

Chapter 1

Introduction: Stochastic Filtering in Finance

This chapter introduces the background to the stochastic filtering problems as applied to various discipline area. This will point out different possible scenarios and solution approaches based on filtering for problems in finance and economics. This will also broadly outline the structures and models of a generalized stochastic filtering problem and the main equations that are needed. The introduction will also set the scene for various terminologies used in this context e.g. Unobserved Components (UC), Dynamic harmonic Regression (DHR), Forward Pass Kalman Filter, and Backward Pass Kalman Filter or Fixed Interval Smoothing (FIS). It will discuss further the extensions available for dealing with non-linear problems e.g. Extended Kalman Filter (EKF), and Unscented Kalman Filter (UKF).

Before ending introduction, it will give details of the contents of the following chapters and how these are organized etc. This relates to various application areas within the disciplines of finance and economics.

Some authors have identified three most successful areas of applications in quantitative finance. The first one relate to the unobserved volatility of any financial asset which is crucial to the development of derivatives products dependent on that asset. The second one concerns factor processes with or without jumps applied in commodities and interest rates. The third one concerns risk-minimizing hedging strategies under partial observations. Of course, there are several other areas of interests particularly in macro-economic issues.

1.1 Filtering Problem

The filtering problem deals with the estimation of a stochastic process X_t that is unobserved based on the past and current measurement of a related process $Y, \{Y_s : 0 \leq s \leq t\}$. The information coming out of the measurement process up to time t is conveniently represented by the σ -algebra, \mathcal{Y}_t generated by $\{Y_s : 0 \leq s \leq t\}$. Davis and Marcus (1981) give excellent introduction to the filtering problem. Besides, for a more complete and mathematical treatments see Lipster and Shiriyayev (1978). On the other hand, Jazwinski (1970) gives a more applied and practical analyses of the filtering issues.

The filtering problem gives the solution to the question of conditional density of the signal $X_t, p(x_t | y_t)$ given the filtration \mathcal{Y}_t . In this context, Chaleyat-Maurel and Michel (1984) prove that the general solution is infinite dimensional. This implies that a general filter cannot be implemented by an algorithm that only uses a finite number of characteristics. In other words, there is no finite memory computer implementation that can be adequate for the filter. However, the most referenced implementation of the filter algorithm, Kalman filter, is a special case for the linear system with Gaussian conditional distribution. In this case, only two parameters mean and the variance, capture the essential attributes of $p(x_t | y_t)$.

1.2 Examples of Filtering Applications

Filtering has long been used in Control Engineering and Signal Processing. For the linear case, the Kalman filter is fast and easy to apply despite the noisiness and the length of the input data. It is an iterative procedure and the model parameters may depend upon a large number of observable and unobservable data.

In finance and economics applications we normally have a temporal time series of observations (Y_t). For example, Wells (1996) uses stock price series, Babbs and Nowman (1999) as well as Pennacchi (1991) use interest rate series, Lautier (2000) and Lautier and Galli (2000) use futures price series. The unobserved time series (X_t) in these cases could be volatility, convenience yield, price of risk etc depending on the

objectives of the researchers. The iterative process allows us to link two consecutive unobservable states using a transition equation and a measurement equation relating to the hidden state. The idea is to proceed in two steps: first we estimate the hidden state, prior distribution, by using all the information up to that time-step. Then using this predicted value together with the new observation, we obtain a conditional estimation of the posterior distribution of the state.

It is useful at this stage to get used to some of the terminologies applicable to different filtering processes depending on the model characteristics. The linear filter, Kalman filters (KF), is applicable to cases when the transition and measurement equations are linear and the noise processes are Gaussian. With the same noise processes if the model equations are non-linear then we refer to this as the Extended Kalman Filter (EKF). Another more recent advance in this scenario is termed as the Unscented Kalman Filter (UKF). To deal with non-Gaussian cases we resort to Particle Filter (PF).

1.3 Linear Kalman Filter

In this section we describe the linear Gaussian filter and the algorithm that is needed for implementation. To assist with the exposition we will use a general specification of the model and will subsequently move on to more traditional specification more suitable for empirical studies.

Let the dynamic process X_t follows a transition equation

$$x_t = f(x_{t-1}, w_t), \quad (1.1)$$

and we also assume that we have a measurement Y_t such that

$$y_t = h(x_t, u_t). \quad (1.2)$$

In the above equations w_t and u_t are two mutually uncorrelated sequences of temporally uncorrelated sequences of Normal random variables with zero mean and covariance matrices Q_t and R_t respectively. Additionally, w_t is uncorrelated with x_{t-1} and u_t is uncorrelated with x_t .

We define the prior process estimate as

$$\hat{x}_{t|t-1} = E[x_t] \quad (1.3)$$

which is the estimate of x_t at time $t-1$ just prior to making the measurement at time t . Similarly, we define the posterior estimate as

$$\hat{x}_{t|t} = E[x_t | y_t] \quad (1.4)$$

which is the estimate at time t after the measurement at t has taken place. We also have the corresponding estimation errors $e_{t|t-1} = x_t - \hat{x}_{t|t-1}$ and $e_t = x_t - \hat{x}_{t|t}$. These give us the estimate of the error covariances as

$$P_{t|t-1} = E[e_{t|t-1} \cdot e'_{t|t-1}], \quad P_{t|t} = E[e_t \cdot e'_t]. \quad (1.5)$$

In order to compute the above mean and the covariance we need the corresponding conditional densities $p(x_t | y_{1:t-1})$ and $p(x_t | y_t)$. These are determined iteratively via transition and measurement updates. The basic idea is to define the probability density function corresponding to the hidden state x_t given all the measurements made up to that time i.e. $y_{1:t}$.

The transition step is based upon Chapman-Kolmogorov equation

$$\begin{aligned} p(x_t | y_{1:t-1}) &= \int p(x_t | x_{t-1}, y_{1:t-1}) p(x_{t-1} | y_{1:t-1}) dx_{t-1} \\ &= \int p(x_t | x_{t-1}) p(x_{t-1} | y_{1:t-1}) dx_{t-1} \end{aligned} \quad (1.6)$$

following the Markov property. The measurement update step is based upon Bayes rule

$$p(x_t | y_{1:t}) = \frac{p(y_t | x_t) p(x_t | y_{1:t-1})}{p(y_t | y_{1:t-1})} \quad (1.7)$$

and $p(y_t | y_{1:t-1}) = \int p(y_t | x_t) p(x_t | y_{1:t-1}) dx_t$. Additional details of this may be found in Jazwinski (1970).

At this point it is instructive to specialize the transition and the measurement equations (1.1) and (1.2) for a linear system and state the updating equations in a form amenable for easier implementation.

Let us focus on a linear state space system with transition equation of the form

$$x_t = T_t x_{t-1} + c_t + w_t \quad (1.8)$$

and the measurement equation

$$y_t = Z_t x_t + d_t + u_t \quad (1.9)$$

where c_t and d_t are possible time dependent vectors of compatible dimensions. Similarly the matrices T and Z_t are of dimensions compatible with the length of the state vector x_t and the measurement vector y_t respectively.

The actual algorithm may be stated as follows

Initialize x_0 and P_0

Transition equations updates for mean and covariance

$$x_{t|t-1} = T_t x_{t-1} + c_t \quad (1.10)$$

$$P_{t|t-1} = T_t P_{t-1} T_t' + Q_t \quad (1.11)$$

We define innovation as

$$v_t = y_t - Z_t x_t - d_t \quad (1.12)$$

Measurement equation update

$$x_{t|t} = x_{t|t-1} + K_t v_t \quad (1.13)$$

$$P_{t|t} = (I - K_t Z_t) P_{t|t-1} \quad (1.14)$$

$$K_t = P_{t|t-1} Z_t' (Z_t P_{t|t-1} Z_t' + R_t)^{-1} \quad (1.15)$$

and I is the identity matrix.

The matrix K_t is the Kalman gain and corresponds to the mean of the conditional distribution of x_t after making the measurement of y_t . It is the quantity that would minimize the mean square error P_t within the class of linear estimators.

This interpretation may be broadened further. Let us consider x as a Normally distributed variable with mean μ_x variance σ_{xx} and z another Normally distributed random variable with mean μ_z and variance σ_{zz} . The covariance between them is $\sigma_{xz} = \sigma_{zx}$. The conditional distribution of $x|z$ is also Normal with $\mu_{x|z} = \mu_x + K(z - \mu_z)$ with Kalman gain corresponding to $K = \sigma_{xz} \sigma_{zz}^{-1}$.

In empirical applications most models would be parameterized and such parameters need to be calibrated to the available time series data. This calibration can be carried out by maximum likelihood method for

conditionally Gaussian models. Let us label the unknown parameters by Θ .

We need to maximize the joint probability density $\prod_1^T p(y_t | y_{1:t-1})$. Since the conditional density is Normal and taking logarithm of the precise form of the function and ignoring the constant term the target likelihood function to be maximized is

$$L(\Theta) = -\sum_1^T \ln(|\Sigma_t|) - \sum_1^T v_t' \Sigma_t^{-1} v_t \quad (1.16)$$

and $\Sigma_t = Z_t P_{t|t-1} Z_t' + R_t$.

1.4 Extended Kalman Filter (EKF)

In this section we discuss how to apply filtering algorithms when the assumption of Gaussian distribution still holds but the state equation and the measurement equation may be non-linear. It is easier to relate this discussion by focusing on the model given by equations (1.1) and (1.2). In the previous section the linear version used the model definition given by the equations (1.8) and (1.9). This section has benefited from Javaheri, Lautier, and Galli (2005).

The actual form of non-linearity given by (1.1) and (1.2) is not important to develop the ideas of Extended Kalman Filter (EKF). The functions in these equations need to be linearized and use the notions of conservation of Normal distribution property within the class of linear functions. We, thus, define the Jacobian matrices of the function $f(\cdot)$ with respect to the state variables and the system noise element and we refer to these as A_t and W_t . Precisely, for every row i and column j ,

$$A_{ij} = \frac{\partial f_i(x_{t-1|t-1}, 0)}{\partial x_j}, \quad W_{ij} = \frac{\partial f_i(x_{t-1|t-1}, 0)}{\partial w_j}. \quad (1.17)$$

Similarly, the Jacobian matrices for the measurement equation may be given by:

$$H_{ij} = \frac{\partial h_i(x_{t|t-1}, 0)}{\partial x_j}, \quad U_{ij} = \frac{\partial h_i(x_{t|t-1}, 0)}{\partial u_j}. \quad (1.18)$$

These Jacobian matrices could be related to the system matrices of the linear system given by (1.8) and (1.9).

The transition and the updating equation of the previous section may now be expressed via these Jacobian matrices.

Transition equations updates for mean and covariance

$$x_{t|t-1} = f(x_{t-1|t-1}, 0) \quad (1.19)$$

$$P_{t|t-1} = A_t P_{t-1} A_t' + W_t Q_{t-1} W_t' \quad (1.20)$$

We define innovation as

$$v_t = y_t - h(x_{t|t-1}, 0) \quad (1.21)$$

Measurement equation updates are

$$x_{t|t} = x_{t|t-1} + K_t v_t \quad (1.22)$$

$$P_{t|t} = (I - K_t H_t) P_{t|t-1} \quad (1.23)$$

$$K_t = P_{t|t-1} H_t' (H_t P_{t|t-1} H_t' + U_t R_t U_t')^{-1} \quad (1.24)$$

where I is the identity matrix.

Once this system of recursive equations are in place, the unknown parameters of the model can be calibrated using likelihood function maximization as outlined in the previous section. It may, however, be more instructive to follow an example due to Théoret, Rostan and El Moussadek (2004).

1.5 Applying EKF to Interest Rate Model

This illustration is based on the interest rate model due to Fong and Vasicek (1992). This two factor short rate model introduces the stochastic variance specification of Ornstein-Uhlenbeck form to the original Vasicek short rate model. This is described by

$$\begin{aligned} dr_t &= \kappa(\mu - r_t)dt + \sqrt{v_t}dW_{r,t} \\ dv_t &= \lambda(\bar{v} - v_t)dt + \tau\sqrt{v_t}dW_{v,t} \end{aligned} \quad (1.25)$$

where $E(dW_{r,t}, dW_{v,t}) = \rho$, i.e. the two noise processes $W_{r,t}$ and $W_{v,t}$ are two correlated Brownian motions associated with the short rate and the short rate volatility processes, respectively.

The model allows stationary mean reverting process for the short rate where the variance of the short rate is also a stationary stochastic process. Here μ is the unconditional average short rate, κ is the speed of mean reversion in the short rate. Similarly, \bar{v} is the average unconditional variance and λ defines the degree of persistence in the variance. Finally, τ is the variance of the unobserved variance process over a short period of time. The hidden volatility requires filtering technique to be extracted from observations on short rates.

The model in (1.25) is in continuous time as well as it is non-linear. Therefore, before applying EKF we need to transform it to discrete and approximate linear form.

The first of the two equations in (1.25) may be written in the differential notation as

$$d(e^{-\kappa t}(r_t - \mu)) = e^{\kappa t} \sqrt{v_t} dW_{r,t}. \quad (1.26)$$

Integrating by parts and assuming the sampling period being Δ this leads to,

$$r_{t+\Delta} = \mu + e^{-\kappa\Delta}(r_t - \mu) + e^{-\kappa\Delta} \int_t^{t+\Delta} e^{\kappa(s-t)} \sqrt{v_s} dW_{r,s}. \quad (1.27)$$

Similarly, for the variance equation,

$$v_{t+\Delta} = \bar{v} + e^{-\lambda\Delta}(v_t - \bar{v}) + e^{-\lambda\Delta} \tau \int_t^{t+\Delta} e^{\lambda(s-t)} \sqrt{v_s} dW_{v,s}. \quad (1.28)$$

If we now define

$$\varepsilon_t(\Delta) = e^{-\kappa\Delta} \int_t^{t+\Delta} e^{\kappa(s-t)} \sqrt{v_s} dW_{r,s} \quad (1.29)$$

and

$$\eta_t(\Delta) = e^{-\lambda\Delta} \int_t^{t+\Delta} e^{\lambda(s-t)} \sqrt{v_s} dW_{v,s} \quad (1.30)$$

then the equations (1.27) and (1.28) can be recast as the discrete form of the Fong and Vasicek model. Then,

$$r_{t+\Delta} = \mu + e^{-\kappa\Delta}(r_t - \mu) + \varepsilon_t(\Delta) \quad (1.31)$$

and

$$v_{t+\Delta} = \bar{v} + e^{-\lambda\Delta} (v_t - \bar{v}) + \eta_t(\Delta). \quad (1.32)$$

Now we focus on (1.29) and (1.30) and approximate the innovations as,

$$\varepsilon_{t\Delta}(\Delta) = e^{-\kappa\Delta} \sqrt{v_{t\Delta}} \sqrt{\Delta} \varepsilon_t, \quad \eta_{t\Delta}(\Delta) = e^{-\lambda\Delta} \tau \sqrt{v_{t\Delta}} \sqrt{\Delta} \eta_t \quad (1.33)$$

and $\varepsilon_t(\Delta)$ and $\eta_t(\Delta)$ are independent standard Normal random variables.

If we define a transformed discrete observed variable as,

$$R_t = e^{\kappa\Delta} (r_{(t+1)\Delta} - \mu) - (r_{t\Delta} - \mu), \quad t=0,1,2,\dots \quad (1.34)$$

and denote $V_t = v_{t\Delta}$ then the discrete state space form of the Fong and Vasicek model results,

$$\begin{aligned} R_t &= \sqrt{\Delta} \sqrt{V_t} \varepsilon_t, \quad t=0,1,2,\dots \\ V_t &= e^{-\lambda\Delta} V_{t-1} + (1 - e^{-\lambda\Delta}) \bar{v} + e^{-\lambda\Delta} \tau \sqrt{\Delta} \sqrt{V_{t-1}} \eta_t, \quad t=1,2,\dots \end{aligned} \quad (1.35)$$

The first of the two equations in (1.35) is the measurement equation, but it is not linear in state variable for Kalman filter to be applicable. With a suitable change of variable this measurement equation could be made linear as, with

$$y_t = \ln \left(\frac{R_t^2}{\Delta} \right) \quad (1.36)$$

the first of equation (1.35) becomes,

$$y_t = \ln V_t + \ln \varepsilon_t^2. \quad (1.37)$$

Thus, the non-linear form of the measurement equation (1.37) and the second of the equation (1.35) create the discrete linear state space version of the Fong and Vasicek model. Since the measurement equation error term in (1.37) is not strictly Normal, it may be approximated as such and the EKF is now applicable.

Additional details of this model and the empirical results may be found in Théoret, Rostan and El Moussadek (2004).

1.6 Unscented Kalman Filter (UKF) for Nonlinear Models

In this section we describe another more recent development in application of Kalman filter to nonlinear systems. This new approach is due to Julier and Uhlmann (1997) and the description in this section follows the article by Javaheri, Lautier, and Galli (2005).

In this approach Jacobians are not computed. Instead the actual nonlinear system is used to approximate the distributions of the state and the measurement variables with a Normal distribution by applying unscented transformation to it. This allows more accurate evaluation of the mean and the covariance compared to that of the EKF method. The UKF method requires augmentation of the state vector with the system and the measurement noises and this leads to expanded state space. This is not required if the system noise and the measurement noise are linearly additive in their respective equations. Let us define this augmented state dimension as $n_a = n_x + n_w + n_u$, where n_a , n_x , n_w , n_u represent the augmented state space, original state space, dimension of the state noise and the dimension of the measurement noise respectively.

The UKF algorithm may be summarized as follows:

First initialize the state vector, x_0 and the state covariance P_0

Second, we need to define two weights to be associated with the mean $\omega_i(m)$, and the covariance $\omega_i(c)$ of the state as,

$$\omega_0(m) = \frac{\lambda}{n_a + \lambda}, \quad \omega_0(c) = \frac{\lambda}{n_a + \lambda} + (1 - \alpha^2 + \beta) \quad (1.38)$$

and for all other elements,

$$\omega_i(m) = \omega_i(c) = \frac{1}{2(n_a + \lambda)}, \quad i = 1, 2, \dots, 2n_a \quad (1.39)$$

where the scaling parameters α , β , κ and $\lambda = \alpha^2(n_a + \kappa) - n_a$ may be tuned for the particular problem. With these setting in place, the algorithm proceeds for each t as:

Third, the augmented state vector is created by concatenating the original state vector, state noise and the measurement noise,

$$\mathbf{x}_{t-1}^a = \begin{bmatrix} \mathbf{x}_{t-1} \\ \mathbf{w}_{t-1} \\ \mathbf{u}_{t-1} \end{bmatrix}. \quad (1.40)$$

Thus the augmented state expectation after a measurement,

$$\mathbf{x}_{t-1|t-1}^a = \mathbb{E}[\mathbf{x}_{t-1}^a | y_{k-1}] = \begin{bmatrix} \mathbf{x}_{t-1|t-1} \\ 0 \\ 0 \end{bmatrix} \quad (1.41)$$

and similarly, the state covariance,

$$\mathbf{P}_{t-1|t-1}^a = \begin{bmatrix} \mathbf{P}_{t-1|t-1} & \mathbf{P}_{xw(t-1|t-1)} & 0 \\ \mathbf{P}_{xw(t-1|t-1)} & \mathbf{P}_{ww(t-1|t-1)} & 0 \\ 0 & 0 & \mathbf{P}_{uu(t-1|t-1)} \end{bmatrix} \quad (1.42)$$

where the subscripts x, w, and u correspond to the original state vector, state noise and the measurement noise parts respectively.

Fourth, to achieve the Normal approximation we need to define what is known as the Sigma Points for Unscented Transformation:

$$\boldsymbol{\chi}_{t-1}^a(0) = \mathbf{x}_{t-1|t-1}^a \quad (1.43)$$

and,

$$\boldsymbol{\chi}_{t-1}^a(i) = \mathbf{x}_{t-1|t-1}^a + \left((na + \lambda) \mathbf{P}_{t-1|t-1}^a \right)_i^{0.5}, \quad i = 1, \dots, na, \quad (1.44)$$

$$\boldsymbol{\chi}_{t-1}^a(i) = \mathbf{x}_{t-1|t-1}^a - \left((na + \lambda) \mathbf{P}_{t-1|t-1}^a \right)_{i-na}^{0.5}, \quad i = na + 1, \dots, 2na \quad (1.45)$$

where the subscripts i and (i - na) in equations (1.44) and (1.45) refer to the corresponding columns of the matrix in the square root.

Fifth, the augmented state equation updates may be stated as,

$$\boldsymbol{\chi}_{t|t-1}(i) = f(\boldsymbol{\chi}_{t-1}^x(i), \boldsymbol{\chi}_{t-1}^w(i)), \quad i = 0, \dots, 2na + 1 \quad (1.46)$$

$$\mathbf{x}_{t|t-1} = \sum_{i=0}^{2na} \boldsymbol{\omega}_i(m) \boldsymbol{\chi}_{t|t-1}(i) \quad (1.47)$$

$$\mathbf{P}_{t|t-1} = \sum_{i=0}^{2na} \boldsymbol{\omega}_i(c) (\boldsymbol{\chi}_{t|t-1}(i) - \mathbf{x}_{t|t-1}) (\boldsymbol{\chi}_{t|t-1}(i) - \mathbf{x}_{t|t-1})' \quad (1.48)$$

and the superscript x and w in the above equations represent the state and the state noise part of the augmented state vector, respectively.

Sixth, the innovation from the measurement equation may be written as,

$$y_{t|t-1}(i) = h(\chi_{t|t-1}(i), \chi_{t-1}^u(i)) \quad (1.49)$$

$$y_{t|t-1} = \sum_{i=0}^{2na} \omega_i(m) y_{t|t-1}(i) \quad (1.50)$$

and,

$$v_t = y_t - y_{t|t-1}. \quad (1.51)$$

Seventh, the measurement updates are,

$$P_{y_t, y_t} = \sum_{i=0}^{2na} \omega_i(c) (y_{t|t-1}(i) - y_{t|t-1}) (y_{t|t-1}(i) - y_{t|t-1})', \quad (1.52)$$

$$P_{\chi_t, y_t} = \sum_{i=0}^{2na} \omega_i(c) (\chi_{t|t-1}(i) - x_{t|t-1}) (y_{t|t-1}(i) - y_{t|t-1})', \quad (1.53)$$

where Kalman gain is given by,

$$K_t = P_{\chi_t, y_t} P_{y_t, y_t}^{-1}. \quad (1.54)$$

Finally, as in the linear system,

$$x_{t|t} = x_{t|t-1} + K_t v_t \quad (1.55)$$

and,

$$P_{t|t} = P_{t|t-1} - K_t P_{y_t, y_t} K_t'. \quad (1.56)$$

This completes all the recursions needed to develop the likelihood function that needs to be maximized in order to estimate the unknown parameters in the model.

In a later chapter we will examine the application of UKF to a practical problem.

1.7 Background to Particle Filter for Non Gaussian Problems

In this section we discuss the basic idea behind particle filter and the next section will develop the filtering algorithm utilizing this concept. Instead of using a Gaussian approximation to infer $(x_t | y_t)$ it employs Monte Carlo Simulation to achieve that. This requires deciding on a proposal distribution, $q(\cdot)$, which is simple and importance sampling as described below.

In term of the proposal distribution the expected value will be given by,

$$\begin{aligned} E[f(x_t)] &= \int f(x_t) p(x_t | y_{1:t}) dx_t \\ &= \int f(x_t) \frac{p(x_t | y_{1:t})}{q(x_t | y_{1:t})} q(x_t | y_{1:t}) dx_t. \end{aligned} \quad (1.57)$$

This may be re-expressed as,

$$\begin{aligned} E[f(x_t)] &= \int f(x_t) \frac{w_t(x_t)}{p(x_t | y_{1:t})} q(x_t | y_{1:t}) dx_t, \\ w_t(x_t) &= \frac{p(x_t | y_{1:t}) p(x_t)}{q(x_t | y_{1:t})}. \end{aligned} \quad (1.58)$$

This is referred to as the non-normalized weight during filtering at step t .

We can convert this to normalized weight as,

$$\begin{aligned} E[f(x_t)] &= \frac{E_q[w_t(x_t) f(x_t)]}{E_q[w_t(x_t)]} \equiv E_q[\tilde{w}_t(x_t) f(x_t)] \\ \tilde{w}_t(x_t) &= \frac{w_t(x_t)}{E_q(w_t(x_t))}. \end{aligned} \quad (1.59)$$

We can now apply Monte Carlo Simulation (with N sampling points) from the distribution of $q(x_t | y_{1:t})$ in discrete framework to compute the above expectation,

$$E[f(x_t)] \approx \sum_{i=1}^N \tilde{w}_t(x_t^i) f(x_t^i), \quad \tilde{w}_t(x_t^i) = \frac{w_t(x_t^i)}{\sum_{i=1}^N w_t(x_t^i)}. \quad (1.60)$$

We now need a mechanism to recursively compute this weight from one time step to the next. This is achieved by assuming that the proposal distribution is Markov. This recursion is,

$$w_t^i = w_{t-1}^i \frac{p(y_t | x_t^i) p(x_t^i | x_{t-1}^i)}{q(x_t^i | x_{t-1}^i, y_{1:t})}. \quad (1.61)$$

This now completely defines the recursion for the importance sampling procedure and it solely relies on filtering and not smoothing as may be applicable to Kalman filter.

We will conclude this section with some observations on practical issues related to implementation of this re-sampling procedure. The variance of the weights calculated above may vary widely and thus several suggestions have been proposed in the literature. Some of these could be found in Javaheri, Lautier, and Galli (2005). One such suggestion is to use Gaussian approximation for the proposal distribution based on the Kalman filter discussed earlier. This has the added advantage of utilizing the most current observation in the observed series. In term of earlier notation this may be specified as,

$$q(x_t | x_{t-1}, y_{1:t}) = N(x_{t|t}, P_{t|t}). \quad (1.62)$$

Further information can be obtained from the cited references.

1.8 Particle Filter Algorithm

In this section we discuss the enhancement to the filtering algorithm needed to accommodate particle filter. This applies to both the EKF and the UKF described earlier.

At $t=0$, we first set x_0 and P_0 similar to earlier algorithms.

For each point of the N sampling points of the simulation we set,

$$x_0^i = x_0 + \sqrt{P_0} z^i \quad (1.63)$$

where z is a standard Normal simulated number. Also, let

$$P_0^i = P_0, \quad w_0^i = \frac{1}{N}.$$

For each of the simulation points (i), follow the steps:

$$x_{t|t}^i = \text{KF}(x_{t-1|t-1}^0)$$

and similarly the associated covariance $P_{t|t}^i$ computed via the EKF or the UKF. KF indicates this association with the EKF or the UKF.

For each simulation point again,

$$\tilde{x}_t^i = x_{t|t}^i + \sqrt{P_0^i} z^i$$

with z being a simulated standard Gaussian number.

Now, calculate the weight for each of the simulation points following,

$$w_t^i = w_{t-1}^i \frac{p(y_t | \tilde{x}_t^i) p(\tilde{x}_t^i | x_{t-1}^i)}{q(\tilde{x}_t^i | x_{t-1}^i, y_{1:t})}$$

where $q(\cdot)$ is the proposal Normal density with mean $x_{t|t}^i$ and covariance $P_{t|t}^i$.

We now need to normalize the weights as,

$$\tilde{w}_t^i = \frac{w_t^i}{\sum_{i=1}^N w_t^i}$$

and resample the points \tilde{x}_t^i , get x_t^i and reset the weights as

$$w_t^i = \tilde{w}_t^i = \frac{1}{N}$$

and repeat the steps for the next time step.

Using the particle filter we can also estimate the unknown model parameters by maximizing the likelihood function as before. Since, by assumption the distribution is not Gaussian in this case; we need some understanding how to formulate the problem.

The likelihood in step t is,

$$l_t = p(y_t | y_{1:t-1}) = \int p(y_t | x_t) p(x_t | y_{1:t-1}) dx_t \quad (1.64)$$

the total likelihood is the product of the above for all t . However, (1.64) may also be written as,

$$l_t = \int p(y_t | x_t) \frac{p(x_t | y_{1:t-1})}{q(x_t | x_{t-1}, y_{1:t})} q(x_t | x_{t-1}, y_{1:t}) dx_t \quad (1.65)$$

and since by construction \tilde{x}_t^i are distributed as $q(\cdot)$, and resetting \tilde{w}_t^i to a constant $(1/N)$ during re-sampling, it is possible to approximate l_t as,

$$\tilde{l}_t = \sum_{i=1}^N w_t^i. \quad (1.66)$$

This gives an alternative interpretation the total likelihood.

1.9 Unobserved Component Models

This section focuses on adaptive, off-line signal processing and forecasting for non-stationary time series described by the following Unobserved Component model (e.g. Young et al, 1998):

$$y_t = T_t + S_t + g(u_t) + e_t \quad (1.67)$$

where, y_t as before is the observed time series, T_t is the low frequency or trend component, S_t is a periodic component, possibly displaying temporal change in both amplitude and phase, $g(u_t)$ captures the influence of a set of exogenous variables and e_t is an irregular component assumed to be Normally distributed with mean zero and variance σ^2 .

In financial economic applications most time series we encounter are non-stationary, or first difference stationary. In these situations, it is customary to define the trend component in (1.67) as a random walk. Cochrane (1988) uses the Beveridge and Nelson (1981) decomposition to express a first-difference stationary process as the sum of (covariance) stationary and random walk components. He argues that a measurement of the size of the random walk component can be a better guide to the proper statistical characterization of the series than a simple unit root test. He proposes a non-parametric method, the Variance Ratio, for determining the magnitudes of the random walk and stationary components of a time series. Further insight may be gained from Cochrane's article. Fama and French (1988) utilize similar concept for

stock market application with non-stationary trend and stationary temporal components. Hatgioannides, Karanasos and Karanassou (2004) also apply the same intuition to term structure of interest rate model.

Here, we proceed to understand the issues relating to filtering in such a set up as in (1.67).

In (1.67) several components of the model are present and, in general, it may lead to identification problem. However, in the special case below the model remains identifiable. This is referred to as the Dynamic Harmonic Regression (DHR), Young (1988).

$$y_t = T_t + S_t + e_t, \quad e_t \sim N(0, \sigma^2). \quad (1.68)$$

In (1.68) the most important unobserved component is the term S_t defined as,

$$S_t = \sum_{i=1}^M \{a_{i,t} \cos(\omega_{i,t}) + b_{i,t} \sin(\omega_{i,t})\} \quad (1.69)$$

where $a_{i,t}$, $b_{i,t}$ are time varying parameters and ω_i represents fundamental and harmonic frequencies associated with the periodicity of the series. The frequency range could be chosen to match the spectral properties of the time series and discussed in detail in Young (1988). The DHR model is, therefore, time dependent extension of the classical Fourier series.

Each of the time varying components and the trend component (T_t) in (1.68) could be represented by a state vector of size 2, $x_i \equiv [l_i, d_i]'$, where the two components represent the changing level and the slope of the associated trend or time varying parameter. Usually, the stochastic evolution of each $x_i \equiv [l_i, d_i]'$ is assumed to be driven by a Generalized Random Walk (GRW) of the form,

$$x_{i,t} = \begin{bmatrix} \alpha & \beta \\ 0 & \gamma \end{bmatrix} x_{i,t-1} + \begin{bmatrix} \delta & 0 \\ 0 & 1 \end{bmatrix} \eta_{i,t}, \quad i=1, \dots, M+1. \quad (1.70)$$

With appropriate constraints on the elements of the matrices in (1.70) different Random Walk models could be realized. For example, Integrated Random Walk would result if $\alpha = \beta = \gamma = 1, \delta = 0$. Scalar Random Walk would result if $\alpha = 1, \beta = \gamma = 0, \delta = 1$. Similarly, Local Linear Trend results if $\alpha = \beta = \gamma = \delta = 1$. Further discussions about this can be found in Young et al (1989).

The overall state space model can be described as an example, assuming that $M=1$. The main idea is to stack the system matrices in (1.70) such a way that the overall state equation is obtained. Similarly, by suitable choice of the measurement matrices the measurement equation of (1.68) is obtained. For $M=1$ this results in the state equation,

$$x_t \equiv \begin{bmatrix} l_{1,t} \\ d_{1,t} \\ l_{2,t} \\ d_{2,t} \end{bmatrix} = \begin{bmatrix} \alpha_1 & \beta_1 & 0 & 0 \\ 0 & \gamma_1 & 0 & 0 \\ 0 & 0 & \alpha_2 & \beta_2 \\ 0 & 0 & 0 & \gamma_2 \end{bmatrix} x_{t-1} + \begin{bmatrix} \delta_1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \delta_2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_1 \\ \eta_2 \\ \eta_2 \end{bmatrix} \quad (1.71)$$

and the measurement equation is,

$$y_t = [1 \ 1 \ 1 \ 1] x_t + e_t. \quad (1.72)$$

For the special case of $M=1$, the equations (1.71) and (1.72) are in the state space form and the algorithm described in an earlier section for Linear Kalman Filter can be applied to estimate the model parameters by maximizing the likelihood function.

To complete our discussions on filtering algorithms we outline here the Backward Pass Filter or Fixed Interval Smoothing algorithm. The algorithm described in section 1.3 in connection with Linear Kalman Filter is also referred to as Forward Pass Filter. For most applications in financial economics the time series of observations have already been made. In other words, the researcher is attempting to fit a model a historical time series. By applying the Forward Pass Filter the model parameters are obtained by maximizing the likelihood function. Also, at the same time, since the complete observation data set is available it is possible to refine the estimates via the Backward Pass Filter by running the procedure from the end of the time series to the beginning.

We define the recursive relations for the Backward Pass Filter with reference to the model in (1.8) and (1.9). Starting with the initial condition at the last measurement point, $y_{T|T}$ and $P_{T|T}$

$$y_{t-1|T} = y_{t-1|t-1} + J_{t-1} (y_{t|T} - y_{t|t-1}) \quad (1.73)$$

$$P_{t-1|T} = P_{t-1|t-1} + J_{t-1} (P_{t|T} - P_{t|t-1}) J'_{t-1} \quad (1.74)$$

where,

$$J_{t-1} = P_{t-1|t-1} T' [P_{t|t-1}]^{-1}. \quad (1.75)$$

It should be clear from the above that to implement the Backward Pass Filter the quantities $y_{t|t}$ and $P_{t|t}$ generated during the Forward Pass Filter must be saved.

This completes our exposition on filtering algorithms and those interested in learning more about applications in engineering (signal processing) the references cited in this section are excellent source.

1.10 Concluding Remarks

We have provided an exposition of various filtering scheme and the main algorithm to implement these for a typical application in the discipline of financial economics. This includes Linear Kalman Filter and its extensions e.g. Extended Kalman Filter applicable to non-linear Gaussian models, Unscented Kalman Filter an alternative to EKF but with more precision as well as Particle Filter to deal with non-Gaussian models. At the end we have also introduced Dynamic Harmonic Regression to cover unobserved component models with time varying parameters and possibly non-stationary time series. The focus has mainly been in the algorithms and the conceptual background to these sophisticated approaches to stochastic modeling.

In the following chapters we are going to discuss various applications that utilize such algorithms. These applications will relate to foreign exchange markets, stock markets, interest rate models as wells as market for fixed income securities. These chapters will include more of economic intuitions in the estimated models – parameters as well as the filtered components. There would be sufficient background to the development of the models for different markets, but the filtering process would rely on the materials in chapter 1.