Filtering Methods

Andras Fulop

ESSEC Business School, Paris fulop@essec.fr

1 Introduction

The methods described in this chapter are applicable to problems where a hidden dynamic Markov process needs to be filtered from the observed data containing some noise. To illustrate the problem, start with an example from fixed income taken from [12]. Figure 1 plots 18 US treasury zero-coupon bond yields with maturities between 1 and 120 months observed in the period 1970-2000. To summarize the information in these time series, analysts often find it useful to extract a small number of dynamic factors that describe most of the variation in the yields. This parsimonious representation can help both in the interpretation of past yield curve movements and in prediction. However, in general, this low-dimensional factor structure is consistent with the observations only if the observed yields are assumed to contain some measurement errors. To uncover the unobserved dynamic factors, one needs to infer them from the noisy observed yields.

Credit risk modeling is the subject of the second example. Six Flags Inc., a large operator of theme parks has been having financial difficulties in the last couple of years. On January 2, 2008 the company reported total assets of 2945 Million USD, total liabilities of 2912 Millions USD and preferred equities with a book value of 285 Millions, consistent with a negative -252 Millions of shareholders' equity on its balance sheets. However, in 2008, the stocks of the company were not worthless, as reported in Figure 2. The main reason for the positive market value of the stock in spite of the large debt is limited liability. In case the value of the company is less than then the face value of the debt when the debt is to be repaid, the stockholders can default in effect handing the company to the debtholders. As a result of this default option, both the equity and the debt can be interpreted as derivatives written on the face value of the firm. The equity holders own a long call option on the value of the firm with an exercise price equal to the face value of debt, while the debt-owners are short of this position. Unfortunately, the observed equity prices are less then perfect signals on the firm value. The first order autocorrelation coefficient

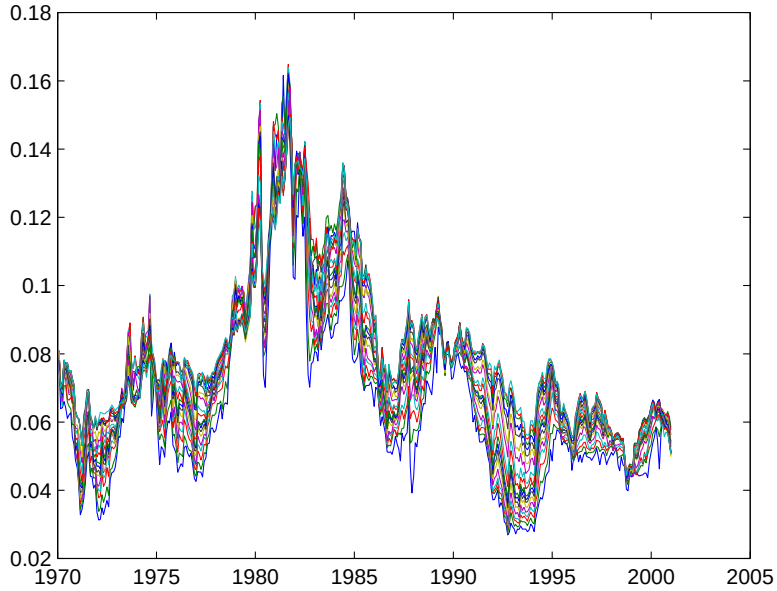

Fig. 1. US treasury zero-coupon-bond yields between 1970-2000

of the equity log-returns is equal to -0.25, showing that a considerable part of the observed price changes are due to transitory microstructure effects unrelated to permanent value innovations. Then, a methodois needed to filter the unobserved asset value of the firm from the noisy observed equity prices.

The chapter begins in Section 2 with the description of the general filtering problem and puts the two examples in the general framework. Section 3 describes Kalman Filtering, applicable to linear normal systems. Here the filtering distributions are normally distributed with a mean and variance that can be recursively updated using the Kalman recursion. The method is applied to the first example on interest rate term structure. Further, some extensions of the Kalman Filter to nonlinear systems are mentioned. Section 4 turns to the general filtering problem, where the dynamic model is nonlinear/nongaussian. Here, the Kalman Filter is not valid any more, but anatheoretical recursions still hold. Unfortunately, these involve integrals that need to be solved numerically. The chapter proceeds by presenting sequential Monte Carlo techniques that have been developed in last 15 year and are routinely used to solve the general filtering problem. It describes the general particle filtering algorithm, where resampling is introduced to tackle sample impoverishment, a pervasive problem in sequential importance sampling. Here and in the remaining part of the chapter, Merton's model with noisy equity observations is used to illus-

Fig. 2. Market Capitalization of Six Flags' equity in 2008

trate the presentation. Section 5 presents various strategies to produce effective proposal distributions, a crucial issue in designing an efficient importance sampling algorithm. The section end by applying the filtering methodoon the second example, in filtering the asset value of Six Flags Inc. from its observed equity prices. Section 6 concludes with a brief discussion of various classical and Bayesian approaches to the estimation of the fixed model parameters in the particle filtering context.

2 The Filtering Problem

Assume that the state of a financial model at time k is described by a random vector x_k whose dynamics follows the transition equation

$$x_{k+1} = Q(x_k, \varepsilon_{k+1}) \tag{1}$$

where $Q()$ is an arbitrary function and ε_k is a sequence of independent random vectors. When x_k is continuous, this defines the conditional probability density $q(x_{k+1} \mid x_k)$. x_k is not directly observable, instead at time k a noisy observation y_k is available, linked to x_k through the measurement equation

$$y_k = G(x_k, \nu_k) \tag{2}$$

where $G()$ is an arbitrary function and ν_k the observation noise is a sequence of random vectors, independent across time and from from ε_k. When y_k is continuous, this defines the conditional probability density $g(y_k \mid x_k)$. Use the following notation

$$x_{0:k} = (x_0, ..., x_k)$$
$$y_{1:k} = (y_1, ..., y_k)$$

Further, assume some prior distribution, $q_0(x_0)$, for the initial state variable. Then, the objective of filtering is to come up with the distribution of the hidden variable, x_k, given the observed data up to k. This quantity is the filtering distribution of x_k and is denoted by $f(x_k \mid y_{1:k})$. In the algorithms that follow these distributions are obtained sequentially, as new observations arrive.

2.1 Uncovering yield curve factors

To tackle the first example in the introduction, this subsection describes a specific factor model of the term structure closely following [12] and shows how it fits into the general filtering framework. Denote by $y(\tau_l)_k$ the zero-coupon yield observations at time k with maturity τ_l. On each observation date k, there are 18 observed yields with maturities ranging between $\tau_1 = 1, \ldots, \tau_{18} = 120$ months. The data-set has monthly observations in the period 1975-2000. To summarize the rich cross-sectional information, the yields are assumed to depend on three common factors $(x_{1,k}, x_{2,k}, x_{3,k})$ and a yield-specific measurement noise $\nu_{l,k}$. This latter is assumed to be standard normal and independent across the yields and through time. This setup leads to the following measurement equations for $l = 1, \ldots, 18$

$$y(\tau_l)_k = x_{1,k} + x_{2,k}\left(\frac{1 - e^{-\lambda\tau_l}}{\lambda\tau_l}\right) + x_{3,k}\left(\frac{1 - e^{-\lambda\tau_l}}{\lambda\tau_l} - e^{-\lambda\tau_l}\right) + \sigma_\nu \nu_{l,k} \quad (3)$$

This factor representation is a version of the Nelson-Siegel [30] parametric form, popular with practitioners. The interpretability of the factors is an attractive feature of this specific parameterization. First, $x_{1,k}$ has the same loading on each yield, so it can be interpreted as a level factor. Second, $x_{2,k}$ affects yields with longer maturities less, hence it is close to a slope factor. Last, $x_{3,k}$ has hump-shaped loadings and plays the role of a curvature factor. The parameter λ determines where the maximum of this hump-shaped pattern lies.

To ensure some degree of time-series consistency and to allow prediction using the model, the factors are assumed to follow independent normal AR(1) processes, resulting in the following transition equations

$$x_{i,k+1} = \mu_i + \gamma_i x_{i,k} + \sigma_{i,x}\epsilon_{i,k+1} \quad (4)$$

where $\epsilon_{i,k+1}$ are independent standard normal variables. Then, if one wants to forecast the future yields for instance, one needs to filter the last value of the unobserved factors, $x_{i,k}$ given the noisy yield observations up to k.

2.2 Finding the value of the firm in Merton's model

[29] laid the foundation to the literature on the structural approach to credit risk modeling. The value of the firm at time t, V_t, is assumed to follow a geometric Brownian motion with respect to the physical probability law that generates the asset values

$$\frac{dV_t}{V_t} = \mu dt + \sigma dW_t$$

The risk-free rate of interest is assumed to be a constant, r. Furthermore, the firm has two classes of claims outstanding – an equity and a zero-coupon debt maturing at time T with face value F. Due to limited liability, equity is a call option on the value of the firm with payout

$$S_T = \max(V_T - F, 0) \tag{5}$$

Then, the equity claim in equation (5) can be priced at time $t < T$ by the standard Black-Scholes option pricing model to yield the following solution:

$$S_t \equiv S(V_t; \sigma, F, r, T-t) = V_t \Phi(d_t) - F e^{-r(T-t)} \Phi(d_t - \sigma\sqrt{T-t}) \tag{6}$$

where

$$d_t = \frac{\log(\frac{V_t}{F}) + (r + \frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}}$$

and $\Phi(\cdot)$ is the standard normal distribution function.

Unfortunately, the asset value of the firm, V_{τ_i} is rarely observable. In contrast, for an exchange listed firm, one can obtain a time series of equity prices denoted by $\mathcal{D}_N = \{S_{\tau_i}, i = 0, \cdots, N\}$ and try to infer the asset value using the equity prices and balance sheet information on debt. If the equity prices are not contaminated by trading noises, the asset value can be obtained by inverting the equity pricing function from equation (6) following [16]. However the observed equity prices may be contaminated by microstructure noise that can be important, especially for smaller firms or firms in financial difficulties. Following [19] the trading noise obeys a multiplicative structure leading to the following measurement equation for the log equity price

$$\log S_{\tau_i} = \log S(V_{\tau_i}; \sigma, F, r, T - \tau_i) + \delta\nu_i \tag{7}$$

where $\{\nu_i, i = 0, N\}$ are i.i.d. standard normal random variables and the nonlinear pricing function $S(V_t; \sigma, F, r, T-t)$ has been given earlier. Since the unobserved asset value process follows the geometric Brownian motion, we can derive its discrete-time form as

$$\log V_{\tau_{i+1}} = \log V_{\tau_i} + (\mu - \frac{\sigma^2}{2})h + \sigma\sqrt{h}\varepsilon_{i+1} \tag{8}$$

where $\{\varepsilon_i, i = 1, N\}$ are i.i.d. standard normal random variables and $h = \tau_i - \tau_{i-1}$ is the observation frequency. Then, one needs to filter the unobserved asset price, V_{τ_i} given the noisy equity observations up to time k in the model defined by the measurement equation (7) and the transition equation (8).

3 Kalman Filtering

When the measurement and the transition equations are normal and linear, the filtering density is normal. Assume that the transition equation is

$$x_k = C + Ax_{k-1} + \varepsilon_k \tag{9}$$

where $\varepsilon_k \sim N(0, Q)$. The measurement equation is also linear and normal:

$$y_k = Hx_k + \nu_k \tag{10}$$

where $\nu_k \sim N(0, R)$. Introduce the following notation for conditional expectations and variances:

$$\begin{aligned} E_s(x_k) &= E(x_k \mid y_{1:s}) \\ V_s(x_k) &= Var(x_k \mid y_{1:s}) \end{aligned}$$

if the initial state x_0 is distributed as $x_0 \sim N(E_0(x_0), V_0(x_0))$, the subsequent filtering distributions are also normally distributed. Further, the first two moments can be sequentially updated by first predicting the distribution of the hidden variable at k given past information up to $k-1$

$$\begin{aligned} E_{k-1}(x_k) &= C + AE_{k-1}(x_{k-1}) \\ V_{k-1}(x_k) &= AV_{k-1}(x_{k-1})A' + Q \end{aligned}$$

Then, the filtering distribution at k is obtained by including the information at k

$$\begin{aligned} K_k &= V_{k-1}(x_k)H' \left(HV_{k-1}(x_k)H' + R\right)^{-1} \\ E_k(x_k) &= E_{k-1}(x_k) + K_k \left(y_k - HE_{k-1}(x_k)\right) \\ V_k(x_k) &= \left(I - K_k H\right) V_{k-1}(x_k) \end{aligned}$$

For a general review of Kalman Filtering see [1]. For an application in yield curve modeling see [6],[12] and another in commodities see [34].

3.1 Application of Kalman Filtering: Uncovering yield curve factors

It is apparent that the first example, on extracting yield curve factors, falls within the realm of Kalman Filtering. In particular both the measurement equation in 3 and the transition equation in 4 are gaussian and linear. To investigate the method on the US zero-coupon yield data-set, the parameters of the model are fitted using maximum likelihood[1]. In-sample, the model-predicted yields have a root mean squared error of around 12 basis points, pointing towards a satisfactory fit. Figure 3 plots the model-implied filtered mean of the factors and the results seem to be in accord with intuition. For example, one can see that the first factor indeed acts as a level factor, with high values when the general level of interest rates is high.

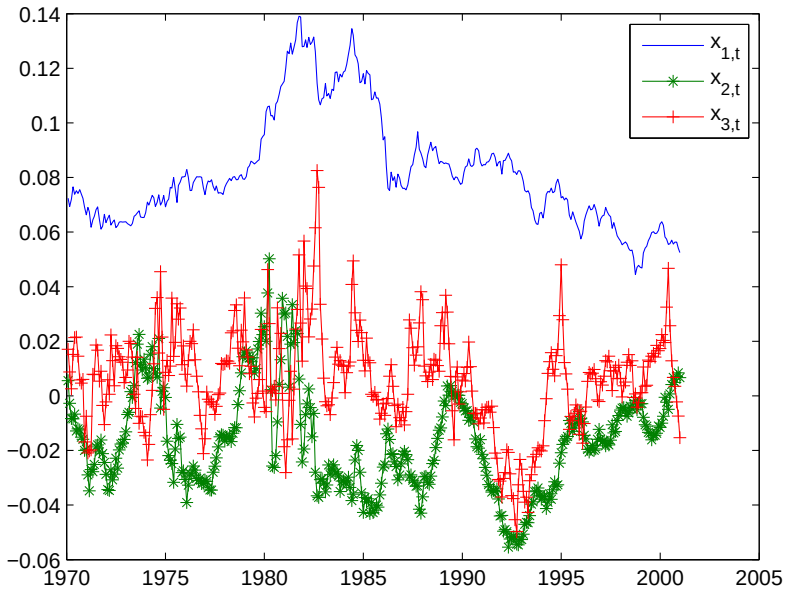

Fig. 3. Time series of filtered yield curve factors

Forecasting is an important application of yield-curve models. Further, investigating the out-of-sample performance of various models may be an even more important check on model validity than the in-sample fit. Hence, following [12], the out-of-sample forecasting performance of the yield curve factor

[1] Note that similarly to [12] $\lambda = 0.0609$ is chosen exogenously. This corresponds to a peak in teh curvature factor at 30 months.

model is compared with two competitors, the first being a naive random walk model while the second is an AR(1) model of the individual yields. All the models are estimated on the data up to 1993 and the quality of their forecasts is investigated on the remaining sample at the 6-months horizon. Figure 4 shows the RMSE of the three forecasts for all maturities and provides evidence that the discipline that the factor model puts on the data considerably helps in prediction. All the results in this subsection are produced by the MATLAB script `DieboldLi_KF.m`.

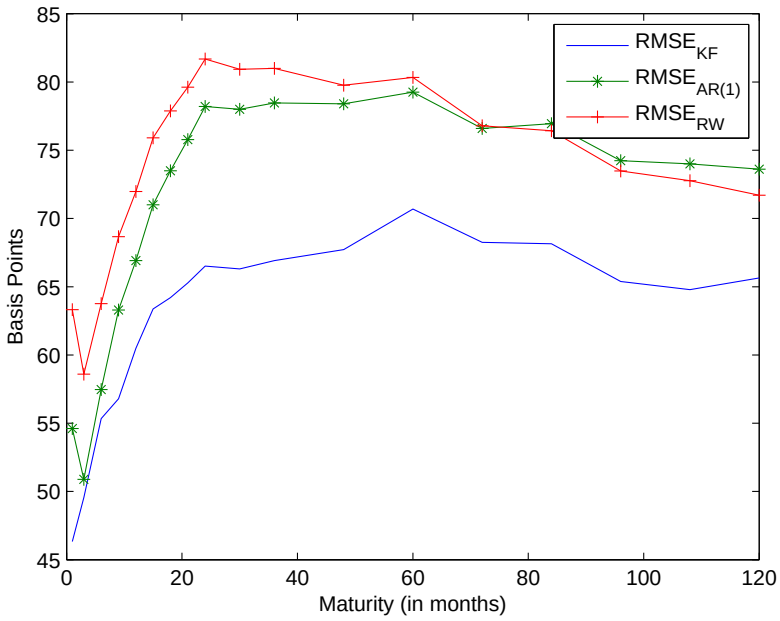

Fig. 4. RMSE of various forecasting methods on the 6-months horizon, 1994-2000

3.2 Extensions

Extended Kalman Filter (EKF)

Often, the financial model of interest is normal, but the transition and measurement equation are not linear. In particular we may have

$$x_k = Q(x_{k-1}, \varepsilon_k) \tag{11}$$
$$y_k = G(x_k, \nu_k) \tag{12}$$

where $Q()$ and $G()$ are differentiable functions and ε_k and ν_k are normally distributed. Then, the Extended Kalman Filter (EKF) approximates this system using a first-order Taylor expansion around $E_{k-1}(x_{k-1})$ and applies Kalman Filtering on the approximating linear system. In finance, this approach is often applied in term structure modeling([10], [20], [21]) and in commodities modeling([36]).

Unscented Kalman Filter (UKF)

The EKF approximates the system only up to a first order and it can provide poor results when the nonlinearity of the measurement or transition equation is serious. An alternative approach that avoids linearization altogether is the Unscented Kalman Filter (UKF). This method approximates the normal filtering distribution using a discrete distribution that matches the mean and covariance matrix of the target gaussian random variable. Then, these points are passed through directly the nonlinear functions to obtain the quantities necessary for the Kalman recursion. In many situations the method provides a higher order approximation to the nonlinear system than the EKF. For a detailed description of the method see [38]. The technique has been applied to currency option pricing by [3].

4 Particle Filtering

4.1 General Filtering Recursion

When the system is non-linear and/or non-gaussian, the filtering distribution may not be normal and the Kalman Filter is not valid any more. To appreciate the difficulty of the task, in the following we describe the sequential filtering problem in the general model described by (1) and (2).

The joint filtering distribution of $x_{0:k}$ given $y_{1:k}$ is

$$f(x_{0:k} \mid y_{1:k}) = \frac{f(x_{0:k}, y_{1:k})}{f(y_{1:k})} = \frac{f(x_{0:k}, y_{1:k})}{L(y_{1:k})}$$

where $L(y_{1:k})$ is the likelihood of the data observed up to k

$$L(y_{1:k}) = \int f(x_{0:k}, y_{1:k}) dx_{0:k}$$

Now derive the recursive formula connecting the filtering distributions at k and $k+1$

$$\begin{aligned} & f(x_{0:k+1} \mid y_{1:k+1}) \\ = & \frac{f(x_{0:k+1}, y_{1:k+1})}{L(y_{1:k+1})} \end{aligned}$$

$$= \frac{g(y_{k+1} \mid x_{k+1})q(x_{k+1} \mid x_k)f(x_{0:k}, y_{1:k})}{L(y_{1:k})} \frac{L(y_{1:k})}{L(y_{1:k+1})}$$
$$= \frac{g(y_{k+1} \mid x_{k+1})q(x_{k+1} \mid x_k)}{f(y_{k+1} \mid y_{1:k})} f(x_{0:k} \mid y_{1:k})$$

This equation gives the recursion of the filtered distributions over the whole path space. Integrating over $x_{0:k-1}$ one gets the following relationship

$$\begin{aligned} & f(x_{k:k+1} \mid y_{1:k+1}) \\ &= \frac{g(y_{k+1} \mid x_{k+1})q(x_{k+1} \mid x_k)}{f(y_{k+1} \mid y_{1:k})} f(x_k \mid y_{1:k}) \\ &\propto g(y_{k+1} \mid x_{k+1})q(x_{k+1} \mid x_k)f(x_k \mid y_{1:k}) \end{aligned}$$

showing that $f(x_{0:k} \mid y_{1:k})$ is a sufficient statistic. Integrating out x_k, one arrives at the filtering distribution of x_{k+1}

$$\begin{aligned} & f(x_{k+1} \mid y_{1:k+1}) \\ &\propto \int g(y_{k+1} \mid x_{k+1})q(x_{k+1} \mid x_k)f(dx_k \mid y_{1:k}) \end{aligned}$$

The Kalman Filter is a special case where this recursion can be executed in closed-form due to the joint normality of the system. In general, the filtering distributions do not belong to a known parametric family and the integration has to be done using numerical methods. In the following a class of simulation-based methods is presented that has been extensively used in the last few years to solve the general filtering task.

4.2 Sequential Importance Sampling

The target is the joint filtering distribution of the hidden states

$$f(x_{0:k} \mid y_{1:k}) \propto \prod_{t=1}^{k} g(y_t \mid x_t)q(x_t \mid x_{t-1})q_0(x_0) \tag{13}$$

Ideally, one would like to sample directly from the densities $g(y_t \mid x_t)q(x_t \mid x_{t-1})$, providing a straightforward recursive Monte Carlo scheme. Unfortunately, due to the complexity of these densities, this is usually not possible. Importance sampling is an approach that can be used in such cases. Here, one draws from a feasible proposal distribution $r(x_{0:k})$ instead of the target and attaches importance weights to the samples to compensate for the discrepancy between the proposal and the target. If the weighted sample is denoted by $(\xi_{0:k}^{(m)}, w_k^{(m)})$ where $m = 1, ..., M$, the samples and weights are obtained as

$$\begin{aligned} \xi_{0:k}^{(m)} &\sim r(x_{0:k}) \\ w_k^{(m)} &= \frac{\prod_{t=1}^{k} g(y_t \mid \xi_t^{(m)})q(\xi_t^{(m)} \mid \xi_{t-1}^{(m)})q_0(\xi_0^{(m)})}{r(\xi_{0:k}^{(m)})} \end{aligned}$$

The expectation $E\left(h(x_{0:k} \mid y_{1:k})\right)$ can be estimated by the estimator

$$\widehat{h} = \frac{\sum_{m=1}^{M} h(\xi_{0:k}^{(m)}) w_k^{(m)}}{\sum_{m=1}^{M} w_k^{(m)}}$$

Using independence of the sample the estimator is asymptotically consistent

$$\widehat{h} - E\left(h(x_{0:k} \mid y_{1:k})\right) \rightarrow_P 0 \text{ as } M \rightarrow \infty$$

and asymptotically normal

$$\sqrt{M}\left[\widehat{h} - E\left(h(x_{0:k} \mid y_{1:k})\right)\right] \rightarrow_D N\left[0, \frac{Var_r\left(h(x_{0:k}) w(x_{0:k})\right)}{\left[E_r\left(w(x_{0:k})\right)\right]^2}\right] \text{ as } M \rightarrow \infty$$

Note that the asymptotic variance can also be estimated using the simulation output, allowing inference on the reliability of the estimate.

The preceding importance sampling algorithm can be made sequential by choosing a recursive structure for the importance sampling distribution, $r(x_{0:k})$:

$$R(x_{0:k}) = \prod_{t=1}^{k} r(x_t \mid y_k, x_{t-1}) r_0(x_0)$$

Then the importance weight w_k can be written as

$$w_k = \prod_{t=1}^{k} \frac{g(y_t \mid x_t) q(x_t \mid x_{t-1})}{r(x_t \mid y_k, x_{t-1})} \frac{q_0(x_0)}{r_0(x_0)}$$

and the importance sampler can be implemented in a sequential manner

Sequential Importance Sampling

- Initial State: Draw an i.i.d. sample $\xi_0^{(m)}, m = 1, ..., M$ from $\xi_0^i \sim r_0(x_0)$ and set
$$w_0^{(m)} = \frac{q_0(\xi_0^{(m)})}{r_0(\xi_0^{(m)})}, m = 1, ..., M$$
- Recursion: For $k = 1, ..., N$
  1. Draw $(\xi_k^{(m)}, m = 1, ..., M)$ from the distribution $\xi_k^{(m)} \sim r(x_k \mid y_k, \xi_{k-1}^{(m)})$
  2. Compute the updated importance weights

$$w_k^{(m)} = w_{k-1}^{(m)} \times \frac{g(y_k \mid \xi_k^{(m)}) q(\xi_k^{(m)} \mid \xi_{k-1}^{(m)})}{r(\xi_k^{(m)} \mid y_t, \xi_{k-1}^{(m)})}$$

This algorithm seems to provide a solution to the recursive filtering problem. Unfortunately after a couple of time steps the normalized weights of most points fall to zero and the weighted sample ceases to provide a reliable representation of the target distribution.

Weight degeneracy in Merton's model

To illustrate the phenomenon mentioned before, consider the performance of the sequential importance sampling algorithm for Merton's model with noisy equity observations. Choose the prior distribution to be a point mass assuming that the initial equity observation S_{τ_0} is observed without any error. Further, use the transition density $f(V_{\tau_{i+1}} \mid V_{\tau_i}^{(m)})$ as the proposal distribution. The procedure that results is:

Sequential Importance Sampling in Merton's model

- Initial State: Set $V_{\tau_0}^{(m)} = S^{-1}(S_{\tau_0})$ where the function $S^{-1}(.)$ is the inverse of the equity pricing function in (6).
- Recursion: For $k = 1, ..., N$
  1. Draw $V_{\tau_k}^{(m)}$ from $f(V_{\tau_k} \mid V_{\tau_{k-1}}^{(m)}, \Theta)$, which can be easily done using equation (8).
  2. Compute the updated importance weights

$$w_k^{(m)} = w_{k-1}^{(m)} f(S_{\tau_k} \mid V_{\tau_k}^{(m)}, \Theta)$$

One measure of the reliability of an importance sampler is the effective sample size, N_{eff}, defined as

$$N_{eff} = \left[\sum_{m=1}^{M} \left(\frac{w_k^{(m)}}{\sum_{m=1}^{M} w_k^{(m)}} \right)^2 \right]^{-1}$$

Roughly speaking, the effective sample size measures the size of an equally-weighted Monte Carlo sample providing the same reliability as the output of the importance sampler. Figure 5 depicts the effective sample sizes for the first few observations in Merton's model obtained by running the SIS algorithm. The model parameters are $(\sigma = 0.2, \mu = 0.1, r = 0.05, \delta = 0.01, F = 100)$. The initial asset value is 60, the initial debt maturity is 3 years, and a year of daily data is generated (i.e. $h = 1/250, n = 250$) and the sample size is $M = 1000$. The MATLAB file producing this figure is `test_MertonSIS.m`. One can observe that by $t = 5$ the effective sample size collapses to one, signaling the deterioration of the filter. The underlying reason behind this phenomenon is that a fixed number of points is used to cover an increasing dimensional space.

4.3 Sequential Importance Sampling with Resampling (SIR or particle filtering)

To deal with the problem of sample impoverishment, [24] suggest to resample the current population of particles using the normalized weights as

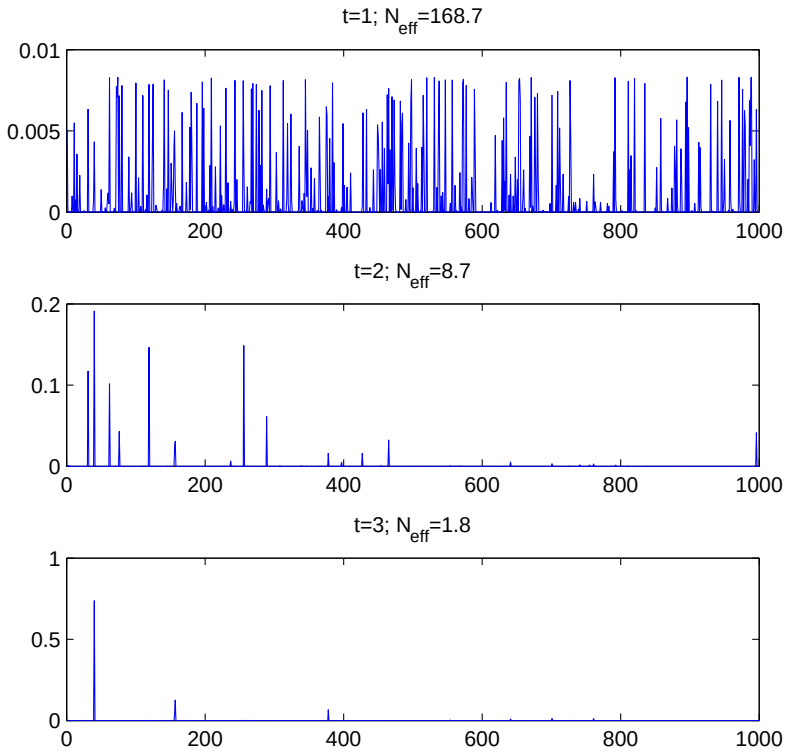

Fig. 5. Normalized importance weights for Sequential Importance Sampling in Merton's model

probabilities of selection. After resampling, all importance weights are reset to one. The intuition behind this procedure is that unlikely trajectories are eliminated and likely ones are multiplied. This yields the following algorithm:

Sequential Importance Sampling with Resampling

- Initial State: Draw an i.i.d. sample $\xi_0^{(m)}$ from $\xi_0^{(m)} \sim r_0(x_0)$ and set $w_0^{(m)} = \frac{q_0(\xi_0^{(m)})}{r_0(\xi_0^{(m)})}, m = 1, ..., M$
- For $k = 1, ..., N$ repeat the next steps
  1. Sampling
     - Draw $(\xi_k^{(m)}, m = 1, ..., M)$ conditionally independently given $(\xi_{0:k-1}^{(m)}, m = 1, ..., M)$ from the distribution $\xi_k^{(m)} \sim r(x_k \mid y_k, \xi_{k-1}^{(m)})$
     - Compute the importance weights

$$w_k^{(m)} = \frac{g(y_k \mid \xi_k^{(m)}) q(\xi_k^{(m)} \mid \xi_{k-1}^{(m)})}{r(\xi_k^{(m)} \mid y_k, \xi_{k-1}^{(m)})}$$

  2. Resampling
     - Draw from the multinomial trial $(I_k^1, ..., I_k^M)$ with probabilities of success

$$\frac{w_k^1}{\sum_{m=1}^M w_k^{(m)}}, ..., \frac{w_k^M}{\sum_{m=1}^M w_k^{(m)}}$$

     - Reset the importance weights $w_k^{(m)}$ to 1;
  3. Trajectory update: $\xi_{0:k}^{(m)} = \xi_{0:k}^{I_k^{(m)}}, m = 1, ..., M$

This approach concentrates on the marginal filtering distribution $f(x_k \mid y_{0:k})$ instead of the joint one, $f(x_{0:k} \mid y_{0:k})$. Resampling helps to achieve a better characterization of the last state of the system at the expense of representing the past of the full hidden path, $x_{0:k}$.

Bootstrap Filter

In the bootstrap filter of [24] the proposal density is chosen to be equal to the transition density

$$r(x_k \mid y_k, x_{k-1}) = q(x_k \mid x_{k-1})$$

In this case the importance weights take a particularly simple form, they simply equal the measurement density

$$w_k^m = \frac{g(y_k \mid \xi_k^m) q(\xi_k^m \mid \xi_{k-1}^m)}{q(\xi_k^m \mid \xi_{k-1}^m)} = g(y_k \mid \xi_k^m)$$

Bootstrap Filter in Merton's model

Bootstrap Filter in Merton's model

- Initial State: Set $V_{\tau_0}^{(m)} = S^{-1}(S_{\tau_0})$ where the function $S^{-1}(.)$ is the inverse of the equity pricing function in equation (6).
- Recursion: For $k = 1, ..., N$
  1. Sampling
     - Draw $V_{\tau_k}^{(m)}$ from $f(V_{\tau_k} \mid V_{\tau_{k-1}}^{(m)}, \Theta)$, using equation(8).
     - Compute the normalized importance weights

$$\pi_k^{(m)} = \frac{w_k^{(m)}}{\sum_{m=1}^{M} w_k^{(m)}} \text{ where } w_k^{(m)} = f(S_{\tau_k} \mid V_{\tau_k}^{(m)}, \Theta)$$

  2. Resample from the weighted sample $\{(V_{\tau_k}^{(m)}, \pi_k^{(m)}); m = 1, \cdots, M\}$ to obtain a new equal-weight sample of size M.

To investigate whether the resampling step successfully deals with sample depletion we repeat the simulation exercise described before on Merton's model, but now we run the bootstrap filter on the simulated data. Panel A of Figure 6 depicts the effective sample sizes (N_{eff}) for a simulated sample path. One can see that now N_{eff} does not collapse as time progresses, so the resampling seems an effective remedy to sample depletion. Panel B reinforces this message by showing that the filter reliably tracks the unobserved asset value path.The MATLAB file producing Figure 6 is `test_MertonBootstrap.m`.

4.4 Theoretical properties of particle filters

The filtering algorithm described above has been shown to possess attractive asymptotic properties as the number of particles, N, goes to infinity. (see [9] for a short introduction to the theory and [11] for a monograph-length treatment) In particular it provides consistent estimates of any filtering quantity

$$\frac{1}{M}\sum_{m=1}^{M} h(\xi_k^m) - E\left(h(x_k \mid y_{1:k})\right) \to 0 \text{ as } M \to \infty$$

Central limit theorems has been proved for particle systems, leading to results of the type

$$\sqrt{M}\left(\frac{1}{M}\sum_{m=1}^{M} h(\xi_k^m) - E(h(x_k \mid y_{1:k}))\right) \to N(0, \sigma_k^2(h)) \text{ as } M \to \infty$$

In general the Monte-Carlo variance, $\sigma_k^2(h)$ increases with the time k. This reflects the fact that as time passes, errors accumulate in the filtering

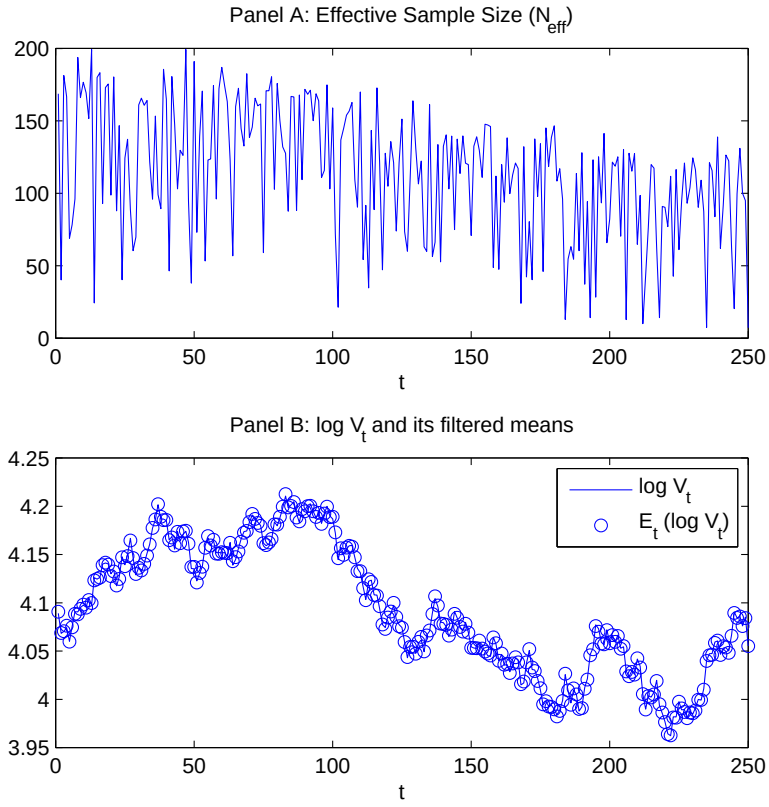

Fig. 6. Bootstrap Filter in Merton's model

recursions. In practice this means that an ever-increasing number of particles is needed to ensure the same quality for the estimates. To rule this out and achieve uniform convergence further assumptions on the forgetting properties of the model are needed.
While these results provide the rate of convergence, $\sqrt{M}$, the constant of convergence, $\sigma_k^2(h)$ is usually not known. This means that in contrast to simple importance sampling, one cannot compute confidence intervals for the estimates.

5 Implementation issues for Particle Filters

5.1 The choice of proposal in SIR

The choice of the proposal distribution is critical for the efficiency of the method. The question is how to best use the information in the next observation in sampling. The optimal choice would be the conditional distribution of the new hidden state given the past hidden state and the new observation:

$$f(x_k \mid y_k, x_{k-1}) \propto g(y_k \mid x_k) q(x_k \mid x_{k-1})$$

As direct sampling from the optimal choice is usually not feasible, approximations are needed. In the following Merton's model is used to illustrate various strategies to obtain efficient proposal distributions. The first approach uses a specific feature of Merton's model by localizing the sampler around the new observation. The second, more generic approach linearizes the model around each particle and uses the optimal sampler of the approximated model as the proposal. The third strategy adapts a parametric family of proposal and picks the best density within this family using the information in the previously sampled particles.

Localized sampling in Merton's model

To illustrate the importance of the proposal density, consider again Merton's model. If one uses the bootstrap filter, the importance weights are

$$f(S_{\tau_{i+1}} \mid V_{\tau_{i+1}}, \Theta) \propto \frac{1}{\delta} \phi \left(\frac{\log S_{\tau_{i+1}} - \log S(V_{\tau_{i+1}}, \Theta)}{\delta} \right)$$

When the microstructure noise δ is small, this density function is peaked, resulting in high variance of the particle weights and a poor representation of the filtering distribution. Intuitively, when the microstructure noise is relatively small, the new observation is very informative on the hidden asset value. This makes the bootstrap sampler that ignores the new observation, a poor choice for the proposal.
However, if the observation is so important, why not totally base the sampler on the observation, forgetting the past? This idea is used in [19] to propose an efficient sampler, localized around the new observed equity price. In particular, [19] suggest to draw from the microstructure noise, ν_k and to use the asset value implied by the noise and the new observation as the sampler. This results in the following algorithm:

Localized Sampling in Merton's model

- Initial State: Set $V_{\tau_0}^{(m)} = S^{-1}(S_{\tau_0})$ where the function $S^{-1}(.)$ is the inverse of the equity pricing function in (6).
- Recursion: For $k = 1, ..., N$
  1. Sampling
     - Draw a standard normal $\nu_k^{(m)}$ and compute $V_{\tau_k}^{(m)} = V_{\tau_k}^*(S_{\tau_k}, \nu_k^{(m)})$ to obtain the pair $(V_{\tau_i}^{(m)}, V_{\tau_k}^{(m)})$, where

$$V_{\tau_k}^*(S_{\tau_k}, \nu_k) = S^{-1}(S_{\tau_k} e^{-\delta\nu_k}; \sigma, F, r, T - \tau_k)$$

     - Compute the importance weights

$$w_k^{(m)} = \frac{f(V_{\tau_k}^{(m)} \mid V_{\tau_{k-1}}^{(m)}, \Theta)}{\Phi(d_{\tau_k}^{*(m)}) e^{\delta\nu_k^{(m)}}}$$

     - Normalize the importance weights

$$\pi_k^{(m)} = \frac{w_k^{(m)}}{\sum_{m=1}^M w_k^{(m)}} \text{ where } w_k^{(m)} = f(S_{\tau_k} \mid V_{\tau_k}^{(m)}, \Theta)$$

  2. Resample from the weighted sample $\{(V_{\tau_k}^{(m)}, \pi_k^{(m)}); m = 1, \cdots, M\}$ to obtain a new equal-weight sample of size M.

Here, using a change of variables formula, the density function of the sampler is

$$\begin{aligned} & g(V_{\tau_k}^{(m)} \mid S_{\tau_k}, V_{\tau_{k-1}}^{(m)}) \\ = & f(V_{\tau_k}^*(S_{\tau_k}, \nu_k^{(m)}) \mid S_{\tau_k}) = \frac{\phi(\nu_k^{(m)}) \Phi(d_{\tau_k}^{*(m)}) e^{\delta\nu_k^{(m)}}}{\delta S_{\tau_k}} \end{aligned}$$

Then, the expression for the importance weights can be derived as

$$\begin{aligned} w_k^{(m)} &= \frac{f(S_{\tau_k} \mid V_{\tau_k}^{(m)}, \Theta) f(V_{\tau_k}^{(m)} \mid V_{\tau_{k-1}}^{(m)}, \Theta)}{g(V_{\tau_k}^{(m)} \mid S_{\tau_k}, V_{\tau_{k-1}}^{(m)})} \\ &= \frac{f(S_{\tau_k} \mid V_{\tau_k}^{(m)}, \Theta) \delta S_{\tau_k} f(V_{\tau_k}^{(m)} \mid V_{\tau_{k-1}}^{(m)}, \Theta)}{\phi(\nu_k^{(m)}) \Phi(d_{\tau_k}^{*(m)}) e^{\delta\nu_k^{(m)}}} \\ &= \frac{f(V_{\tau_k}^{(m)} \mid V_{\tau_{k-1}}^{(m)}, \Theta)}{\Phi(d_{\tau_k}^{*(m)}) e^{\delta\nu_k^{(m)}}} \end{aligned}$$

Table 1 shows the efficient sample sizes for the bootstrap filter and the localized sampler for different values of the measurement noise standard deviation, δ. The values are averages taken through time and across 20

simulations, run at different random seeds. The sample size $M = 1000$ and all the other simulation parameters are as described before. The MATLAB file producing the table is `test_MertonLocalized.m`. Overall, the localized sampler seems to perform much better than the bootstrap filter reflected in the much higher effective sample sizes. Further, as δ decreases, the performance of the bootstrap filter deteriorates while that of the localized filter actually gets better. The reason for this phenomenon is that for smaller values of δ, the relative importance of the new observation is higher in determining the location of the new unobserved asset value. Then, the the localized sampler that ignores the past overperforms the bootstrap filter that ignores the new observation.

Table 1. Effective Sample Size for the Localized Sampler and the Bootstrap Filter in Merton's model

	$\delta = 0.0005$	$\delta = 0.005$	$\delta = 0.01$	$\delta = 0.02$
N_{eff}(Localized)	999.9	993.0	974.1	916.9
N_{eff}(Bootstrap)	6.4	61.4	121.1	230.4

Using local linearization to generate the proposal in Merton's model

The localized sampler described in the previous section completely ignores the past. An alternative approach is to follow the advice of [15] and use a local linear approximation of the model to generate a proposal. Here, both the past and the new observation is used to come up with a proposal distribution at the price of the bias due to the linearization. In Merton' model, the only non-linearity is in the measurement equation (7). Linearizing this equation around the conditional expected value yields the approximate measurement equation:

$$\log S_{\tau_k} \sim A\left(\log V^{*(m)}\right) + B\left(\log V^{*(m)}\right) \times \left(\log V_{\tau_k} - \log V^{*(m)}\right)$$

where

$$\begin{aligned}
\log V^{*(m)} &= (\mu - \frac{\sigma^2}{2})h + \log V^{(m)}_{\tau_{k-1}} \\
A\left(\log V^{*(m)}\right) &= \log S\left(e^{\log V^{*(m)}}; \sigma, F, r, T - \tau_k\right) \\
B\left(\log V^{*(m)}\right) &= \frac{V^* \Phi(d^{*(m)})}{S\left(e^{\log V^{*(m)}}; \sigma, F, r, T - \tau_k\right)}
\end{aligned}$$

By local normality of this system, the conditional distribution of $\log V^{(m)}_{\tau_k}$ given $\log S_{\tau_k}$ is

$$\log V_{\tau_k}^{(m)} \sim N\left(\mu(\log V_{\tau_{k-1}}^{(m)}), \sigma^2(\log V_{\tau_{k-1}}^{(m)})\right)$$

where

$$\mu(\log V_{\tau_{k-1}}^{(m)}) = \log V^{*(m)} + \frac{B\left(\log V^{*(m)}\right)\sigma^2 h}{B^2\left(\log V^{*(m)}\right)\sigma^2 h + \delta^2} \times \left(\log S_{\tau_k} - A\left(\log V^{*(m)}\right)\right)$$

$$\sigma^2(\log V_{\tau_{k-1}}^{(m)}) = \sigma^2 h - \frac{B^2\left(\log V^{*(m)}\right)\sigma^4 h^2}{B^2\left(\log V^{*(m)}\right)\sigma^2 h + \delta^2}$$

The expression of the importance weights is

$$\begin{aligned} w_k^{(m)} &= \frac{f(S_{\tau_k} \mid V_{\tau_k}^{(m)}, \Theta) f(V_{\tau_k}^{(m)} \mid V_{\tau_{k-1}}^{(m)}, \Theta)}{g(V_{\tau_k}^{(m)} \mid S_{\tau_k}, V_{\tau_{k-1}}^{(m)})} \\ &= \frac{f(S_{\tau_k} \mid V_{\tau_k}^{(m)}, \Theta) f(V_{\tau_k}^{(m)} \mid V_{\tau_{k-1}}^{(m)}, \Theta)}{\phi\left(\frac{\log V_{\tau_k}^{(m)} - \mu(\log V_{\tau_{k-1}}^{(m)})}{\sigma(\log V_{\tau_{k-1}}^{(m)})}\right) / \sigma(\log V_{\tau_{k-1}}^{(m)})} \end{aligned}$$

Table 2 compares this linearized proposal with the bootstrap filter. The MATLAB file producing the table is `test_MertonLinearized.m`. The linearized filter performs much better, with results that are comparable to the localized sampler described before. Instead of using local linearization, [37] suggests using the unscented Kalman Filter for proposal generation.

Table 2. Effective Sample Size for the Localized Sampler and the Bootstrap Filter in Merton's model

	$\delta = 0.0005$	$\delta = 0.005$	$\delta = 0.01$	$\delta = 0.02$
N_{eff}(Linearized)	607.5	966.7	979.1	955.0
N_{eff}(Bootstrap)	6.4	61.4	121.1	230.4

Using adaptation to tune the proposal in Merton's model

Another generic approach to improve the efficiency of the particle filtering algorithm is to use the filter output to adapt the proposal. [8] suggests to implement this idea by choosing a parametric family of proposal distribution and then optimize the parameters using the particles from the filter. To illustrate this method, consider the adaptation of the bootstrap filter in Merton's model. In particular, assume that the following family of proposal distribution is chosen:

$$\log V_{\tau_k}^{(m)} \sim N\left((\mu - \frac{\sigma^2}{2})h + \log V_{\tau_{k-1}}^{(m)} + \gamma_{1,k}, \sigma^2 h \gamma_{2,k}\right) \tag{14}$$

Setting $\gamma_{1,k} = 0$ and $\gamma_{2,k} = 1$ one obtains bootstrap filter. In general $\gamma_{1,k}$ and $\gamma_{2,k}$ can be varied in order to find a proposal that is as close as possible to the target distribution, $f(\log V_{\tau_k}, \log V_{\tau_{k-1}} \mid \mathcal{D}_k)$. One appropriate metric to measure closeness between probability distributions is the Kullback-Leibler (K-L) distance. In the present context, if $r(\log V_{\tau_k} \mid \gamma, \log V_{\tau_{k-1}})$ is the parametric proposal conditional on $\log V_{\tau_{k-1}}$, then the overall proposal over the pair $(\log V_{\tau_k}, \log V_{\tau_{k-1}})$ is $r(\log V_{\tau_k} \mid \gamma, \log V_{\tau_{k-1}}) f(\log V_{\tau_{k-1}} \mid \mathcal{D}_{k-1})$. The K-L distance of this proposal from the target is defined as

$$
\begin{aligned}
&D_{KL}\left(f(\log V_{\tau_k}, \log V_{\tau_{k-1}} \mid \mathcal{D}_k) \| r(\log V_{\tau_k} \mid \gamma, \log V_{\tau_{k-1}}) f(\log V_{\tau_{k-1}} \mid \mathcal{D}_{k-1})\right) \\
&= \int_{\{\log V_{\tau_k}, \log V_{\tau_{k-1}}\}} f(\log V_{\tau_k}, \log V_{\tau_{k-1}} \mid \mathcal{D}_k) \log\left(\frac{f(\log V_{\tau_k}, \log V_{\tau_{k-1}} \mid \mathcal{D}_k)}{r(\log V_{\tau_k} \mid \gamma, \log V_{\tau_{k-1}}) f(\log V_{\tau_{k-1}} \mid \mathcal{D}_{k-1})}\right)
\end{aligned}
$$

Then, the "best" proposal within the parametric family is the one that minimizes the K-L distance to $f(\log V_{\tau_k}, \log V_{\tau_{k-1}} \mid \mathcal{D}_k)$. This is achieved by γ_k^* solving

$$
\gamma_k^* = arg\max_{\gamma} \int_{\{\log V_{\tau_k}, \log V_{\tau_{k-1}}\}} f(\log V_{\tau_k}, \log V_{\tau_{k-1}} \mid \mathcal{D}_k) \log r(\log V_{\tau_k} \mid \gamma, \log V_{\tau_{k-1}}) \tag{15}
$$

This optimization problem is unfeasible as the integral is not known in closed form. However, if one has a normalized weighted sample $(\pi_k^{(m)}, \log V_{\tau_k}^{(m)}, \log V_{\tau_{k-1}}^{(m)}, m = 1, ..., M)$ representing $f(\log V_{\tau_k}, \log V_{\tau_{k-1}} \mid \mathcal{D}_k)$ from a prior run of a particle filter, the problem can be approximated by

$$
\gamma_k^* = arg\max_{\gamma} \sum_{i=1}^{M} \pi_k^{(m)} \log r(\log V_{\tau_k}^{(m)} \mid \gamma, \log V_{\tau_{k-1}}^{(m)}) \tag{16}
$$

In the example in Merton's model with the choice of the proposal family as in (14) , the optimization problem becomes

$$
(\gamma_{1,k}^*, \gamma_{2,k}^*) = arg\max_{\gamma_{1,k}, \gamma_{2,k}} \sum_{i=1}^{M} \pi_k^{(m)} \left(-\frac{(\log V_{\tau_k}^{(m)} - (\mu - \frac{\sigma^2}{2})h - \log V_{\tau_{k-1}}^{(m)} - \gamma_{1,k})^2}{2\gamma_{1,k}, \sigma^2 h \gamma_{2,k}} - \frac{\log(\gamma_{1,k}, \sigma^2 h \gamma_{2,k})}{2}\right)
$$

This can be solved in one step, yielding

$$
\gamma_{1,k}^* = \sum_{i=1}^{M} \pi_k^{(m)} \left(\log V_{\tau_k}^{(m)} - (\mu - \frac{\sigma^2}{2})h - \log V_{\tau_{k-1}}^{(m)}\right) \tag{17}
$$

$$
\gamma_{2,k}^* = \frac{\sum_{i=1}^{M} \pi_k^{(m)} \left(\log V_{\tau_k}^{(m)} - (\mu - \frac{\sigma^2}{2})h - \log V_{\tau_{k-1}}^{(m)} - \gamma_{1,k}^*\right)^2}{\sigma^2 h} \tag{18}
$$

The algorithm is initialized by running the bootstrap filter (setting $(\gamma_{1,k}^{(0)} = 0, \gamma_{2,k}^{(0)} = 1)$) and then the filter is adapted by the procedure described above.

Adapted Bootstrap Filter in Merton's model

- Initial State: Set $V_{\tau_0}^{(m)} = S^{-1}(S_{\tau_0})$ where the function $S^{-1}(.)$ is the inverse of the equity pricing function in equation (6).
- Run the bootstrap filter, providing $(\gamma_{1,k}^{(1)}, \gamma_{2,k}^{(1)})$ using (17-18)
- Adapt the filter: For $j = 1, ..., N_{iter}$
  - Recursion: For $k = 1, ..., N$
    1. Sampling
       - Draw $\log V_{\tau_k}^{(m)}$ from $r(V_{\tau_k}^{(m)} \mid \gamma_{1,k}^{(j)}, \gamma_{2,k}^{(j)})$

$$\log V_{\tau_k}^{(m)} \sim N\left((\mu - \frac{\sigma^2}{2})h + \log V_{\tau_{k-1}}^{(m)} + \gamma_{1,k}^{(j)}, \sigma^2 h \gamma_{2,k}^{(j)}\right)$$

       - Compute the normalized importance weights

$$\pi_k^{(m)} = \frac{w_k^{(m)}}{\sum_{m=1}^{M} w_k^{(m)}}$$

       where

$$w_k^{(m)} = \frac{f(S_{\tau_k} \mid V_{\tau_k}^{(m)}, \Theta) f(V_{\tau_k}^{(m)} \mid V_{\tau_{k-1}}^{(m)}, \Theta)}{r(V_{\tau_k}^{(m)} \mid \gamma_{1,k}^{(j)}, \gamma_{2,k}^{(j)}, V_{\tau_{k-1}}^{(m)})}$$

    2. Compute the new value of the adaptation parameters: $(\gamma_{1,k}^{(j+1)}, \gamma_{2,k}^{(j+1)})$ using the new weighted sample and (17-18). To avoid spurious results due to a poor particle set, $\gamma_{2,k}$ is updated only when $N_{eff}(k) >= 5$.
    3. Resample from the weighted sample $\{(V_{\tau_k}^{(m)}, \pi_k^{(m)}); m = 1, \cdots, M\}$ to obtain a new equal-weight sample of size M.

As $M \to \infty$, the approximating optimization problem in (16) converges to the true problem in (15). Thus if M is large enough, setting $N_{iter} = 1$ would already achieve the optimal parameters. However for finite M, the initial particle approximation may be poor and running a couple more iterations can yield further improvements.

Table 3 reports the results of this algorithm with $N_{iter} = 4$ and $M = 1000$, with all the other simulation parameters set as in the examples before. The MATLAB file producing the table is `test_MertonAdapt.m`. Adaptation yields great improvements in the algorithm, providing acceptable results even when δ is small and the likelihood function is very peaked. In accordance with theory, most of the improvement takes place in the first iteration. Substantial further improvements are achieved only when the initial sampler is very poor, the case of small δ. In more complicated problems, wider parametric families could be used for adaptation. In

particular, using the adaptive D-kernel method of [4] and [5] would allow the use of general mixture classes.

Table 3. Effective Sample Size for the Adapted Bootstrap Filter in Merton's model

	$\delta = 0.0005$	$\delta = 0.005$	$\delta = 0.01$	$\delta = 0.02$
N_{eff}(Iteration 0)	6.4	61.4	121.1	230.4
N_{eff}(Iteration 1)	252.6	520.3	537.4	557.8
N_{eff}(Iteration 2)	457.4	542.0	546.5	557.9
N_{eff}(Iteration 3)	506.7	543.54	545.8	559.7
N_{eff}(Iteration 4)	523.9	544.3	547.51	557.6

5.2 Other variations on the filtering algorithm

When the future observation is very informative on the present state, it may be better to resample the present particles before propagating them forward. This idea is used in the Auxiliary Particle Filter by [33] and investigated theoretically in [13]. More sophisticated resampling routines have been proposed to reduce the variance of multinomial resampling. Some examples are residual resampling [28] or stratified resampling [27].

5.3 Application of Particle Filtering: Obtaining the asset and debt value of Six Flags

In the second example described before the objective is to obtain the unobserved asset value of Six Flags in 2008 using the noisy time series of equity. The application of Merton's model necessitates some assumptions on the inputs of the model. The face value of debt is chosen to be the sum of total liabilities and preferred equity (as this latter is more senior than common equity) yielding $F = 3197.9$ (the unit is Million USD). The maturity of debt is chosen to be 1 years, while the risk-free rate is set to 2.7%, the one-year zero-coupon yield on treasuries at the beginning of 2008. Last, to run the filter one needs estimates of the model parameters (μ, σ, δ). The estimation of the drift is unreliable using one year of data, so the drift is simply set equal to the riskfree rate. The other two parameters σ and δ are estimated in a Bayesian procedure using importance sampling and a flat prior. The posterior means are used as point estimates yielding $\sigma = 0.075$ and $\delta = 0.0117$. Panel A of Figure 7 reports the filtered asset values while Panel B the filtered yield spread on the debt (liabilities+preferred equity) of the firm. The localized filter with $M = 1000$ particles was used to produce the results. One can see that in the second half of 2008 when the asset value of the company decreased, the spread becomes more sensitive to changes in

the asset value. This can be explained by the fact that by this stage, the equity buffer that protects the debt-holders is more depleted. To understand the uncertainty created by the noise in the equity prices, Figure 8 plots the 90% confidence interval of the yield spread of Six Value, ranging from 7 basis points at the beginning of the year to 12 basis points at the end of the period. The figures have been produced using the MATLAB script `SixFlags.m`.

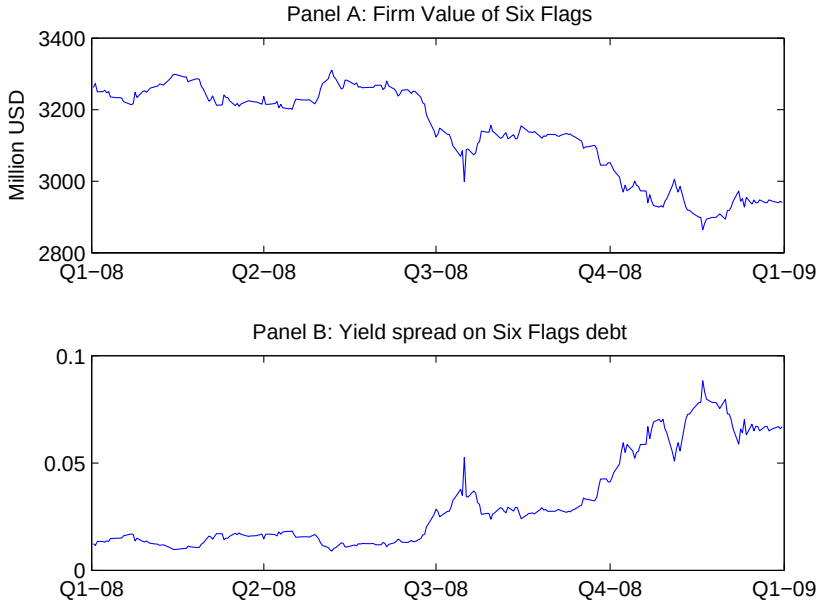

Fig. 7. Filtered asset value and yield spread on Six Flags' debt

6 Outlook

While the algorithms described in this chapter provide reliable sequential inference on the unobservable dynamic states, the important task of estimating the fixed parameters of the financial model has proved to be a formidable task.
In a classical setting, the problem stems from the irregularity of the likelihood surface. The individual likelihood function, $f(y_k \mid y_{1:k-1}, \theta)$, can be estimated pointwise using the particle filter as

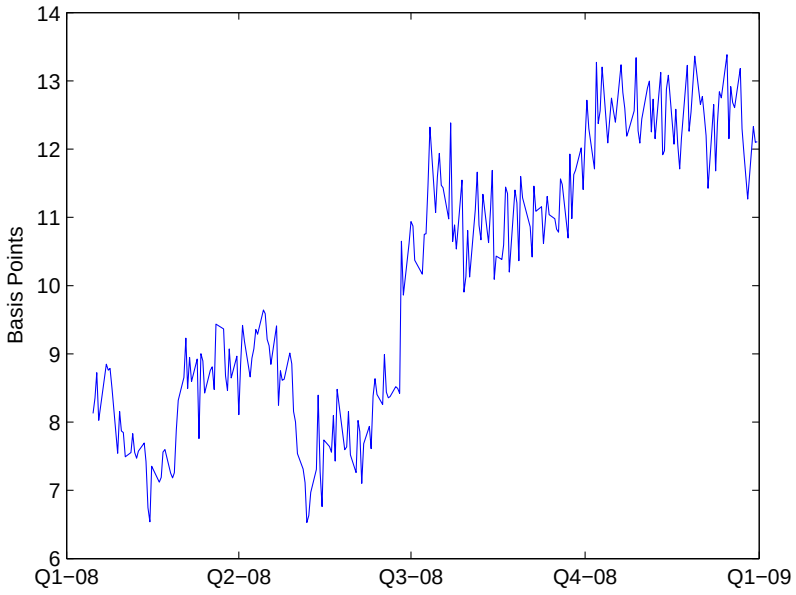

Fig. 8. 90% confidence interval on the Yield Spread of Six Flag

$$\hat{f}(y_k \mid y_{1:k-1}, \theta) \approx \sum_{i=1}^{M} w_k^{(m)}(\theta)$$

yielding an estimate of the sample :

$$\hat{l}(y_{1:N} \mid \theta) = \sum_{k=1}^{N} \log \hat{f}(y_k \mid y_{1:k-1}, \theta)$$

However, $\hat{l}(y_{1:N} \mid \theta)$ is an inherently irregular function of the fixed model parameters, θ. Figure 9 illustrates this phenomenon by plotting the estimated likelihood function of a simulated data sample in Merton's model for different values of the asset volatility parameter, σ. The local wiggles one observes here result from the resampling step and make both the usual gradient-based optimization routines unusable and inference based on the numerical derivatives of the likelihood function problematic.
There are several ways in the literature to circumvent this problem. [32] proposes to use a smooth resampling routine that makes the likelihood function regular. [19] apply the method to estimate the parameters of Merton's model with noisy equity prices, while [7] use it in fitting equity option prices with different stochastic volatility models. Unfortunately, the approach only works when the hidden state is one-dimensional.

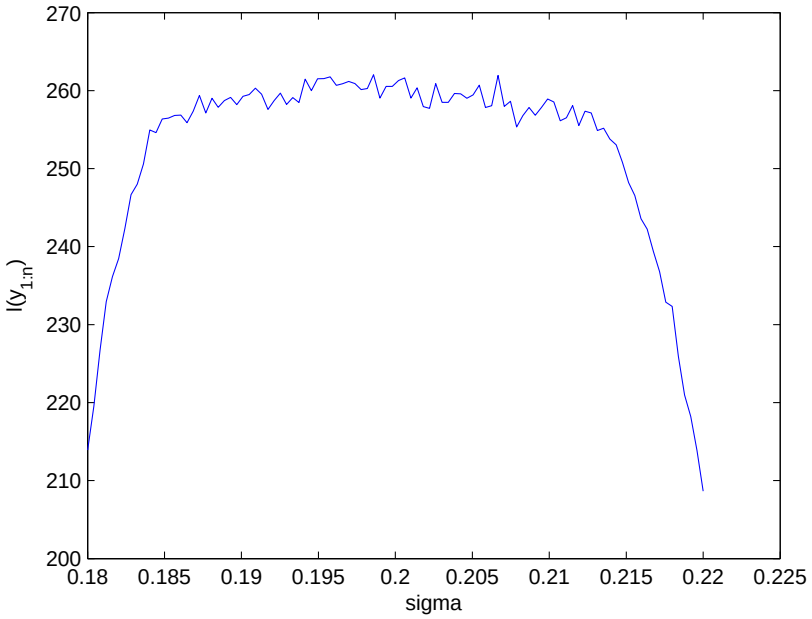

Fig. 9. Irregularity of the Likelihood Function in Merton's model

An alternative approach that works even when x_k is multi-dimensional is the Monte-Carlo Expectation-Maximization (MCEM) algorithm. Here the irregularity of the filter becomes inconsequential for obtaining parameter estimates, because filtering and optimization are disentangled. In particular, while the particle filter is used to approximate the necessary expectations in the E-step, the particles are kept unchanged in the M-step where optimization is implemented. Further, [31] show that it is sufficient to use fixed-lag smoothing in the E-step with a relatively small number of lags. This is important because in particle filtering the inference on the recent past is more reliable than the representation of the distant past, a result of the repeated use of the resampling step. To deal with the problem of inference, [18] proposes the use the sample cross-products of the individual smoothed scores and a Newey-West correction. [17] apply the MCEM algorithm to the estimation of a jump-diffusion model with high-frequency data and microstructure noise, while [23] uses it to estimate intra-daily patterns of transaction costs and volatilities on the credit default swap market.
In a Bayesian setting, one could simply try to include the fixed parameters in the state-space and perform particle filtering on the extended state-space. This is very attractive as it would allow joint sequential inference on the states and the fixed parameters. Unfortunately it is well-known that this

algorithm is unreliable. The underlying reason is that the extended dynamic system is not forgetting its past due to the inclusion of the fixed parameters, thus the Monte-Carlo errors committed in each stage quickly accumulate. Extending the work of [35], [26] suggest tracking the filtering distribution of some sufficient statistics to perform sequential inference on the parameters. [25] apply this approach to examine the predictability of the stock market and optimal portfolio allocation. The key limitation is that the method can only be applied to models that admit a finite-dimensional sufficient statistic structure for the fixed parameters. Instead of attempting sequential inference, [2] suggest inserting particle filters into an MCMC algorithm as a proposal-generating mechanism. The method is illustrated on different economic and financial models by [22].

References

1. Anderson BDO, Moore JB (1979) Optimal filtering. Prentice-Hall, Englewood Cliffs, N.J.
2. Andrieu C, Doucet A, Holenstein A (2010) Particle markov chain monte carlo. Journal of Royal Statistical Society B 72:1-33
3. Bakshi G, Carr P, Wu L (2008) Stochastic risk premiums, stochastic skewness in currency options, and stochastic discount factors in international economies. Journal of Financial Economics 87:132–156
4. Cappe O, Douc R, Gullin A, Marin JM, Robert CP (2007) Convergence of adaptive mixtures of importance sampling schemes. Annals of Statistics 35:420–448
5. Cappe O, Douc R, Gullin A, Marin JM, Robert CP (2008) Adaptive importance sampling in general mixture classes. forthcoming Statistics and Computing
6. Christensen JHE, Diebold FX, Rudebusch GD (2007) The affine arbitrage-free class of Nelson-Siegel term structure models. NBER Working Paper No. 13611
7. Christoffersen PF, Jacobs K, Mimouni K (2008) Models for S&P 500 dynamics: Evidence from realized volatility, daily returns, and option prices. Manuscript, McGill University
8. Cornebise J, Moulines E, Olsson J (2008) Adaptive methods for sequential importance sampling with application to state space models. forthcoming Statistics and Computing
9. Crisan D, Doucet A (2002) A survey of convergence results on particle filtering methods for practitioners. IEEE Transactions on Signal Processing 50:736–746, 2002
10. De Jong F (2000) Time series and cross-section information in affine term-structure models. Journal of Business & Economic Statistics 18:300–314
11. Del Moral P (2004)Feynman-Kac formulae genealogical and interacting particle systems with applications. Springer, New York
12. Diebold FX, Li C (2006) Forecasting the term structure of government bond yields. Journal of Econometrics 130:337364
13. Doucet A, Johansen AM (2008) A note on auxiliary particle filters. forthcoming in Statistics and Probability Letters

14. Doucet A, De Freitas N, Gordon N (eds) (2001) Sequential Monte Carlo Methods in Practice. Springer-Verlag
15. Doucet A, Godsill SJ, Andrieu C (2000) On sequential monte carlo sampling methods for bayesian filtering. Statistics and Computing 10:197–208
16. Duan JC (1994) Maximum likelihood estimation using price data of the derivative contract. Mathematical Finance 4:155–167
17. Duan JC, Fulop A (2007) How frequently does the stock price jump? – an analysis of high-frequency data with microstructure noises. Working Paper
18. Duan JC, Fulop A (2009) A stable estimator for the information matrix under EM. forthcoming Statistics and Computing
19. Duan JC, Fulop A (2009) Estimating the structural credit risk model when equity prices are contaminated by trading noises. Journal of Econometrics 150:288-296
20. Duan JC, Simonato JG (1999) Estimating and testing exponential-affine term structure models by Kalman filter. Review of Quantitative Finance and Accounting 13:111–35
21. Duffee GE (2002) Term premia and interest rate forecasts in affine models. Journal of Finance 57:405–443
22. Flury T, and Shephard N (2008) Bayesian inference based only on simulated likelihood: Particle filter analysis of dynamic economic models. Technical report, Nuffield College, Oxford University
23. Fulop A, Lescourret L (2008) Intra-daily variations in volatility and transaction costs in the Credit Default Swap market. Working Paper, 2009
24. Gordon NJ, Salmond DJ, Smith AFM (1993) A novel approach to non-linear and non-gaussian bayesian state estimation. IEEE Proceedings F 140:107–113
25. Johannes M, Korteweg AG, Polson N (2008) Sequential learning, predictive regressions, and optimal portfolio returns. Working Paper
26. Johannes M, Polson N (2006) Exact particle filtering and parameter learning. Working Paper
27. Kitagawa G (1996) Monte carlo filter and smoother for non-gaussian nonlinear state space models. Journal of Computational and Graphical Statistics 5:1–25
28. Liu JS (2001) Monte Carlo Strategies in Scientific Computing. Springer, New York
29. Merton RC (1974) On the pricing of corporate debt: The risk structure of interest rates. Journal of Finance 29:449–470
30. Nelson, CR, Siegel, AF (1987) Parsimonious modeling of yield curve. Journal of Business 60:473489
31. Olsson J, Cappe R, Douc R, Moulines E (2008) Sequential monte carlo smoothing with application to parameter estimation in non-linear state space models. Bernoulli 14:155–179
32. Pitt M (2002) Smooth particle filters likelihood evaluation and maximisation. Working Paper, University of Warwick
33. Pitt M , Shephard N (1999) Filtering via simulation: auxiliary particle filter. Journal of the American Statistical Association, 94:590–599
34. Schwartz E, Smith JE (2000) Short-term variations and long-term dynamics in commodity prices. Management Science 46:893–911
35. Storvik, G (2002) Particle filters in state space models with the presence of unknown static parameters. IEEE Transactions on Signal Processing 50:281-289.

36. Trolle AB, and Schwartz E (2008) Unspanned stochastic volatility and the pricing of commodity derivatives. Working Paper
37. van der Merwe R, Doucet A, De Freitas N, Wan E (2000) The unscented particle filter
38. van der Merwe R, Wan E (2000) The unscented kalman filter for nonlinear estimation. Proceedings of Symposium 2000 on Adaptive Systems for Signal Processing, Communication and Control