Filtering

Stochastic filtering concerns the estimation of the state of a randomly evolving system that is only indirectly observed and the observations are, furthermore, affected by noise. The primary examples in finance concern factor models, where some factors are not directly observable. We review stochastic filtering in discrete and continuous time in linear and nonlinear models and describe some applications to pricing and portfolio management.

Key words: Incomplete information, stochastic filtering, Bayesian estimation, analytic filter solutions, particle filters, factor models, pricing, portfolio management.

The filtering problem

Consider a randomly evolving system, the state of which is denoted by x_t and this state may not be directly observable. Denote by y_t the observation at time $t \in [0,T]$ (x_t and y_t may be vector-valued): y_t is supposed to be probabilistically related with x_t. For instance, y_t may represent a noisy measurement of x_t.

The process x_t is generally supposed to evolve in a Markovian way according to a given (a priori) distribution $p(x_t \mid x_s)$, $s \leq t$. The dynamics of y_t are given in terms of the process x_t; a general assumption is that, given x_t, the process y_t is independent of its past and so one may consider as given the distribution $p(y_t \mid x_t)$. The information on x_t at a given $t \in [0,T]$ is thus represented by the past and present observations of y_t, i.e. by $y_0^t := \{y_s;\, s \leq t\}$ or, equivalently, by the filtration $\mathcal{F}_t^y := \sigma\{y_s;\, s \leq t\}$. This information, combined with the a-priori dynamics of x given by $p(x_t \mid x_s)$ can, via a Bayes'-type formula, be synthesized in the conditional or posterior distribution $p(x_t \mid y_0^t)$ of x_t, given y_0^t, and this distribution is called the *filter distribution*.

The filtering problem consists now in determining, possibly in a recursive way, the filter distribution at each $t \leq T$. It can also be seen as a dynamic extension of Bayes' statistics: for $x_t \equiv x$ an unknown parameter, the dynamic model for x given by $p(x_t \mid x_s)$ reduces to a prior distribution for x and the filter $p(x \mid y_0^t)$ is then simply the posterior distribution of x, given the observations y_s, $s \leq t$.

In many applications it suffices to determine a synthetic value of the filter distribution $p(x_t \mid y_0^t)$. In particular, given an (integrable) function $f(\cdot)$, one may want to compute

$$E\{f(x_t) \mid y_0^t\} = E\{f(x_t) \mid \mathcal{F}_t^y\} = \int f(x)\, dp(x \mid y_0^t) \qquad (1)$$

The quantity in (1) may be seen as the best estimate of $f(x_t)$, given y_0^t, with respect to the mean square error criterion in the sense that $E\{(E\{f(x_t \mid y_0^t)\} - f(x_t))^2\} \leq E\{(g(y_0^t) - f(x_t))^2\}$ for all measurable (and integrable) functions $g(y_0^t)$ of the available information. In this sense one may also consider $E\{f(x_t) \mid \mathcal{F}_t^y\}$ as the *optimal filter* for $f(x_t)$. Notice that, determining $E\{f(x_t) \mid \mathcal{F}_t^y\}$ is no more restrictive than determining the entire filter distribution $p(x_t \mid y_0^t)$; in fact, by taking $f(x) = e^{i\lambda x}$ for a generic λ, the $E\{f(x_t) \mid \mathcal{F}_t^y\}$ in (1) leads to the conditional characteristic function of x_t given y_0^t.

Related to the filtering problem are the *prediction problem*, namely that of determining $p(x_t \mid y_0^s)$ for $s < t$, and the *interpolation or smoothing problem* concerning $p(x_t \mid y_0^s)$ for $t < s$. Given the Bayesian nature of the filtering problem, one can also consider the so-called *combined filtering and parameter estimation problem*: if the dynamics $p(x_t \mid x_s)$ for x include an unknown parameter θ, one may consider the problem of determining the joint conditional distribution $p(x_t, \theta \mid \mathcal{F}_t^y\}$.

Models for the filtering problem

To solve a given filtering problem one has to specify the two basic inputs, namely $p(x_t \mid x_s)$ and $p(y_t \mid x_t)$. A classical model in discrete time is

$$\begin{cases} x_{t+1} &= a(t, x_t) + b(t, x_t)\, w_t \\ y_t &= c(t, x_t) + v_t \end{cases} \qquad (2)$$

where w_t and v_t are (independent) sequences of independent random variables and the distribution of x_0 is given. Notice that in (2) the process x_t is Markov and y_t represent indirect observations of x_t, affected by additive noise.

The continuous time counterpart is

$$\begin{cases} dx_t &= a(t, x_t)dt + b(t, x_t)\, dw_t \\ dy_t &= c(t, x_t)dt + dv_t \end{cases} \qquad (3)$$

and notice that, here, y_t represents the cumulative observations up to t. These basic models allow for various extensions: x_t may e.g. be a jump-diffusion process or a Markov process with a finite number of states, characterized by its transition intensities. Also the observations may more generally be a jump-diffusion such as

$$dy_t = c(t, x_t)dt + dv_t + dN_t \qquad (4)$$

where N_t is a doubly stochastic Poisson process, the intensity $\lambda_t = \lambda(x_t)$ of which depends on x_t. Further generalizations are of course possible.

Analytic solutions of the filtering problem

Discrete time

By the Markov property of the process x_t and the fact that, given x_t, the process y_t is independent of its past, with the use of Bayes' formula one obtains easily the following two-step recursions

$$\begin{cases} p(x_t \mid y_0^{t-1}) & = & \int p(x_t \mid x_{t-1})\, dp(x_{t-1} \mid y_0^{t-1}) \\ p(x_t \mid y_0^t) & \propto & p(y_t \mid x_t) p(x_t \mid y_0^{t-1}) \end{cases} \tag{5}$$

where $\propto$ denotes "proportional to" and the first step corresponds to the *prediction step* while the second one is the *updating step*. The recursions start with $p(x_0 \mid y_0^0) = p(x_0)$. Although (5) represents a fully recursive relation, its actual computation is made difficult not only by the presence of the integral in x_{t-1}, but also by the fact that this integral is parametrized by x_t that in general takes infinitely many values. Depending on the model one can however obtain explicit solutions as will be shown below. The most general such situation arises when one can find a finitely parametrized class of distributions of x_t that is closed under the operator implicit in (5), namely such that, whenever $p(x_{t-1} \mid y_0^{t-1})$ belongs to this class, then also $p(x_t \mid y_0^t)$ belongs to it. A classical case is the linear conditionally Gaussian case that corresponds to a model of the form

$$\begin{cases} x_{t+1} & = & A_t(y_0^t) x_t + B_t(y_0^t)\, w_t \\ y_t & = & C_t(y_0^t)\, x_t + R_t(y_0^t)\, v_t \end{cases} \tag{6}$$

where the coefficients may depend on the entire past of the observations y_t and w_t, v_t are independent i.i.d. sequences of standard Gaussian random variables. For such a model $p(x_t \mid y_0^t)$ is Gaussian at each t and therefore characterized by mean and (co)variance that can be recursively computed by the well-known *Kalman-Bucy filter*. Denoting

$$\begin{array}{c} \hat{x}_{t|t-1} := E\{x_t \mid y_0^{t-1}\} \;;\; \hat{x}_{t|t} := E\{x_t \mid y_0^t\} \\ P_{t|t-1} := E\{(x_t - \hat{x}_{t|t-1})(x_t - \hat{x}_{t|t-1})' \mid y_0^{t-1}\} \;; \\ P_{t|t} := E\{(x_t - \hat{x}_{t|t})(x_t - \hat{x}_{t|t})' \mid y_0^t\} \end{array} \tag{7}$$

the Kalman-Bucy filter is given by (dropping for simplicity the dependence on y_0^t)

$$\begin{cases} \hat{x}_{t|t-1} & = & A_{t-1}\, \hat{x}_{t-1|t-1} \\ P_{t|t-1} & = & A_{t-1} P_{t-1|t-1} A'_{t-1} + B_{t-1} B'_{t-1} \end{cases} \tag{8}$$

which represents the *prediction step*, and

$$\begin{array}{ccl} \hat{x}_{t|t} & = & \hat{x}_{t|t-1} + L_t[y_t - C_t \hat{x}_{t|t-1}] \\ P_{t|t} & = & P_{t|t-1} - L_t C_t P_{t|t-1} \end{array} \tag{9}$$

which represents the *updating step* with $\hat{x}_{0|-1}$ the mean of x_0 and $P_{0|-1}$ its variance. Furthermore,

$$L_t := P_{t|t-1} C'_t [C_t P_{t|t-1} C'_t + R_t R'_t]^{-1} \;. \tag{10}$$

Notice that in the prediction step the estimate of x_t is propagated one step further on the basis of the given a priori dynamics of x_t, while in the updating step one takes into account the additional information coming from the current observation. A crucial role in the updating step (9) is played by

$$\begin{array}{l} y_t - C_t \hat{x}_{t|t-1} = y_t - C_t A_{t-1} \hat{x}_{t-1|t-1} \\ \quad = y_t - C_t E\{x_t \mid y_0^{t-1}\} = y_t - E\{y_t \mid y_0^{t-1}\} \end{array} \tag{11}$$

which represents the new information given by y_t with respect to its best estimate $E\{y_t \mid y_0^{t-1}\}$ and is therefore called *innovation*.

The Kalman-Bucy filter has been extremely successful and has been applied also to Gaussian models that are nonlinear by simply linearizing the nonlinear coefficient functions around the current best estimate of x_t. In this way one obtains an approximate filter, called *extended Kalman filter*.

Exact solutions for the discrete time filtering problem can also be obtained for the case when x_t is a finite state Markov chain with, say, N states defined by its transition probability matrix. In this case the filter is characterized by its conditional state probability vector that we denote by $\pi_t = (\pi_t^1, \cdots, \pi_t^N)$ with $\pi_t^i := P\{x_t = i \mid \mathcal{F}_t^y\}$.

Continuous time

For the solution of a general continuous time problem we have two main approaches, namely the *innovations approach* that extends the innovation representation of the Kalman filter where, combining (8) and (9), this latter representation is given by

$$\hat{x}_{t|t} = A_{t-1}\, \hat{x}_{t-1|t-1} + L_t[y_t - C_t A_{t-1}\, \hat{x}_{t-1|t-1}],$$

and the so-called *reference probability approach*. For the sake of brevity we discuss here only the innovations approach *(Kushner-Stratonovich equation)* and we do it for the case of model (3) mentioning briefly possible extensions to other cases. For the reference probability approach (Zakai equation) we refer to the literature (for instance [8], [19]).

We denote by $\mathcal{L}$ the generator of the Markov diffusion x_t in (3) i.e., assuming $x \in \mathbb{R}^n$, for a function $\phi(t,x) \in \mathbb{C}^{1,2}$ we have

$$\mathcal{L}\phi(t,x) = a(t,x)\phi_x(t,x) + \frac{1}{2} \sum_{i,j=1}^{n} \sigma_{ij}(t,x) \phi_{x_i x_j}(t,x) \tag{12}$$

with $\sigma(t,x) := b(t,x)b'(t,x)$. Furthermore, for a generic (integrable) $f(\cdot)$, we let $\hat{f}_t := E\{f(x_t) \mid \mathcal{F}_t^y\}$. The innovations approach now leads in case of model (3) to the following dynamics, also called *Kushner-Stratonovich equation* (see e.g. [19],[8])

$$d\hat{f}_t = \widehat{\mathcal{L}f(x_t)}dt + [\widehat{c(t,x_t)f}(x_t) - \widehat{c(t,x_t)}\hat{f}_t]'[dy_t - \widehat{c(t,x_t)}dt] \tag{13}$$

which (see (3)) is based on the innovations $dy_t - \widehat{c(t,x_t)}dt = dy_t - E\{dy_t \mid \mathcal{F}_t^y\}$. In addition to the stochastic integral, the main difficulty with (13) is that, to compute $\hat{f}$, one needs $\widehat{cf}$ which in turn requires $\widehat{c^2 f}$ and so on. In other words, (13) is not a closed system of stochastic differential equations. Again, for particular models, (13) leads to a closed system as it happens with the linear-Gaussian version of (3) that leads to the continuous time Kalmann-Bucy filter, which is analogous to its discrete time counterpart. A further case arises when x_t is finite-state Markov with transition intensity matrix $Q = \{q_{ij}\},\ i,j = 1,\cdots,N$. Putting $\pi_t(i) := P\{x_t = i \mid \mathcal{F}_t^y\}$ and taking as $f(\cdot)$ the indicator function of the various values of x_t, (13) becomes (replacing $\mathcal{L}$ by Q)

$$\begin{array}{rcl} d\pi_t(j) & = & \sum_{i=1}^N \pi_t(i)q_{ij}dt + \pi_t(j)[c(t,j) - \sum_{i=1}^N \pi_t(i)c(t,i)] \\ & \cdot & [dy_t - \sum_{i=1}^N \pi_t(i)c(t,i)dt] \end{array} \tag{14}$$

For more results when x_t is finite-state Markov we refer to [10], in particular see [11].

We just mention that one can write the dynamics of $\hat{f}_t$ also in the case of jump-diffusion observations as in (4) (see [17]) and one can furthermore obtain an evolution equation, a stochastic PDE, for the conditional density $p(x_t) = p(x_t \mid y_0^t)$, whenever it exists, that involves the formal adjoint $\mathcal{L}^*$ of the $\mathcal{L}$ in (12) (see [19]).

Numerical solutions of the filtering problem

As we have seen, an explicit analytic solution to the filtering problem can be obtained only for special models so that, remaining within analytic solutions, in general one has to use an approximation approach. As already mentioned, one such approximation consists in linearizing the nonlinear model, both in discrete and continuous time, and this leads to the extended Kalman filter. Another approach consists in approximating the original model by one where x_t is finite-state Markov. The latter approach goes back mainly to H.Kushner and co-workers, see e.g. [18] (for a financial application see also [13]). A more direct numerical approach is simulation-based and given by the so-called *particle approach to filtering* that has been successfully introduced more recently and that we are going to summarize next.

Simulation-based solution (particle filters)

Being simulation-based, this solution method as such is applicable only to discrete time models; continuous time models have first to be discretized in time. There are various variants of particle filters but, analogously to the analytical approaches, they all proceed along two steps, a prediction step and an updating step and at each step the relevant distribution (predictive and filter distribution respectively) is approximated by a discrete probability measure supported by a finite number of points. These approaches vary mainly in the updating step.

A simple version of a particle filter is as follows (see [3]): in the generic period $t-1$ approximate $p(x_{t-1} \mid y_0^{t-1})$ by a discrete distribution $((x_{t-1}^1, p_{t-1}^1), \cdots, (x_{t-1}^L, p_{t-1}^L))$ where p_{t-1}^i is the probability that $x_{t-1} = x_{t-1}^i$. Consider each location x_{t-1}^i as the position of a "particle".

i) *Prediction step:* propagate over one time period each of the particles $x_{t-1}^i \to \hat{x}_t^i$ using the given (discrete time) evolution dynamics of x_t: referring to model (2) just simulate independent trajectories of x_t starting from the various x_{t-1}^i. This leads to an approximation of $p(x_t \mid y_0^{t-1})$ by the discrete distribution $((\hat{x}_t^1, \hat{p}_t^1), \cdots, (\hat{x}_t^L, \hat{p}_t^L))$ where one puts $\hat{p}_t^i = p_{t-1}^i$.

ii) *Updating step:* update the weights using the new observation y_t by putting $p_t^i = c p_{t-1}^i p(y_t \mid \hat{x}_t^i)$ where c is the normalization constant (see the second relation in (5) for an analogy).

Notice that $p(y_t \mid \hat{x}_t^i)$ may be viewed as the likelihood of particle $\hat{x}_t^i$, given the observation y_t, so that in the updating step one weighs each particle according to its likelihood. There exist various improvements of this basic setup. There are also variants, where in the updating step each particle is made to branch into a random number of offsprings, where the mean number of offsprings is taken to be proportional to the likelihood of that position. In this latter variant the number of particles increases and one can show that, under certain assumptions, the empirical distribution of the particles converges to the true filter distribution. There is a vast literature on particle filters, we just mention [5] and, in particular, [1].

Filtering in Finance

There are various situations in finance where filtering problems may arise, but one typical situation is given by factor models. These models have proven to be useful for capturing the complicated nonlinear dynamics of real asset prices, while at the same time being parsimonious and numerically tractable. In addition, with Markovian factor processes, Markov-process techniques can be fruitfully applied. In many financial applications of factor models the investors have only incomplete information about the actual state of the factors and this may induce model risk. In fact, even if the factors are associated with economic quantities, some of them are difficult to observe precisely. Furthermore, abstract factors without economic interpretation are often included in the specification of a model in order to increase its flexibility. Under incomplete information of the factors, their values have to be inferred from observable quantities and this is where filtering comes in as an appropriate tool.

Most financial problems concern pricing as well as portfolio management, in particular hedging and portfolio optimization. While portfolio management is performed under the physical measure, for pricing one has to use a martingale measure. Filtering problems in finance may therefore be considered under the physical or the martingale measures, or under both (see [22]). In what follows we shall discuss filtering for pricing problems, with examples from term structure and credit risk, as well as for portfolio management. More general aspect can be found e.g. in the recent papers [6], [7], [23].

Filtering in pricing problems

This section is to a large extent based on [14]. In Markovian factor models the price of an asset at a generic time t can, under full observation of the factors, be expressed as an instantaneous function $\Psi(t, x_t)$ of time and the value of the factors. Let $\mathcal{G}_t$ denote the full filtration that measures all the processes of interest and let $\mathcal{F}_t \subset \mathcal{G}_t$ be a subfiltration representing the information of an investor. What is an arbitrage-free price in the filtration $\mathcal{F}_t$? Assume the asset to be priced is a European derivative with maturity T and claim $H \in \mathcal{F}_T$. Let N be a numeraire, adapted to the investor filtration $\mathcal{F}_t$ and let Q^N be the corresponding martingale measure. One can easily prove the following

Lemma 0.1. *Let $\Psi(t, x_t) = N_t E^{Q^N}\left\{\frac{H}{N_T} \mid \mathcal{G}_t\right\}$ be the arbitrage-free price of the claim H under the full information $\mathcal{G}_t$ and $\hat{\Psi}(t) = N_t E^{Q^N}\left\{\frac{H}{N_T} \mid \mathcal{F}_t\right\}$ the corresponding arbitrage-free price in the investor filtration. It then follows that*

$$\hat{\Psi}(t) = E^{Q^N}\left\{\Psi(t, x_t) \mid \mathcal{F}_t\right\} \tag{15}$$

Furthermore, if the savings account $B_t = \exp\{\int_0^t r_s ds\}$ with corresponding martingale measure Q is $\mathcal{F}_t-$adapted, then

$$\hat{\Psi}(t) = E^{Q}\left\{\Psi(t, x_t) \mid \mathcal{F}_t\right\} \tag{16}$$

We thus see that, in order to compute the right hand sides in (15) or (16), namely the price of a derivative under restricted information given its price under full information, one has to solve the filtering problem for x_t given $\mathcal{F}_t$ under a martingale measure. We present now two examples.

Example 0.2. *(Term structure of interests)* The example is a simplified version adapted from [15]. Consider a factor model for the term structure where the unobserved (multivariate) factor process x_t satisfies the linear-Gaussian model

$$dx_t = Fx_t dt + D dw_t \tag{17}$$

In this case the term structure is exponentially affine in x_t and one has

$$p(t, T; x_t) = \exp[A(t, T) - B(t, T)\, x_t] \tag{18}$$

with $A(t, T), B(t, T)$ satisfying well-known first order ordinary differential equations in order to exclude arbitrage. Passing to log-prices for the bonds, one gets the linear relationship $y_t^T := \log p(t, T; x_t) = A(t, T) - B(t, T)x_t$. Assume now that investors cannot observe x_t, but they can observe the short rate and the log-prices of a finite number n of zero-coupon bonds, perturbed by additive noise. This leads to a system of the form

$$\left\{\begin{array}{lcl} dx_t &=& Fx_t dt + D\, dw_t \\ dr_t &=& \left(\alpha_t^0 + \boldsymbol{\beta}_t^0 x_t\right) dt + \sigma_t^0 dw_t + dv_t^0 \\ dy_t^i &=& \left(\alpha_t^i + \boldsymbol{\beta}_t^i x_t\right) dt + \sigma_t^i dw_t + (T_i - t) dv_t^i \,; \quad i = 1, \cdots, n \end{array}\right. \tag{19}$$

where v_t^i, $i = 0, \cdots, n$ are independent Wiener processes and the coefficients are related to those in (17) and (18). The time-dependent volatility in the perturbations of the log-prices reflects the fact that it tends to zero as time approaches maturity.

From the filtering point of view the system (19) is a linear-Gaussian model with x_t unobserved and the observations given by (r_t, y_t^i). We shall thus put $\mathcal{F}_t = \sigma\{r_s, y_s^i;\, s \le t,\, i = 1, \cdots, n\}$. The filter distribution is Gaussian and, via the Kalman filter, one can obtain its conditional mean m_t and (co)variance Σ_t. Applying Lemma 0.1 and using the moment-generating function of a Gaussian r.v., we obtain as arbitrage-free price, in the investor filtration, of an illiquid bond with maturity T the following

$$\begin{array}{rcl} \hat{p}(t, T) &=& E\{p(t, T; x_t) \mid \mathcal{F}_t\} \\ &=& \exp[A(t, T)]\, E\{\exp[-B(t, T)x_t] \mid \mathcal{F}_t\} \\ &=& \exp[A(t, T) - B(t, T)m_t + \frac{1}{2}B(t, T)\Sigma_t B'(t, T)] \end{array} \tag{20}$$

For the given setup the expectation is under the martingale measure Q with the money market account B_t as numeraire. To apply Lemma 0.1 we need the numeraire to be observable and this contrasts with the assumption that r_t is observable only in noise. This difficulty can be overcome (see [14]) but, by suitably changing the drifts in (19) (corresponding to a translation of w_t), one may however consider the model (19) also under a martingale measure for which the numeraire is different from B_t and observable.

A further filter application to the term structure of interest rates can be found in [2].

Example 0.3. *(Credit risk)* One of the main issues in credit risk is the modeling of the dynamic evolution of the default state of a given portfolio. To formalize the problem, given a portfolio of m obligors, let $y_t := (y_{t,1}, \cdots, y_{t,m})$ be the default indicator process where $y_{t,i} := \mathbf{1}_{\{\tau_i \leq t\}}$ with τ_i the random default time of obligor i; $i = 1, \cdots, m$. In line with the factor modeling philosophy it is natural to assume that default intensities depend on an unobservable latent process x_t. In particular, if $\lambda_i(t)$ is the default intensity of obligor i; $i = 1, \cdots, m$, assume $\lambda_i(t) = \lambda_i(x_t)$. Note that this generates *information-driven contagion*: it is in fact well known that the intensities with respect to $\mathcal{F}_t$ are given by $\hat{\lambda}_i(t) = E\{\lambda_i(x_t) \mid \mathcal{F}_t\}$. Hence the news that an obligor has defaulted leads, via filtering, to an update of the distribution of x_t and thus to a jump in the default intensities of the still surviving obligors. In this context we shall consider the pricing of illiquid credit derivatives on the basis of the investor filtration supposed to be given by the default history and noisily observed prices of liquid credit derivatives.

We assume that, conditionally on x_t, the defaults are independent with intensities $\lambda_i(x_t)$ and that (x_t, y_t) is jointly Markov. A credit derivative has the payoff linked to default events in a given reference portfolio and so one can think of it as a r.v. $H \in \mathcal{F}_T^y$ with T the maturity. Its full information price at the generic $t \leq T$, i.e. in the filtration $\mathcal{G}_t$ that measures also x_t, is given by $\tilde{H}_t = E\{e^{-r(T-t)} H \mid \mathcal{G}_t\}$ where r is the short rate and the expectation is under a given martingale measure Q. By the Markov property of (x_t, y_t) one gets a representation of the form

$$\tilde{H}_t = E\{e^{-r(T-t)} H \mid \mathcal{G}_t\} := a(t, x_t, y_t) \tag{21}$$

for a suitable $a(\cdot)$. In addition to the default history we assume that the investor filtration includes also noisy observations of liquid credit derivatives. In view of (21) it is reasonable to model such observations as

$$dz_t = \gamma(t, x_t, y_t)dt + d\beta_t \tag{22}$$

where the various quantities may also be column vectors, β_t is an independent Wiener process and $\gamma(\cdot)$ is a function of the type of $a(\cdot)$ in (21). The investor filtration is then $\mathcal{F}_t = \mathcal{F}_t^y \vee \mathcal{F}_t^z$. The price at $t < T$ of the credit derivative in the investor filtration is now $H_t = E\{e^{-r(T-t)} H \mid \mathcal{F}_t\}$ and by Lemma 0.1 we have

$$H_t = E\{e^{-r(T-t)} H \mid \mathcal{F}_t\} = E\{a(t, x_t, y_t) \mid \mathcal{F}_t\} \tag{23}$$

Again, if one knows the price $a(t, x_t, y_t)$ in $\mathcal{G}_t$, one can thus obtain the price in $\mathcal{F}_t$ by computing the right hand side in (23) and for this we need the filter distribution of x_t given $\mathcal{F}_t$.

To define the corresponding filtering problem we need a more precise model for (x_t, y_t) (the process z_t is already given by (22)). Since y_t is a jump process, the model cannot be one of those for which we had described an explicit analytic solution. Without entering into details, we refer to [13] (see also [14]), where a jump-diffusion model is considered that allows for common jumps between x_t and y_t. In [13] it is shown that an arbitrarily good approximation to the filter solution can be obtained both analytically, as well as by particle filtering.

We conclude this section with a couple of additional remarks:

1. Traditional credit risk models are either structural models or reduced-form (intensity-based) models. Example 0.3 belongs to the latter class. In *structural models* the default of the generic obligor/firm i is defined as the first passage time of the asset value $V_i(t)$ of the firm at a given (possibly stochastic) barrier $K_i(t)$, i.e.

$$\tau_i = \inf\{t \geq 0 \mid V_i(t) \leq K_t(t)\} \tag{24}$$

In such a context, filtering problems may arise when either $V_i(t)$ of $K_i(t)$ or both are not exactly known/observable (see e.g. [9]).

2. *Can a structural model also be seen as a reduced-form model?* At first sight this is not clear since τ_i in (24) is predictable, while in intensity-based models it is totally inaccessible. It turns however out (see e.g. [16]) that, while τ_i in (24) is predictable with respect to the full filtration (measuring also $V_i(t)$ and $K_i(t)$), it becomes totally inaccessible in the smaller investor filtration that, say, does not measure $V_i(t)$ and it admits furthermore an intensity.

Filtering in portfolio management problems

Rather than presenting a general treatment (for this we refer to [21] and the references therein), we discuss here two specific examples in models with unobserved factors, one in discrete time and one in continuous time. Contrary to the previous section on pricing, here we shall work under the physical measure P.

A discrete time case

To motivate the model, start from the classical continuous time asset price model $dS_t = S_t[adt + x_t dw_t]$ where w_t is Wiener and x_t is the non directly observable volatility process (factor). For $y_t := \log S_t$ one then has

$$dy_t = \left(a - \frac{1}{2}x_t^2\right) dt + x_t dw_t \tag{25}$$

Passing to discrete time with step δ, let for $t = 0, \cdots, T$ the process x_t be a Markov chain with m states $x^1, \cdots, x^m$ (may result from a time discretization of a continuous time x_t) and

$$y_t = y_{t-1} + \left(a - \frac{1}{2}x_{t-1}^2\right)\delta + x_{t-1}\sqrt{\delta}\varepsilon_t \tag{26}$$

with ε_t i.i.d. standard Gaussian as it results from (25) by applying the Euler-Maruyama scheme. Notice that (x_t, y_t) is Markov. Having for simplicity only one stock to invest in, denote by ϕ_t the number of shares of stock held in the portfolio in period t with the rest invested in a riskless bond B_t (for simplicity assume $r = 0$). The corresponding self financed wealth process then evolves according to

$$V_{t+1}^{\phi} = V_t^{\phi} + \phi_t\left(e^{y_{t+1}} - e^{y_t}\right) := F\left(V_t^{\phi}, \phi_t, y_t, y_{t+1}\right) \tag{27}$$

and ϕ_t is supposed adapted to $\mathcal{F}_t^y$; denote by $\mathcal{A}$ the class of such strategies. Given a horizon T, consider the following investment criterion

$$\begin{aligned} J_{opt}(V_0) &= \sup_{\phi\in\mathcal{A}} J(V_0,\phi) \\ &= \sup_{\phi\in\mathcal{A}} E\left\{\sum_{t=0}^{T-1} r_t(x_t, y_t, V_t^{\phi}, \phi_t) + f(x_T, y_T, V_T^{\phi})\right\} \end{aligned} \tag{28}$$

which, besides portfolio optimization, includes also hedging problems. Problem (26),(27),(28) is now a stochastic control problem under partial/incomplete information given that x_t is an unobservable factor process.

A standard approach to dynamic optimization problems under partial information is to transform them into corresponding complete information ones whereby x_t is replaced by its filter distribution given $\mathcal{F}_t^y$. Letting then $\pi_t^i := P\{x_t = x^i \mid \mathcal{F}_t^y\},\ i = 1, \cdots, m$ we shall first adapt the filter dynamics in (5) to our situation to derive a recursive relation for $\pi_t = (\pi_t^1, \cdots, \pi_t^m)$. Being x_t finite-state Markov, $p(x_{t+1} \mid x_t)$ is given by the transition probability matrix and the integral in (5) reduces to a sum. On the other hand, $p(y_t \mid x_t)$ in (5) corresponds to the model in (2) that does not include our model (26) for y_t. One can however easily see that (26) leads to a distribution of the form $p(y_t \mid x_{t-1}, y_{t-1})$ and (5) can be adapted to become here

$$\left\{\begin{array}{lcl} \pi_0 &=& \mu \quad \text{initial distribution for } x_t) \\ \pi_t^i &\propto& \sum_{j=1}^m p\left(y_t \mid x_{t-1} = j, y_{t-1}\right) p\left(x_t = i \mid x_{t-1} = j\right)\, \pi_{t-1}^j \end{array}\right. \tag{29}$$

In addition we may consider the law of y_t conditional on $(\pi_{t-1}, y_{t-1}) = (\pi, y)$ that is given by

$$Q_t(\pi, y, dy') = \sum_{i,j=1}^m p\left(y' \mid x_{t-1} = j, y\right)\, p\left(x_t = i \mid x_{t-1} = j\right)\, \pi^j. \tag{30}$$

From (29), (30) it follows easily that (π_t, y_t) is a sufficient statistic and an $\mathcal{F}_t^y-$Markov process.

To transform the original partial information problem with criterion (28) into a corresponding complete observation problem, put $\hat{r}_t(\pi, y, v, \phi) = \sum_{i=1}^m r_t(x^i, y, v, \phi)\pi^i$ and $\hat{f}(\pi, y, v) = \sum_{i=1}^m f(x^i, y, v)\pi^i$ so that, by double conditioning, one obtains

$$\begin{aligned} J(V_0,\phi) = E\Big\{&\textstyle\sum_{t=0}^{T-1} E\left\{r_t(x_t, y_t, V_t^{\phi}, \phi_t) \mid \mathcal{F}_t^y\right\} \\ &+ E\left\{f(x_T, y_T, V_T^{\phi}) \mid \mathcal{F}_T^y\right\}\Big\} \\ = E\Big\{&\textstyle\sum_{t=0}^{T-1} \hat{r}_t(\pi_t, y_t, V_t^{\phi}, \phi_t) + \hat{f}(\pi_T, y_T, V_T^{\phi})\Big\} \end{aligned} \tag{31}$$

Due to the Markov property of (π_t, y_t) one can write the following (backwards) dynamic programming recursions

$$\left\{\begin{array}{l} u_T(\pi, y, v) = \hat{f}(\pi, y, v) \\ u_t(\pi, y, v) = sup_{\phi\in A}\big[\hat{r}_t(\pi, y, v, \phi) \\ \quad +E\left\{u_{t+1}(\pi_{t+1}, y_{t+1}, F(v, \phi, y, y_{t+1})) \mid (\pi_t, y_t) = (\pi, y)\right\}\big] \end{array}\right. \tag{32}$$

where the function $F(\cdot)$ was defined in (27), and ϕ here refers to the generic choice of $\phi = \phi_t$ in period t. It leads to the optimal investment strategy ϕ^* and the optimal value $J_{opt}(V_0) = u_0(\mu, y_0, V_0)$. It can in fact be shown that the strategy and value thus obtained are optimal also for the original incomplete information problem when ϕ there is required to be $\mathcal{F}_t^y-$adapted.

To actually compute the recursions in (32) one needs the conditional law of (π_{t+1}, y_{t+1}) given (π_t, y_t), which can be deduced from (29) and (30). In this context notice that, even if x is $m-$valued, π_t takes values in the $m-$dimensional simplex that is $\infty-$valued. To actually perform the calculation one needs an approximation leading to a finite-valued process (π_t, y_t) and to this effect various approaches have appeared in the literature (for an approach with numerical results see [4]).

A continuous time case

Consider the following market model where x_t is an unobserved factor process and S_t is the price of a single risky asset

$$\left\{\begin{array}{lcl} dx_t &=& F_t(x_t)dt + R_t(x_t)dM_t \\ dS_t &=& S_t\left[a_t(S_t, x_t)dt + \sigma_t(S_t)dw_t\right] \end{array}\right. \tag{33}$$

with w_t a Wiener process and M_t a not necessarily continuous martingale, independent of w_t. Since in continuous time $\int_0^t \sigma_s^2 ds$ can be estimated by the empirical quadratic variation of S_t, in order not to have degeneracy in the filter to be derived below for x_t, we do not let $\sigma(\cdot)$ depend also on x_t. For the riskless asset we assume for simplicity that its price is $B_t \equiv const$ (short rate $r = 0$). In what follows it will be convenient to consider log-prices $y_t = \log S_t$ for which

$$\begin{aligned} dy_t &= \left[a_t(S_t,x_t) - \frac{1}{2}\sigma_t^2(S_t)\right] dt + \sigma(S_t) dw_t \\ &:= A_t(y_t,x_t)dt + B(y_t)dw_t \end{aligned} \quad (34)$$

Investing in this market in a self-financing way and denoting by ρ_t the fraction of wealth invested in the risky asset, we have from $\frac{dV_t}{V_t} = \rho_t \frac{dS_t}{S_t} = \rho_t \frac{de^{y_t}}{e^{y_t}}$ that

$$dV_t = V_t \left[\rho_t \left(A_t(y_t,x_t) + \frac{1}{2}B_t^2(y_t)\right) dt + \rho_t B_t(y_t) dw_t\right] \quad (35)$$

We want to consider the problem of maximization of expected utility from terminal wealth, without consumption, and with a power utility function. Combining (33), (34) and (35) we obtain the following portfolio optimization problem under incomplete information where the factor process x_t is not observed and where we shall require that ρ_t is $\mathcal{F}_t^Y-$adapted

$$\begin{cases} dx_t = F_t(x_t)dt + R_t(x_t)dM_t \quad \text{(unobserved)} \\ dy_t = A_t(y_t,x_t)dt + B(y_t)dw_t \quad \text{(observed)} \\ dV_t = V_t \left[\rho_t \left(A_t(y_t,x_t) + \frac{1}{2}B_t^2(y_t)\right) dt + \rho_t B_t(y_t) dw_t\right] \\ \sup_\rho E\left\{(V_T)^\mu\right\}, \quad \mu \in (0,1) \end{cases} \quad (36)$$

As in the previous discrete time case, we shall now transform this problem into a corresponding one under complete information thereby replacing the unobserved state variable x_t by its filter distribution given $\mathcal{F}_t^y$, namely $\pi_t(x) := p(x_t \mid \mathcal{F}_t^y)_{x_t=x}$. Even if x_t is finite-dimensional, $\pi_t(\cdot)$ is $\infty-$dimensional. We have seen above cases where the filter distribution is finitely parametrized, namely the linear-Gaussian case (*Kalman filter*) and when x_t is finite-state Markov. The parameters characterizing the filter were seen to evolve over time driven by the innovations process (see (9),(11) and (14)). In what follows we shall then assume that the filter is parametrized by a vector process $\xi_t \in \mathbb{R}^p$, i.e. $\pi_t(x) := p(x_t \mid \mathcal{F}_t^y)_{x_t=x} = \pi(x;\xi_t)$ and that ξ_t satisfies

$$d\xi_t = \beta_t(y_t,\xi_t)dt + \eta_t(y_t,\xi_t)d\bar{w}_t \quad (37)$$

where $\bar{w}_t$ is Wiener and given by the innovations process. We shall now specify this innovations process $\bar{w}_t$ for our general model (36). To this effect, putting $A_t(y_t,\xi_t) := \int A_t(y_t,x)\, d\pi_t(x;\xi_t)$, let

$$d\bar{w}_t := B_t^{-1}(y_t)\left[dy_t - A_t(y_t,\xi_t)\, dt\right] \quad (38)$$

and notice that, replacing dy_t from (34), this definition implies a translation of the original $(P,\mathcal{F}_t)-$Wiener w_t, namely

$$d\bar{w}_t = dw_t + B_t^{-1}(y_t)\left[A_t(y_t,x_t) - A_t(y_t,\xi_t)\right] dt \quad (39)$$

and thus the implicit change of measure $P \rightarrow \bar{P}$ with

$$\begin{aligned} \frac{d\bar{P}}{dP}_{\mid \mathcal{F}_T} = \exp\{ &\int_0^T \left[A_t(y_t,\xi_t) - A_t(y_t,x_t)\right] B_t^{-1}(y_t) dw_t \\ &- \frac{1}{2}\int_0^T \left[A_t(y_t,\xi_t) - A_t(y_t,x_t)\right]^2 B_t^{-2}(y_t) dt\} \end{aligned} \quad (40)$$

We obtain thus as complete information problem corresponding to (36) the following one, which is defined on the space $(\Omega,\mathcal{F},\mathcal{F}_t,\bar{P})$ with Wiener $\bar{w}_t$:

$$\begin{cases} d\xi_t = \beta_t(y_t,\xi_t)dt + \eta_t(y_t,\xi_t)\, d\bar{w}_t \\ dy_t = A_t(y_t,\xi_t)dt + B_t(y_t)d\bar{w}_t \\ dV_t = V_t \left[\rho_t \left(A_t(y_t,\xi_t) + \frac{1}{2}B_t^2(y_t)\right) dt + \rho_t B_t(y_t) d\bar{w}_t\right] \\ \sup_\rho \bar{E}\left\{(V_T)^\mu\right\}, \quad \mu \in (0,1) \end{cases} \quad (41)$$

One can now use methods for complete information problems to solve (41) and it can also be shown that the solution to (41) gives a solution of the original problem for which ρ_t was assumed $\mathcal{F}_t^y-$adapted.

We just remark that also other reformulations of the incomplete information problem as a complete information one are possible (see e.g. [20]).

A final comment concerns hedging under incomplete information (incomplete market). When using the quadratic hedging criterion namely $\min_\rho E_{S_0,V_0}\left\{(H_T - V_T^\rho)^2\right\}$, its quadratic nature implies that, if $\phi_t^*(x_t,y_t)$ is the optimal strategy (number of units invested in the risky asset) under complete information also of x_t then, under the partial information $\mathcal{F}_t^y$, the optimal strategy is simply the projection $E\{\phi_t^*(x_t,y_t) \mid \mathcal{F}_t^y\}$ that can be computed on the basis of the filter of x_t given $\mathcal{F}_t^y$ (see [12]).

Wolfgang Runggaldier

References

[1] A. Bain, D. Crisan, *Fundamentals of Stochastic Filtering.* Series: Stochastic Modelling and Applied Probability , Vol. 60, Springer Science+Business Media, New York, 2009.

[2] R. Bhar, C. Chiarella, H. Hung, and W. Runggaldier, The volatility of the instantaneous spot interest rate implied by arbitrage pricing - a dynamic Bayesian approach. *Automatica*, 42:1381–1393, 2005.

[3] A. Budhiraja, L. Chen, and C. Lee, A survey of nonlinear methods for nonlinear filtering problems. *Physica D*, 230:27–36, 2007.

[4] M. Corsi, H. Pham, and W.J. Runggaldier, Numerical Approximation by Quantization of Control Problems in Finance under Partial Observations. To appear in : Mathematical Modeling and Numerical Methods in Finance. Handbook of Numerical Analysis, Vol XV. (A.Bensoussan, Q.Zhang, eds.) Elsevier 2008.

[5] D. Crisan, P. Del Moral, and T. Lyons, Interacting particle systems approximations of the Kushner-Stratonovich equation. *Advances in Applied Probability*, 31:819–838, 1999.

[6] J. Cvitanic, R. Liptser, and B. Rozovski, A filtering approach to tracking volatility from prices observed at random times. *The Annals of Applied Probability*, 16:1633–1652, 2006.

[7] J. Cvitanic, B. Rozovski and I. Zaliapin, Numerical estimation of volatility values from discretely observed diffusion data. *Journal of Computational Finance*, 9: 1–36, 2006.

[8] M.H.A. Davis, and S.I. Marcus, An Introduction to Nonlinear Filtering. In: M.Hazewinkel and J.C.Willems eds., *Stochastic Systems: The Mathematics of Filtering and Identification and Applications*: 53–75. D.Reidel, Dordrecht, 1981.

[9] D. Duffie and D. Lando, Term structure of credit risk with incomplete accounting observations. *Econometrica*, 69:633–664, 2001.

[10] R. J. Elliott, New finite-dimensional filters and smoothers for noisily observed Markov chains. *IEEE Trans. Info. theory*, IT-39: 265–271, 1993.

[11] R.J. Elliott, L. Aggoun, and J.B. Moore, *Hidden Markov Models: Estimation and Control.* Series: Applications of Mathematics, Vol.29, Springer-Verlag, Berlin-Heidelberg-New York, 1994.

[12] R. Frey and W. Runggaldier, Risk-minimizing hedging strategies under restricted information: the case of stochastic volatility models observed only at discrete random times. *Math. Methods Operations Res.*, 50 (3):339–350, 1999.

[13] R. Frey and W. Runggaldier, Credit risk and incomplete information: a nonlinear filtering approach. Preprint, Universität Leipzig, 2008. Available from www.math.uni-leipzig.de/%7Efrey/publications-frey.html.

[14] R. Frey and W. Runggaldier, Nonlinear Filtering in Models for Interest-Rate and Credit Risk. To appear in: *Handbook of Nonlinear Filtering* (D.Crisan and B.Rozovski eds.). Oxford University Press.

[15] A. Gombani, S. Jaschke, and W. Runggaldier, A filtered no arbitrage model for term structures with noisy data. *Stochastic Processes and Applications*, 115: 381–400, 2005.

[16] R. Jarrow and P. Protter, Structural versus reduced-form models: a new information based perspective. *Journal of Investment management*, 2:1–10, 2004.

[17] W. Kliemann, G. Koch, and F. Marchetti, On the unnormalized solution of the filtering problem with counting process observations. *IEEE*, IT-36: 1415–1425, 1990.

[18] H.J. Kushner, and P. Dupuis, *Numerical Methods for Stochastic Control Problems in Continuous Time.* Series: Applications of Mathematics, Vol.24, Springer, New York, 1992.

[19] R.S. Liptser, and A.N. Shiryaev, *Statistics of random processes, Vols. I, II.* Series: Applications of Mathematics; Stochastic Modelling and Applied Probability. Springer-Verlag, Berlin, 2001.

[20] H. Nagai, and W.J. Runggaldier, PDE approach to utility maximization for market models with hidden Markov factors. In: *Seminar on Stochastic Analysis, Random Fields and Applications V* (R.C.Dalang, M.Dozzi, F.Russo, eds.). Progress in Probability, Vol.59, Birkhuser Verlag, 2008, pp.493–506.

[21] H. Pham, Portfolio optimization under partial observation: theoretical and numerical aspects. To appear in: *Handbook of Nonlinear Filtering* (D.Crisan and B.Rozovski eds.). Oxford University Press.

[22] W. J. Runggaldier, Estimation via stochastic filtering in financial market models. In: G. Yin, and Q. Zhang, editors, *Mathematics of Finance. Contemporary Mathematics, Vol. 351*: 309–318. AMS, 2004.

[23] Y. Zeng, A partially observed model for micromovement of asset prices with Bayes estimation via filtering. *Mathematical Finance*, 13: 411–444, 2003.