1}|X_t}(x;u) = p_{\mathcal{N}(x,\sigma^2)}(\rho u)$. Bayes' formula reads now

$$p_{X_{t+1}|Y_0^{t+1}}(x) \quad = \quad \frac{p_{\mathcal{N}(0,\exp(x+\gamma))}(y)\int_{-\infty}^{+\infty} p_{\mathcal{N}(x,\sigma^2)}(\rho u)p_{X_t|Y_0^t}(u) \quad du}{N(y)}. \tag{8}$$

This is the exact solution of our filtering problem. However, this is very difficult to compute. Assume for example that we can deal with the numerical integration involved above. The problem is that in order to obtain the density at time $t+1$, given the density a time t, one has to update the given density point by point in the whole real line. In the next section we suggest a finite dimensional filter which approximates the exact filter found in this section.

2.4 A projection filter for the stochastic volatility model

Consider now the family $EM(c)$ of exponential densities defined in section (2.2). More specifically, we take the exponential manifold $EP(m) := \{p(\cdot,\theta) : \theta \in \Theta \subset \mathbb{R}^m\}$, with m an even positive integer and with a linear combination of the monomials $x, x^2, \ldots, x^m$ in the exponent:

$$p(x,\theta) = \exp\{\theta_1 x + \ldots + \theta_m x^m - \psi(\theta)\}, \quad \theta_m < 0. \tag{9}$$

In section (2.2) we showed that in order to approximate the density $p = p_{X_t|Y_0^t}$ with a density $p(\cdot,\theta)$ of $EM(c)$, it suffices to find the density in $EM(c)$ such that the c_i-expectations of p and $p(\cdot,\theta)$ match. With our specific manifold $EM(c) = EP(m)$, these expectations are exactly the first m moments of the exponential density. Then, in computing the projection filter, we update only the first m moments. Suppose we have computed the projection filter at time t via the expectation parameters $\eta_1(t), \ldots, \eta_m(t)$. Bayes' formula yields

$$\eta_j(t+1) = \frac{\int_{-\infty}^{+\infty} x^j p_{\mathcal{N}(0,\exp(x+\gamma))}(y)\int_{-\infty}^{+\infty} p_{\mathcal{N}(x,\sigma^2)}(\rho u)p(u;\eta(t)) \quad du \quad dx}{\int_{-\infty}^{+\infty} p_{\mathcal{N}(0,\exp(\xi+\gamma))}(y)\int_{-\infty}^{+\infty} p_{\mathcal{N}(\xi,\sigma^2)}(\rho u)p(u;\eta(t)) \quad du \quad d\xi}, \quad j = 1,..,m$$

which permits to update the expectation parameters. Then the new density $p(\cdot;\eta(t+1))$ may be computed recursively from the previous one $p(\cdot;\eta(t))$. If one prefers to avoid normalization at every step, one can use the scheme

$$\alpha_j(t+1) = \int_{-\infty}^{+\infty} x^j p_{\mathcal{N}(0,\exp(x+\gamma))}(y) \int_{-\infty}^{+\infty} p_{\mathcal{N}(x,\sigma^2)}(\rho u) q(u;\alpha(t)) \ du \ \ dx, \quad j=0,...,m, \tag{10}$$

$$\eta_i = \alpha_i/\alpha_0, \quad i=1,...,m, \tag{11}$$

where $q(\cdot;\alpha)$ is the unnormalized exponential density of the family $\{\exp(\theta_0+\theta_1 x+\ldots+\theta_m x^m);\theta_m<0\}$, characterized by the unnormalized expectation parameters α_i, $i=0,1,..,m$. Initially, at $t=0$, one can take $\alpha_0(0)=1, \alpha_i(0)=\eta_i(0)$ $i=1,\ldots,m$. By expanding this last expression one obtains

$$\begin{aligned}\alpha_j(t+1) = & \int_{-\infty}^{+\infty} \{x^j \exp[-\frac{x+\gamma}{2} - \frac{1}{2\sigma^2}x^2 - \frac{1}{2}y^2 e^{-x-\gamma}] \\ & \int_{-\infty}^{+\infty} \exp[-\frac{1}{2\sigma^2}(-2\rho x u+\rho^2 u^2)] q(u;\alpha(t)) \ du \ \}dx, \quad j=0,...,m.\end{aligned} \tag{12}$$

This last equation yields the evolution of the $m+1$ parameters α characterizing the projection filter for $EP(m)$. However, there are some problems in implementing this equation. Mainly, we need a way to express the exponential density $p(\cdot;\eta)$ explicitly from the knowledge of the η. Actually, from the theory of exponential families (see Brigo (1996) [6], Chapter 3 and references given therein) we know that the expectation parameters η characterize the densities of $EP(m)$, but we do not know a direct way to express the densities on the basis of such parameters. On the contrary, from (9) it is clear that the canonical parameters θ permit to express the densities of $EP(m)$ explicitly. In Brigo (1996) [6] (lemma 3.3.3) we give a recursive formula for $EP(m)$ which allows one to compute the last expectation parameter η_m and the higher order moments $\eta_{m+i}=E_{p(\cdot,\theta)}\{x^{m+i}\}$ for all nonnegative integers i, on the basis of the canonical parameters θ and of the first $m-1$ expectation parameters $\eta_1,...,\eta_{m-1}$. Define the matrix $M(\eta)$ as follows:

$$M_{i,j}(\eta) := \eta_{i+j}, \quad i,j=1,2,\ldots,m. \tag{13}$$

It is easy to verify that lemma (3.3.3) of Brigo (1996) [6] implies the following formula:

$$\begin{bmatrix} \theta_1 \\ 2\theta_2 \\ \vdots \\ m\theta_m \end{bmatrix} = -M(\eta)^{-1} \begin{bmatrix} 2\eta_1 \\ 3\eta_2 \\ \vdots \\ (m+1)\eta_m \end{bmatrix}. \tag{14}$$

From this last equation it follows that we can recover *algebraically* the canonical parameters θ from the knowledge of the moments $\eta_1,\ldots,\eta_{2m}$ up to order $2m$. Then we can compute the projection filter according to the following scheme:

(i) Given the initial density $p(x, \theta(0)) = p_{X_0}(x)$, set $t = 0$.

(ii) Assign $t := t + 1$.

(iii) Compute the first m moments of the new projection filter density at time t via the formula

$$\begin{aligned}
\alpha_j(t) &= \int_{-\infty}^{+\infty} \{x^j \exp[-\frac{x+\gamma}{2} - \frac{1}{2\sigma^2}x^2 - \frac{1}{2}y^2 e^{-x-\gamma}] \\
&\quad \int_{-\infty}^{+\infty} \exp[-\frac{1}{2\sigma^2}(-2\rho x u + \rho^2 u^2)] p(u; \theta(t-1))\ du\ \} dx, \quad j = 0, ..., m, \\
\eta_i(t) &= \alpha_i(t)/\alpha_0(t), \quad i = 1, ..., m.
\end{aligned}$$

(iv) Recover the canonical parameters $\theta(t)$ from the moments $\eta_1(t), \ldots, \eta_m(t)$ (What is the best way of doing this is still under investigation).

(v) Estimate the stochastic volatility by evaluating numerically the integral

$$E_{p(\theta(t))}\{\exp(\frac{x+\gamma}{2})\} = \int_{-\infty}^{+\infty} \exp(\frac{x+\gamma}{2})\ p(x, \theta(t))\ dx.$$

(vi) Start again from **(ii)**.

A possible problem in applying the above scheme is that for the integrals appearing in (iii) and (v) there are apparently no closed form expressions while the *numerical* integration is a subtle problem in this case. One of the difficulties in the numerical evaluation of the above integrals is that if the filter performs very well then the resulting density becomes very peaked, so that special numerical integration techniques are required. This problem is currently under investigation.

A possible heuristic answer to the problem under investigation in point (iv) is to replace points (iii) and (iv) by the following:

(iii.a) Compute the first **2m** moments of the new projection filter density at time t (j and i range now up to $2m$).

(iv.a) Recover the canonical parameters $\theta(t)$ from the moments $\eta_1(t), \ldots, \eta_{2m}(t)$ by using (14).

For a study of the behaviour of such a heuristic procedure, in a slightly different context, and for a comparison to several alternatives, including a Newton method, see Borwein and Huang (1995) [3]. Further investigations into this so called polynomial moment problem are called for. Better insight into the geometry of the manifolds $EP(m)$ is likely to be helpful, especially to understand the behaviour of the various algorithms at the boundary of the manifold where θ_m is close to zero.

Concerning the scheme as a whole, difficulties in numerical integration in the various steps are still present. A good performance of the above scheme is not guaranteed and it should be tested on simulations. We hope to return to this matter in future research work.

3 A state space approach to estimate CIR models of the term structure of interest rates

We consider one of the most popular models of the term-structure of interest rates: the multi-factor Cox-Ingersoll-Ross (CIR) model. In this model one assumes the instantaneous spot interest-rate r to be the sum of K factors X which follow a square-root process under the objective probability measure $\mathbf{P}$:

$$r_t = X_t^1 + ... + X_t^K, \quad dX_t^j = k_j(\theta_j - X_t^j)dt + \sigma_j \sqrt{X_t^j} dW_t^j \ , \quad j = 1, \ldots, K \ . \tag{15}$$

Let $\{\mathcal{F}_t, \ \ t \geq 0\}$ be the filtration representing the information available through time. With some reasonable requirements on the parameters k, θ and σ, this model yields an almost surely positive spot-rate r_t for all $t \geq 0$. This is generally considered as one of the main advantages of the CIR model. The term structure is expressed by specifying the *price* $P_t(T)$ at any time t for a *bond* which pays 1 at the maturity time $t + T$. In order to be able to price such bonds and specify the term structure of interest rates, one needs to specify the attitude towards risk. This is done by specifying the so-called equivalent martingale measure $\mathbf{Q}$ or risk neutral measure. For simplicity, this measure is taken of a form such that under $\mathbf{Q}$ the factors X still follow a square root process of the CIR type:

$$\frac{d\mathbf{Q}}{d\mathbf{P}}|\mathcal{F}_t = \exp\left\{ -\sum_{j=1}^{K} \left[\frac{\lambda_j^2}{2\sigma_j^2} \int_0^t X_s^j \ ds + \frac{\lambda_j}{\sigma_j} \int_0^t \sqrt{X_s^j} \ dW_s^j \right] \right\}$$

Under the risk-neutral measure $\mathbf{Q}$ the factors follow the equation

$$dX_t^j = [k_j \ \theta_j - (k_j + \lambda_j)X_t^j]dt + \sigma_j \sqrt{X_t^j} d\widetilde{W_t^j} \ , \quad j = 1, \ldots, K \ ,$$

where $\widetilde{W}$ is a standard Brownian motion under the risk-neutral measure $\mathbf{Q}$. The attitude towards risk can be tuned by the parameters $\lambda_1, ..., \lambda_K$, the so called *market prices of risk*. Set $\alpha = (\lambda, k, \theta, \sigma)$. *Yields* are given by

$$\begin{aligned}
y_t(T, \alpha) &:= \frac{-\log \ P_t(T)}{T} = -\frac{1}{T} \sum_{j=1}^{K} [\log \ \phi(\alpha_j, T) - \psi(\alpha_j, T) \ X_t^j], \\
\phi(\alpha_j, T) &= \left[\frac{2\sqrt{h} \ \exp\{(k_j + \lambda_j + \sqrt{h})T/2\}}{2\sqrt{h} + (k_j + \lambda_j + \sqrt{h})(\exp\{T\sqrt{h}\} - 1)} \right]^{2k_j\theta_j/\sigma_j^2} , \\
\psi(\alpha_j, T) &= \frac{2(\exp\{T\sqrt{h}\} - 1)}{2\sqrt{h} + (k_j + \lambda_j + \sqrt{h})(\exp\{T\sqrt{h}\} - 1)} , \\
h &= (k_j + \lambda_j)^2 + 2\sigma_j^2 \ .
\end{aligned}$$

which are affine functions of the factors X. This is a second advantage of the CIR model: it yields an affine term-structure.

Once this type of model has been established, one is confronted with the task of estimating the model parameters $\alpha = (\lambda, k, \theta, \sigma)$ on the basis of the available information. This problem is usually treated in two ways, as explained in Geyer and Pichler (1996) [15].

1) The cross section approach: One fits the quantities $y_t(T, \alpha)$ given above to observed yields in different periods of time, finding in each period the parameter values for which the model yields $y_t(T, \alpha)$ are closest to the actually observed yields in that period. The main objections to this approach are that the parameter estimates in general are not the same in different periods of time, and the fact that even if they were the same, the real dynamics of the spot rate r need not follow the CIR structure.

2) The time series approach: One fits the SDE's for the X's (usually for only one factor) to observable proxies of X_i (e.g. prices of T-bills or money-market rates). This approach raises the following objection: fitting to different proxies usually produces different estimates for the same parameters, so as to be inconsistent with the no-arbitrage conditions. Moreover, this approach does not use available information coming from observed yields.

The following state space approach answers the above objections by using both the *CIR dynamics* and the observed *yields' cross section* without the above inconsistencies.

The idea can be described as follows: assume that the observed yields differ from the yields $y_t(T, \alpha)$ prescribed by the model by a white noise process whose variance δ^2 is a new parameter to be estimated. This noise process can be viewed as a tool for taking into account market imperfections and deviations from the true model. Among the possible advantages of the state-space approach (over the pure cross-section approach and the time-series approach) stated by Geyer and Pichler (1996) [15] we recall the following:

- There is no need to rely on proxies for the factors X, contrary to the time-series approach;

- It is possible to estimate the parameters themselves rather than non-invertible functions of them;

- It is possible to estimate the factors X themselves, not only the parameters of the model;

- Measurement errors are taken into account explicitly.

Let us formalize the observation process as follows: τ_t is the vector of the n_t maturities at time t, ϵ is a discrete-time white noise process, and Y is the process of observed yields, where the capital letter is used to distinguish between actually observed yields Y and the yields y of the CIR model.

$$\tau_t := [T_t^1, ..., T_t^{n_t}]^T \ , \ \chi_i(\alpha, \tau_t) = -\frac{1}{T_t^i} \sum_{j=1}^{K} \log \phi(\alpha_j, T_t^i), \ \ \Psi_{i,j}(\alpha, \tau_t) = \frac{\psi(\alpha_j, T_t^i)}{T_t^i}$$

$$Y_t^i := y_t(T_t^i) + \delta_i \epsilon_t^i = \chi_i(\alpha, \tau_t) + \Psi_{i,\cdot}(\alpha, \tau_t) X_t + \delta_i \ \epsilon_t, \quad i = 1..n_t \ . \tag{16}$$

In vector form the observation process reads $Y_t = \chi(\alpha, \tau_t) + \Psi(\alpha, \tau_t) X_t + \mathrm{Diag}(\delta_1, .., \delta_{n_t}) \ \epsilon_t$, where the dimension n_t of the vector varies over time with the number of maturities.

Now there are essentially two main possibilities for introducing filtering theory in this setup.

3.1 Completely Bayesian approach

The first approach is completely Bayesian, and is used in system identification. It consists of viewing the parameters as new state variables in order to reduce the problem to a nonlinear filtering problem. Set

$$(X_t^{K+j}, X_t^{2K+j}, X_t^{3K+j}, X_t^{4K+j}, X_t^{5K+i}) := (k_j, \theta_j, \sigma_j, \lambda_j, \delta_i), \quad j = 1, ..., k, \quad i = 1, ..., n_t \ .$$

In such a way, the equations of the system (15,16), including the new state variables are:

$$dX_t^{K+r} = 0, \qquad r = 1, \ldots, 4K + n_t,$$

$$dX_t^j = X_t^{K+j}(X_t^{2K+j} - X_t^j)dt + X_t^{3K+j}\sqrt{X_t^j}\, dW_t^j \ , \qquad j = 1, \ldots, K,$$

$$\begin{aligned} Y_m^1 = &-\frac{1}{T_m^1}\sum_{j=1}^{K}[\log \ \phi(X_m^{4K+j}, X_m^{K+j}, X_m^{3K+j}, X_m^{2K+j}, T_m^1) \\ &-\psi(X_m^{4K+j}, X_m^{K+j}, X_m^{3K+j}, X_m^{2K+j}, T_m^1)X_m^j] + X_m^{5K+1} \ \epsilon_m^1 \\ &\vdots \\ Y_m^{n_m} = &-\frac{1}{T_m^{n_m}}\sum_{j=1}^{K}[\log \ \phi(X_m^{4K+j}, X_m^{K+j}, X_m^{3K+j}, X_m^{2K+j}, T_m^{n_m}) \\ &-\psi(X_m^{4K+j}, X_m^{K+j}, X_m^{3K+j}, X_m^{2K+j}, T_m^{n_m})X_m^j] + X_m^{5K+n_m} \ \epsilon_m^{n_m} \end{aligned}$$

This is a filtering problem with continuous time state X and discrete time observations Y, as described for example in Jazwinski (1970) [19]. Indeed, the unobserved signal is X, and the observation process Y consists of a deterministic functional of X plus some noise $X \ \epsilon$. Notice that the noise is state dependent, since components of the state X appear in front of the white noise process ϵ. The above filtering problem is nonlinear, and as such is infinite dimensional. An approximation of its solution can be considered. For example, one can use the extended Kalman filter (see again Jazwinski (1970) [19]) even though no general analytical result on the quality of the filter estimates is available. Justifications of the use of this filter are usually based on heuristics.

3.2 Quasi Maximum Likelihood

This method is based on an approximate computation of the likelihood function. Consider equations (15) for the factors of the CIR model. One of the advantages of square root processes like X is that they yield closed formulas for the mean and the variance of the factors themselves. This is somewhat helpful in establishing approximations, although nonlinearities in (15) imply that mean and variance are not sufficient to characterize the probabilistic behaviour of the factors X, contrary to the linear case. Indeed, the factor X^j features a non-central χ^2 transition density. Define $\widehat{X}^j_{t|s} = E\{X^j_t | Y_1, \ldots, Y_s\}$ and $V^{j,j}_{t|s} = E\{(X^j_t - \widehat{X}^j_{t|s})^2 | Y_1, \ldots, Y_s\}$ for $j = 1, \ldots, K$ and for any $0 \leq s \leq t$. From the above considerations it follows easily that between two observations, for $m \leq t < m+1$, the *prediction* step is given by

$$\begin{aligned} \widehat{X}^j_{m+1|m} &= \theta_j[1 - \exp(-k_j)] + \exp(-k_j)\widehat{X}^j_{m|m}, \\ V^{j,j}_{m+1|m} &= \sigma_j^2 \frac{1 - \exp(-k_j)}{k_j}[\theta_j \frac{(1 - \exp(-k_j))}{2} + \exp(-k_j)\widehat{X}^j_{m|m}] + \exp(-2k_j)V^{j,j}_{m|m}. \end{aligned} \tag{17}$$

Notice that even if at a certain time the conditional density $p^j_{m|m}$ of X^j_m given $Y_1, \ldots, Y_m$ were Gaussian, i.e.

$$p^j_{m|m} \sim \mathcal{N}(\widehat{X}^j_{m|m}, V_{m|m}),$$

the prediction step would lead us out of the Gaussian family:

$$p^j_{m+1|m} \not\sim \mathcal{N}(\widehat{X}^j_{m+1|m}, V^{j,j}_{m+1|m}) \ .$$

Therefore, $p^j_{m+1|m}$ is not Gaussian and its mean $\widehat{X}^j_{m+1|m}$ and variance $V^{j,j}_{m+1|m}$ are not enough to activate the *correction step* (Bayes' formula) leading to the conditional density $p^j_{m+1|m+1}$. In order to avoid such difficulties, one can *replace* the real $p^j_{m+1|m}$ by $\mathcal{N}(\widehat{X}^j_{m+1|m}, V^{j,j}_{m+1|m})$, i.e. replace the density $p^j_{m+1|m}$ by a Gaussian density sharing its first two moments. This is actually what is done in Geyer and Pichler [15]. As we remarked earlier in Section 2.2, this amounts to replacing $p^j_{m+1|m}$ by its best approximation, in the Kullback–Leibler sense, of the Gaussian family. Therefore the approximate filter used here can be interpreted as a Gaussian projection filter! By this approximation, it follows that the approximated *correction* at $t = m+1$, when Y_{m+1} is available, is given by Bayes' formula and can be summarized by

$$\begin{aligned} &\Delta_m := \mathrm{Diag}(\delta_1, .., \delta_{n_m}), \\ &\widehat{X}_{m+1|m+1} = \{\widehat{X}_{m+1|m} + V_{m+1|m}\Psi^T(\alpha, \tau_{m+1}) \left[\Psi(\alpha, \tau_{m+1})V_{m+1|m}\Psi(\alpha, \tau_{m+1})^T + \Delta^2_{m+1}\right]^{-1} \\ &\qquad\qquad \left(Y_{m+1} - \chi(\alpha, \tau_{m+1}) - \Psi(\alpha, \tau_{m+1})\widehat{X}_{m+1|m}\right)\}^+ \ , \\ &V_{m+1|m+1} = \left[\Psi^T(\alpha, \tau_{m+1})\Delta^{-2}_{m+1}\Psi(\alpha, \tau_{m+1}) + V^{-1}_{m+1|m}\right]^{-1} \ . \end{aligned} \tag{18}$$

The symbol $\{\,\cdot\,\}^+$ in the above equation denotes the positive part. It is applied in order to make sure that the approximate conditional mean $\widehat{X}$ be positive. We can now calculate the quasi-likelihood function as follows: Set $\beta = (\alpha, \delta)$ and compute

$$p_{Y_1,\dots,Y_n}(y_1, \dots, y_n; \beta) = p_{Y_n - \widehat{Y}_{n|n-1}}(y_n - \widehat{Y}_{n|n-1}; \beta) p_{Y_{n-1} - \widehat{Y}_{n-1|n-2}}(y_{n-1} - \widehat{Y}_{n-1|n-2}; \beta) \cdots$$

$$\cdots p_{Y_1 - \widehat{Y}_1}(y_1 - \widehat{Y}_1; \beta)$$

$$= p_{\mathcal{N}(0, \Psi(\alpha,\tau_n) V_{n|n-1} \Psi(\alpha,\tau_n)^T + \Delta_n^2)}(y_n - \widehat{Y}_{n|n-1})$$

$$p_{\mathcal{N}(0, \Psi(\alpha,\tau_{n-1}) V_{n-1|n-2} \Psi(\alpha,\tau_{n-1})^T + \Delta_{n-1}^2)}(y_{n-1} - \widehat{Y}_{n-1|n-2}) \cdots p_{\mathcal{N}(0, \Psi(\alpha,\tau_1) V_1 \Psi(\alpha,\tau_1)^T + \Delta_1^2)}(y_1 - \widehat{Y}_1).$$

This function can be computed (and maximized) once we know $\widehat{Y}$ and V for all β. These quantities can be obtained for every possible value of β from the above recursion (17, 18). Of course, in practice numerical simulation techniques are required to maximize the quasi-likelihood.

The two unanswered questions about this approach are:

- How good is the Kullback-Leibler projection on the Gaussian family used after the prediction step?
- How good is taking $\{\quad\}^+$ in the correction?

In order to deal appropriately with the first of these questions one can make use of the concept of *projection residual* that was developed for the continuous time case in Brigo, Hanzon and Le Gland (1995) [7]. This concept can actually be used here, because the approximate filter used in [15] has in fact the interpretation of a continuous time Gaussian Projection Filter for a continuous time signal observed in discrete time. Of course the question about taking $\{\quad\}^+$ arises because here one works with Gaussian densities. In order to avoid this problem one could try to work with a class of densities which have their support on the non-negative real halfline and work out the Projection Filter, for the model under investigation here, by using such a class of densities.

4 Risk–minimizing hedging strategies under partial observation

We shortly report the result of Fischer, Platen, and Runggaldier (1996) [13]. This is a significant case where filtering theory fits nicely a mathematical-finance setup. A financial market is considered over a time interval $[0, T]$ with a risky asset, whose price is denoted by S, and a bond, whose price is assumed identically equal to one. Under a martingale measure, we write

$$B_t = 1,$$

$$dS_t = \sigma_t(Z_t)S_t dW_t$$
$$dy_t = A_t S_t dt + D_t dV_t.$$

Let $\mathcal{F}_t = \sigma\{S_u, Z_u : \quad u \leq t\}$ be the information represented by observation of S and Z up to time t. The process Z is a hidden Markov process (representing the state of the economy) with transition intensity matrix Λ. Let N_t be the number of jumps of Z (number of changes in the economy) up to time t. The process y_t represent observation of S_t in additive noise, reflecting the possibility that not all indicated prices are actually traded. Our observation process is denoted by $Y_t := [y_t, \quad N_t]$. Denote by $\mathcal{Y}_t := \sigma\{Y_s, \ s \leq t\}$ the information represented by observation of S and Z up to time t. We assume that S_T is fully observed. We consider a contingent claim $H = H(S_T)$ to be priced at all $t < T$. We will consider two cases: full observations $\{\mathcal{F}_t \ : t \geq 0\}$ available, and partial observations $\{\mathcal{Y}_t \ : t \geq 0\}$ available. In both cases we are dealing with an incomplete market, since there are more sources of randomness than traded risky assets. Then perfect hedging with self-financing portfolios is not possible in general. We can still try to determine a mean self financing hedging strategy that minimizes a risk criterion related to the lack of self-financing.

We begin by the case with full observations. The main ingredient is the Kunita - Watanabe decomposition. We are looking for a strategy (ξ_t, η_t) (ξ_t amount of stock, η_t amount of bond) such that

i) ξ_t is $\mathcal{F}_t$ predictable, η_t is $\mathcal{F}_t$ adapted, and

$$E\{\int_0^T |\xi_t|^2 \sigma_t(Z_t)^2 S_t^2 dt\} < \infty \ .$$

ii) $\xi_T \ S_T + \eta_T \ 1 = H$ (final value of the strategy equals the claim)

iii) $\xi_t \ S_t + \eta_t \ 1 - \int_0^t \xi_u \ dS_u =: C_t(\xi, \eta)$ (value - gains = constant) is a martingale (mean-constant);

iv) minimizes $E\{(C_T - C_t)^2 | \mathcal{F}_t\}$ for each t (quadratic criterion) among all other strategies as in (i), (ii), (iii).

The solution of this problem was derived by Föllmer and Schweizer (1991) [14]. They proved, among other results, that if $H \in L_2(\mathcal{F}_T, Q)$ (Q is a martingale measure for S), then

$$\xi_t^* = \xi_t^H, \quad \eta_t^* = E\{H|\mathcal{F}_t\} - \xi_t^* S_t, \text{ where}$$

$$E\{H|\mathcal{F}_t\} = EH + \int_0^t \xi_u^H dS_u + L_t^H$$

is the Kunita-Watanabe decomposition (L is a martingale, orthogonal to S).

In the case of partial observations, points (i), (iii) and (iv) are replaced respectively by

i) ξ_t is $\mathcal{Y}_t$ predictable, η_t is $\mathcal{Y}_t$ adapted, and

$$E\{\int_0^T |\xi_t|^2 \sigma_t(Z_t)^2 S_t^2 dt | \mathcal{Y}_0\} < \infty \ .$$

iii) $E\{\xi_t \ S_t + \eta_t \ 1 - \int_0^t \xi_u \ dS_u | \mathcal{Y}_t\} = E\{C_t(\xi, \eta) | \mathcal{Y}_t\}$ is a $(\mathcal{Y}, Q)-$martingale;

iv) minimizes $E\{(C_T - C_t)^2 | \mathcal{Y}_t\}$ among all other strategies as in (i), (ii), (iii).

The solution of this second problem was given by Schweizer (1994) [25], see also Di Masi, Platen and Runggaldier (1995) [12] .

$$E\{H|\mathcal{F}_t\} = EH + \int_0^t \xi_u^H dS_u + L_t^H,$$

$$\xi_t^{\mathcal{Y}} = \frac{E\{\xi_t^H \ \sigma_t^2(Z_t) \ S_t^2 | \mathcal{Y}_t\}}{E\{\sigma_t^2(Z_t) \ S_t^2 | \mathcal{Y}_t\}}, \quad \eta_t^{\mathcal{Y}} = E\{H|\mathcal{Y}_t\} - \xi_t^{\mathcal{Y}} S_t.$$

How can one compute ξ^H and $E\{H|\mathcal{Y}_t\}$ explicitly ? The solution of this problem was given by Di Masi, Kabanov and Runggaldier (1994) [11]. If H has polynomial growth, then

$$\xi_t^H = \xi_t^H(S_t, Z_t) = \frac{\partial}{\partial S} u_t(S_t, Z_t), \quad E\{H|\mathcal{Y}_t\} = E\{u_t(S_t, Z_t)|\mathcal{Y}_t\},$$

where $u_t(x, i) = E\{H|S_t = x, Z_t = i\}$ solves

$$\partial_t u_t(x, i) + \frac{1}{2}\sigma^2(i) x^2 \frac{\partial^2}{\partial x^2} u_t(x, i) + \sum_j \Lambda_{ij} u_t(x, j) = 0, \quad u_T(x, i) = H(x)$$

The $\mathcal{Y}-$mean self-financing strategy can be computed via the conditional distribution of the unobserved state (S_t, Z_t) given the observations $\mathcal{Y}_t$. This is the filtering problem treated by Miller and Runggaldier (1996) [24].

5 Acknowledgements

This paper was originally presented at the *International Workshop on the Interplay between Insurance, Finance and Control*, held at the Mathematical Center of the University of Aarhus on February 25 – March 1, 1997. The first named author wishes to thank Aleardo Adotti, head of the Product Development Group of IMI Bank and Renzo G. Avesani, head of Risk Management and Research at Cariplo Bank, for encouraging the prosecution of his research activities even in their most mathematical aspects.

References

[1] Amari, S-I.(1985). *Differential Geometric Methods in Statistics.* Lecture Notes in Statistics, 28. Springer-Verlag, Berlin.

[2] Andersen, T.G. (1994). Stochastic autoregressive volatility: a framework for volatility modelling. *Mathematical Finance* 4, 75-102.

[3] Borwein, J.M., and Huang, W.Z. (1995). A fast heuristic method for polynomial moment problems with Boltzmann-Shannon entropy. *SIAM J. Optimization* 5, 68-99.

[4] Brigo, D. (1995). On the nice behaviour of the Gaussian projection filter with small observation noise. *Systems & Control Letters* 26, 363–370.

[5] Brigo, D. (1996). New results on the Gaussian projection filter with small observation noise. *Systems & Control Letters* 28, 273–279

[6] Brigo, D. (1996). *Filtering by Projection on the Manifold of Exponential Densities*, PhD Thesis, Free University Amsterdam, 1996.

[7] Brigo, D. , Hanzon, B. , and Le Gland, F. (1995). A differential geometric approach to nonlinear filtering: the projection filter. To appear in IEEE Transactions on Automatic Control (extended version available on the internet at URL: ftp://ftp.irisa.fr/techreports/1995/PI-914.ps.Z).

[8] Brigo, D. , Hanzon, B. , and Le Gland, F. (1997). Approximate filtering by projection on the manifold of exponential densities. Accepted for publication in *Bernoulli.*

[9] Chaleyat-Maurel , M. , and Michel, D. (1984). Des resultats de non-existence de filtre de dimension finie. *Stochastics* 13, 83–102.

[10] Davis, M.H.A. , and Marcus, S.I. (1981). An introduction to nonlinear filtering. In: *Stochastic systems: The mathematics of filtering and identification and applications*, 53–75. Reidel, Dordrecht.

[11] Di Masi, G.B., Kabanov, Yu. M., and Runggaldier, W.J. (1994). Mean-variance hedging of options on stocks with Markov volatilities. *Theory of probability and its applications* 39, 211–222.

[12] Di Masi, G.B., Platen, E., and Runggaldier, W.J. (1995). Hedging of options under discrete observations on assets with stochastic volatility. In: *Seminar on stochastic analysis, random fields and applications, progress in probability* 36, 359–364. Birkhäuser.

[13] Fischer, P. , Platen, E. , and Runggaldier, W.J. (1996). Risk–minimizing hedging strategies under partial observation, submitted to the *Proceedings of the 1996 Ascona Seminar on Stochastic Analysis, Random Fields and Applications.*

[14] Föllmer, H. , and Schweizer, M. (1991). Hedging of contingent claims under incomplete information. In: *Applied stochastic analysis, stochastic monographs* 5, 389–414. Gordon and Breach, London.

[15] Geyer, A.L.J, and Pichler, S. (1996). A state–space approach to estimate and test multi–factor Cox–Ingersoll–Ross models of the term structure. *Preliminary draft*

[16] Hanzon, B. (1987). A differential-geometric approach to approximate nonlinear filtering. In: C.T.J. Dodson (ed), Geometrization of Statistical Theory, pp. 219-224, ULDM Publications, University of Lancaster.

[17] Hanzon, B., and Hut, R. (1991). New results on the Projection Filter. *Proceedings of the European Control Conference*. Grenoble, Vol. I, 623-628.

[18] Jacquier, E. , Polson, N. G. , and Rossi, P.E. (1994). Bayesian analysis of stochastic volatility models. *Journal of Economics and Business Statistics* 12, 371–417.

[19] Jazwinski, A.H. (1970). *Stochastic Processes and Filtering Theory*. Academic Press, New York.

[20] Kagan, A.M. , Linnik, Y.V., and Rao, C.R. (1973). *Characterization problems in Mathematical Statistics*. John Wiley and Sons, New York.

[21] Liptser, R.S. , and Shiryayev, A.N. (1978). *Statistics of Random Processes I, General Theory*. Springer Verlag, Berlin.

[22] Mahieu, R., and Schotman, P. (1997). An empirical application of stochastic volatility models. *Working paper*.

[23] Melino, A., and Turnbull, S.M. (1990). Pricing foreign currency options with stochastic volatility. *Journal of Econometrics* 45, 239-265.

[24] Miller, B.M. , and Runggaldier, W.J. (1996). Kalman filtering for linear systems with coefficients driven by a hidden Markov jump process. *Submitted for publication.*

[25] Schweizer, M. (1994). Risk minimizing hedging strategies under restricted information. *Mathematical finance* 4, 327–342.