Filtering in Finance

Alireza Javaheri, RBC Capital Markets
Delphine Lautier, Université Paris IX, Ecole Nationale Supérieure des Mines de Paris
Alain Galli, Ecole Nationale Supérieure des Mines de Paris

Abstract

In this article we present an introduction to various Filtering algorithms and some of their applications to the world of Quantitative Finance. We shall first mention the fundamental case of Gaussian noises where we obtain the well-known Kalman Filter. Because of common nonlinearities, we will be discussing the Extended Kalman Filter (EKF) as well as the Unscented Kalman Filter (UKF) similar to Kushner's Nonlinear Filter. We also tackle the subject of Non-Gaussian filters and describe the Particle Filtering (PF) algorithm. Lastly, we will apply the filters to the term structure model of commodity prices and the stochastic volatility model.

1 Filtering

The concept of filtering has long been used in Control Engineering and Signal Processing. Filtering is an *iterative* process that enables us to estimate a model's parameters when the latter relies upon a large quantity of observable *and* unobservable data. The Kalman Filter is fast and easy to implement, despite the length and noisiness of the input data.

We suppose we have a temporal time-series of observable data $\mathbf{z}_k$ (*e.g.* stock prices as in Javaheri (2002), Wells (1996), interest rates as in Babbs and Nowman (1999), Pennacchi (1991), futures prices as in Lautier (2000), Lautier and Galli (2000)) and a model using some unobservable time-series $\mathbf{x}_k$ (*e.g.* volatility, correlation, convenience yield) where the index k corresponds to the time-step. This will allow us to construct an algorithm containing a transition equation linking two consecutive unobservable states, and a measurement equation relating the observed data to this hidden state.

The idea is to proceed in two steps: first we estimate the hidden state *a priori* by using all the information prior to that time-step. Then using this predicted value together with the new observation, we obtain a conditional *a posteriori* estimation of the state.

In what follows we shall first tackle linear and nonlinear equations with Gaussian noises. We then will extend this idea to the Non-Gaussian case.

Further, we shall provide a mean to estimate the model parameters via the maximization of the *likelihood* function.

1.1 The Simple and Extended Kalman Filters

1.1.1 Background and Notations

In this section we describe both the traditional Kalman Filter used for linear systems and its extension to nonlinear systems known as the Extended Kalman Filter (EKF). The latter is based upon a first order linearization of the transition and measurement equations and therefore would coincide with the traditional filter when the equations are linear. For a detailed introduction, see Harvey (1989) or Welch and Bishop (2002).

Given a dynamic process $\mathbf{x}_k$ following a transition equation

$$\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{w}_k) \qquad (1)$$

we suppose we have a measurement $\mathbf{z}_k$ such that

$$\mathbf{z}_k = \mathbf{h}(\mathbf{x}_k, \mathbf{u}_k) \qquad (2)$$

where $\mathbf{w}_k$ and $\mathbf{u}_k$ are two mutually-uncorrelated sequences of temporally-uncorrelated Normal random-variables with zero means and covariance matrices $\mathbf{Q}_k$, $\mathbf{R}_k$ respectively[4]. Moreover, $\mathbf{w}_k$ is uncorrelated with $\mathbf{x}_{k-1}$ and $\mathbf{u}_k$ uncorrelated with $\mathbf{x}_k$.

We denote the dimension of $\mathbf{x}_k$ as n_x, the dimension of $\mathbf{w}_k$ as n_w and so on.

We define the *a priori* process estimate as

$$\hat{\mathbf{x}}_k^- = E[\mathbf{x}_k] \tag{3}$$

which is the estimation at time step $k-1$ prior to the step k measurement.

Similarly, we define the *a posteriori* estimate

$$\hat{\mathbf{x}}_k = E[\mathbf{x}_k|\mathbf{z}_k] \tag{4}$$

which is the estimation at time step k after the measurement.

We also have the corresponding estimation errors $\mathbf{e}_k^- = \mathbf{x}_k - \hat{\mathbf{x}}_k^-$ and $\mathbf{e}_k = \mathbf{x}_k - \hat{\mathbf{x}}_k$ and the estimate error covariances

$$\mathbf{P}_k^- = E\left[\mathbf{e}_k^- \mathbf{e}_k^{-t}\right]$$

$$\mathbf{P}_k = E\left[\mathbf{e}_k \mathbf{e}_k^t\right] \tag{5}$$

where the superscript t corresponds to the transpose operator.

In order to evaluate the above means and covariances we will need the conditional densities $p(\mathbf{x}_k|\mathbf{z}_{k-1})$ and $p(\mathbf{x}_k|\mathbf{z}_k)$, which are determined iteratively via the Time Update and Measurement Update equations. The basic idea is to find the probability density function corresponding to a hidden state x_k at time step k given all the observations $z_{1:k}$ up to that time.

The *Time Update* step is based upon the Chapman-Kolmogorov equation

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{z}_{1:k-1})p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})d\mathbf{x}_{k-1}$$
$$= \int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})d\mathbf{x}_{k-1} \tag{6}$$

via the Markov property.

The *Measurement Update* step is based upon the Bayes rule

$$p(\mathbf{x}_k|\mathbf{z}_{1:k}) = \frac{p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{z}_{1:k-1})}{p(\mathbf{z}_k|\mathbf{z}_{1:k-1})} \tag{7}$$

with

$$p(\mathbf{z}_k|\mathbf{z}_{1:k-1}) = \int p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{z}_{1:k-1})d\mathbf{x}_k$$

A proof of the above is given in the appendix.

EKF is based on the linearization of the transition and measurement equations, and uses the conservation of Normal property within the class of linear functions. We therefore define the Jacobian matrices of $\mathbf{f}$ with respect to the state variable and the system noise as $\mathbf{A}_k$ and $\mathbf{W}_k$ respectively; and similarly for $\mathbf{h}$, as $\mathbf{H}_k$ and $\mathbf{U}_k$ respectively.

More accurately, for every row i and column j we have

$$\mathbf{A}_{ij} = \partial \mathbf{f}_i/\partial \mathbf{x}_j(\hat{\mathbf{x}}_{k-1}, 0), \mathbf{W}_{ij} = \partial \mathbf{f}_i/\partial \mathbf{w}_j(\hat{\mathbf{x}}_{k-1}, 0),$$
$$\mathbf{H}_{ij} = \partial \mathbf{h}_i/\partial \mathbf{x}_j(\hat{\mathbf{x}}_k^-, 0), \mathbf{U}_{ij} = \partial \mathbf{h}_i/\partial \mathbf{u}_j(\hat{\mathbf{x}}_k^-, 0)$$

Needless to say for a linear system, the function matrices are *equal* to these Jacobians. This is the case for the simple Kalman Filter.

1.1.2 The Algorithm
The actual algorithm could be implemented as follows:

(1) Initialization of $\mathbf{x}_0$ and $\mathbf{P}_0$

For k in $1...N$

(2) Time Update (Prediction) equations

$$\hat{\mathbf{x}}_k^- = \mathbf{f}(\hat{\mathbf{x}}_{k-1}, 0) \tag{8}$$

and

$$\mathbf{P}_k^- = \mathbf{A}_k \mathbf{P}_{k-1} \mathbf{A}_k^t + \mathbf{W}_k \mathbf{Q}_{k-1} \mathbf{W}_k^t \tag{9}$$

(3-a) Innovation: We define

$$\hat{\mathbf{z}}_k^- = \mathbf{h}(\hat{\mathbf{x}}_k^-, 0)$$

and

$$\nu_k = \mathbf{z}_k - \hat{\mathbf{z}}_k^-$$

as the innovation process.

(3-b) Measurement Update (Filtering) equations

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k \nu_k \tag{10}$$

and

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)\mathbf{P}_k^- \tag{11}$$

with

$$\mathbf{K}_k = \mathbf{P}_k^- \mathbf{H}_k^t (\mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^t + \mathbf{U}_k \mathbf{R}_k \mathbf{U}_k^t)^{-1} \tag{12}$$

and $\mathbf{I}$ the Identity matrix.

The above Kalman gain $\mathbf{K}_k$ corresponds to the mean of the conditional distribution of $\mathbf{x}_k$ upon the observation $\mathbf{z}_k$ or equivalently, the matrix that would minimize the mean square error $\mathbf{P}_k$ *within the class of linear estimators*.

This interpretation is based upon the following observation. Having x a Normally distributed random-variable with a mean m_x and variance P_{xx}, and z another Normally distributed random-variable with a mean m_z and variance P_{zz}, and having $P_{zx} = P_{xz}$ the covariance between x and z, the conditional distribution of $x|z$ is also Normal with

$$m_{x|z} = m_x + K(z - m_z)$$

with

$$K = P_{xz} P_{zz}^{-1}$$

which corresponds to our Kalman gain.

1.1.3 Parameter Estimation
For a parameter-set Ψ in the model, the calibration could be carried out via a Maximum Likelihood Estimator (MLE) or in case of conditionally Gaussian state variables, a Quasi-Maximum Likelihood (QML) algorithm.

To do this we need to maximize $\prod_{k=1}^{N} p(z_k|z_{1:k-1})$, and given the Normal form of this probability density function, taking the logarithm, changing the signs and ignoring constant terms, this would be equivalent to minimizing over the parameter-set Ψ

$$L_{1:N} = \sum_{k=1}^{N} ln(P_{z_k z_k}) + \frac{z_k - m_{z_k}}{P_{z_k z_k}} \tag{13}$$

For the KF or EKF we have

$$m_{z_k} = \hat{\mathbf{z}}_k^-$$

and

$$P_{z_k z_k} = \mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^t + \mathbf{U}_k \mathbf{R}_k \mathbf{U}_k^t$$

1.2 The Unscented Kalman Filter and Kushner's Nonlinear Filter

1.2.1 Background and Notations

Recently Julier and Uhlmann (1997) proposed a new extension of the Kalman Filter to Nonlinear systems, different from the EKF. The new method called the Unscented Kalman Filter (UKF) will calculate the mean to a higher order of accuracy than the EKF, and the covariance to the same order of accuracy.

Unlike the EKF, this method does *not* require any Jacobian calculation since it does not approximate the nonlinear functions of the process and the observation. Indeed it uses the true nonlinear models but approximates the distribution of the state random variable $\mathbf{x}_k$ (as well as the observation $\mathbf{z}_k$) with a Normal distribution by applying an *Unscented Transformation* to it.

In order to be able to apply this Gaussian approximation (unless we have $\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}) + \mathbf{w}_k$ and $\mathbf{z}_k = \mathbf{h}(\mathbf{x}_k) + \mathbf{u}_k$, *i.e.* unless the equations are linear in noise) we will need to *augment* the state space by concatenating the noises to it. This augmented state will have a dimension $n_a = n_x + n_w + n_u$.

1.2.2 The Algorithm

The UKF algorithm could be written in the following way:

(1-a) Initialization: Similarly to the EKF, we start with an initial choice for the state vector $\hat{\mathbf{x}}_0 = E[\mathbf{x}_0]$ and its covariance matrix $\mathbf{P}_0 = E[(\mathbf{x}_0 - \hat{\mathbf{x}}_0)(\mathbf{x}_0 - \hat{\mathbf{x}}_0)^t]$.

We also define the weights $W_i^{(m)}$ and $W_i^{(c)}$ as

$$W_0^{(m)} = \frac{\lambda}{n_a + \lambda}$$

and

$$W_0^{(c)} = \frac{\lambda}{n_a + \lambda} + (1 - \alpha^2 + \beta)$$

and for $i = 1...2n_a$

$$W_i^{(m)} = W_i^{(c)} = \frac{1}{2(n_a + \lambda)} \tag{14}$$

where the scaling parameters α, β, κ and $\lambda = \alpha^2(n_a + \kappa) - n_a$ will be chosen for tuning.

For k in $1...N$

(1-b) State Space Augmentation: As mentioned earlier, we concatenate the state vector with the system noise and the observation noise, and create an augmented state vector for each time-step

$$\mathbf{x}_{k-1}^a = \begin{bmatrix} \mathbf{x}_{k-1} \\ \mathbf{w}_{k-1} \\ \mathbf{u}_{k-1} \end{bmatrix}$$

and therefore

$$\hat{\mathbf{x}}_{k-1}^a = E[\mathbf{x}_{k-1}^a | \mathbf{z}_k] = \begin{bmatrix} \hat{\mathbf{x}}_{k-1} \\ 0 \\ 0 \end{bmatrix}$$

and

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

(1-c) The Unscented Transformation: Following this, in order to use the Normal approximation, we need to construct the corresponding *Sigma Points* through the Unscented Transformation:

$$\chi_{k-1}^a(0) = \hat{\mathbf{x}}_{k-1}^a$$

For $i = 1...n_a$

$$\chi_{k-1}^a(i) = \hat{\mathbf{x}}_{k-1}^a + (\sqrt{(n_a + \lambda)\mathbf{P}_{k-1}^a})_i$$

and for $i = n_a + 1...2n_a$

$$\chi_{k-1}^a(i) = \hat{\mathbf{x}}_{k-1}^a - (\sqrt{(n_a + \lambda)\mathbf{P}_{k-1}^a})_{i-n_a} \tag{15}$$

where the above subscripts i and $i - n_a$ correspond to the ith and $i - n_a^{th}$ columns of the square-root matrix[5].

(2) Time Update equations are

$$\chi_{k|k-1}(i) = \mathbf{f}(\chi_{k-1}^x(i), \chi_{k-1}^w(i)) \tag{16}$$

for $i = 0...2n_a + 1$ and

$$\hat{\mathbf{x}}_k^- = \sum_{i=0}^{2n_a} W_i^{(m)} \chi_{k|k-1}(i) \tag{17}$$

and

$$\mathbf{P}_k^- = \sum_{i=0}^{2n_a} W_i^{(c)} (\chi_{k|k-1}(i) - \hat{\mathbf{x}}_k^-)(\chi_{k|k-1}(i) - \hat{\mathbf{x}}_k^-)^t \tag{18}$$

The superscripts x and w correspond to the state and system-noise portions of the augmented state respectively.

(3-a) Innovation: We define

$$\mathbf{z}_{k|k-1}(i) = \mathbf{h}(\chi_{k|k-1}(i), \chi_{k-1}^u(i)) \tag{19}$$

and

$$\hat{\mathbf{z}}_k^- = \sum_{i=0}^{2n_a} W_i^{(m)} \mathbf{z}_{k|k-1}(i) \tag{20}$$

and as before

$$\nu_k = \mathbf{z}_k - \hat{\mathbf{z}}_k^-$$

(3-b) Measurement Update

$$\mathbf{P}_{z_k z_k} = \sum_{i=0}^{2n_a} W_i^{(c)} (\mathbf{Z}_{k|k-1}(i) - \hat{\mathbf{z}}_k^-)(\mathbf{Z}_{k|k-1}(i) - \hat{\mathbf{z}}_k^-)^t$$

and

$$\mathbf{P}_{x_k z_k} = \sum_{i=0}^{2n_a} W_i^{(c)} (\chi_{k|k-1}(i) - \hat{\mathbf{x}}_k^-)(\mathbf{Z}_{k|k-1}(i) - \hat{\mathbf{z}}_k^-)^t \qquad (21)$$

which gives us the Kalman Gain

$$\mathbf{K}_k = \mathbf{P}_{x_k z_k} \mathbf{P}_{z_k z_k}^{-1}$$

and we have as previously

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k \nu_k \qquad (23)$$

as well as

$$\mathbf{P}_k = \mathbf{P}_k^- - \mathbf{K}_k \mathbf{P}_{z_k z_k} \mathbf{K}_k^t \qquad (23)$$

which completes the Measurement Update Equations.

1.2.3 Parameter Estimation
The maximization of the likelihood could be done exactly as in (13) taking $\hat{\mathbf{z}}_k^-$ and $P_{z_k z_k}$ defined as above.

1.2.4 Analogy with Kushner's Nonlinear Filter
It would be interesting to compare this algorithm to Kushner's Nonlinear Filter[6] (NLF) based on an approximation of the conditional distribution as explained in Kushner (1967), Kushner and Budhiraja (2000). In this approach, the authors suggest using a Normal approximation to the densities $p(\mathbf{x}_k|\mathbf{z}_{k-1})$ and $p(\mathbf{x}_k|\mathbf{z}_k)$. They then use the fact that a Normal distribution is entirely determined via its first two moments, which reduces the calculations considerably.

They finally rewrite the moment calculation equations (3), (4) and (5) using the above $p(\mathbf{x}_k|\mathbf{z}_{k-1})$ and $p(\mathbf{x}_k|\mathbf{z}_k)$, after calculating these conditional densities via the time and measurement update equations (6) and (7). All integrals could be evaluated via Gaussian Quadratures[7].

Note that when $h(x, u)$ is strongly nonlinear, the Gauss Hermite integration is not efficient for evaluating the moments of the measurement update equation, since the term $p(z_k|x_k)$ contains the exponent $z_k - h(x_k)$. The iterative methods based on the idea of importance sampling proposed in Kushner (1967), Kushner and Budhiraja (2000) correct this problem at the price of a strong increase in computation time. As suggested in Ito and Xiong (2000), one way to avoid this integration would be to make the additional hypothesis that $x_k, h(x_k)|z_{1:k-1}$ is Gaussian.

When $n_x = 1$ and $\lambda = 2$, the numeric integration in the UKF will correspond to a Gauss-Hermite Quadrature of order 3. However in the UKF we can tune the filter and reduce the higher term errors via the previously mentioned parameters α and β.

As the Kushner paper indicates, having an N-dimensional Normal random-variable $\mathbf{X} = \mathcal{N}(\mathbf{m}, \mathbf{P})$ with $\mathbf{m}$ and $\mathbf{P}$ the corresponding mean and covariance, for a polynomial G of degree $2M - 1$ we can write

$$E[G(\mathbf{X})] = \frac{1}{(2\pi)^{\frac{N}{2}} |\mathbf{P}|^{\frac{1}{2}}} \int_{R^N} G(\mathbf{y}) exp[-\frac{(\mathbf{y} - \mathbf{m})^t \mathbf{P}^{-1}(\mathbf{y} - \mathbf{m})}{2}] d\mathbf{y}$$

which is equal to

$$E[G(\mathbf{X})] = \sum_{i_1=1}^{M} ... \sum_{i_N=1}^{M} w_{i_1}...w_{i_N} G(\mathbf{m} + \sqrt{\mathbf{P}} \zeta)$$

where $\zeta^t = (\zeta_{i_1} \quad ... \quad \zeta_{i_N})$ is the vector of the Gauss-Hermite roots of order M and $w_{i_1}...w_{i_N}$ are the corresponding weights.

Note that even if both Kushner's NLF and UKF use Gaussian Quadratures, UKF only uses $2N + 1$ sigma points, while NLF needs M^N points for the computation of the integrals.

More accurately, for a Quadrature-order M and an N-dimensional (possibly augmented) variable, the sigma-points are defined for $j = 1...N$ and $i_{j=1}...M$ as

$$\chi_{k-1}^a(i_1, ..., i_N) = \hat{\mathbf{x}}_{k-1}^a + \sqrt{\mathbf{P}_{k-1}^a} \zeta(i_1, ..., i_N)$$

where this square-root corresponds to the Cholesky factorization. Similarly to the UKF, we have the Time Update equations

$$\chi_{k|k-1}(i_1, ..., i_N) = \mathbf{f}\left(\chi_{k-1}^x(i_1, ..., i_N), \chi_{k-1}^w(i_1, ..., i_N)\right)$$

but now

$$\hat{\mathbf{x}}_k^- = \sum_{i_1=1}^{M} ... \sum_{i_N=1}^{M} w_{i_1}...w_{i_N} \chi_{k|k-1}(i_1, ..., i_N)$$

and

$$\mathbf{P}_k^- = \sum_{i_1=1}^{M} ... \sum_{i_N=1}^{M} w_{i_1}...w_{i_N} (\chi_{k|k-1}(i_1, ..., i_N) - \hat{\mathbf{x}}_k^-)(\chi_{k|k-1}(i_1, ..., i_N) - \hat{\mathbf{x}}_k^-)^t$$

and similarly for the measurement update equations.

1.3 The Non-Gaussian Case: The Particle Filter
It is possible to generalize the algorithm for the fundamental Gaussian case to one applicable to any distribution.

1.3.1 Background and Notations
In this approach, we use Markov-Chain Monte-Carlo simulations instead of using a Gaussian approximation for $(x_k|z_k)$ as the Kalman or Kushner Filters do. A detailed description is given in Doucet, De Freitas and Gordon (2001).

The idea is based on the *Importance Sampling* technique:
We can calculate an expected value

$$E[f(x_k)] = \int f(x_k) p(x_k|z_{1:k}) dx_k \qquad (24)$$

by using a known and simple *proposal* distribution $q()$.

More precisely, it is possible to write

$$E[f(x_k)] = \int f(x_k) \frac{p(x_k|z_{1:k})}{q(x_k|z_{1:k})} q(x_k|z_{1:k}) dx_k$$

which could be also written as

$$E[f(x_k)] = \int f(x_k) \frac{w_k(x_k)}{p(z_{1:k})} q(x_k|z_{1:k}) dx_k \qquad (25)$$

with

$$w_k(x_k) = \frac{p(z_{1:k}|x_k)p(x_k)}{q(x_k|z_{1:k})}$$

defined as the filtering *non-normalized weight* as step k.

We therefore have

$$E[f(x_k)] = \frac{E_q[w_k(x_k)f(x_k)]}{E_q[w_k(x_k)]} = E_q[\tilde{w}_k(x_k)f(x_k)] \qquad (26)$$

with

$$\tilde{w}_k(x_k) = \frac{w_k(x_k)}{E_q[w_k(x_k)]}$$

defined as the filtering *normalized weight* as step k.

Using Monte-Carlo sampling from the distribution $q(x_k|z_{1:k})$ we can write in the discrete framework:

$$E[f(x_k)] \approx \sum_{i=1}^{N_{sims}} \tilde{w}_k(x_k^{(i)}) f(x_k^{(i)}) \qquad (27)$$

with again

$$\tilde{w}_k(x_k^{(i)}) = \frac{w_k(x_k^{(i)})}{\sum_{j=1}^{N_{sims}} w_k(x_k^{(j)})}$$

Now supposing that our proposal distribution $q()$ satisfies the Markov property, it can be shown that w_k verifies the recursive identity

$$w_k^{(i)} = w_{k-1}^{(i)} \frac{p(z_k|x_k^{(i)})p(x_k^{(i)}|x_{k-1}^{(i)})}{q(x_k^{(i)}|x_{k-1}^{(i)}, z_{1:k})} \qquad (28)$$

which completes the *Sequential Importance Sampling* algorithm. It is important to note that this means that the state x_k cannot depend on future observations, *i.e.* we are dealing with Filtering and not Smoothing[8].

One major issue with this algorithm is that the variance of the weights increases randomly over time. In order to solve this problem, we could use a *Resampling* algorithm which would map our unequally weighted x_k's to a new set of equally weighted sample points. Different methods have been suggested for this[9].

Needless to say, the choice of the proposal distribution is crucial. Many suggest using

$$q(x_k|x_{k-1}, z_{1:k}) = p(x_k|x_{k-1})$$

since it will give us a simple weight identity

$$w_k^{(i)} = w_{k-1}^{(i)} p(z_k|x_k^{(i)})$$

However this choice of the proposal distribution does *not* take into account our most recent observation z_k at all and therefore could become inefficient.

Hence the idea of using a Gaussian Approximation for the proposal, and in particular an approximation based on the Kalman Filter, in order to incorporate the observations.

We therefore will have

$$q(x_k|x_{k-1}, z_{1:k}) = \mathcal{N}(\hat{x}_k, P_k) \qquad (29)$$

using the same notations as in the section on the Kalman Filter. Such filters are sometimes referred to as the Extended Particle Filter (EPF) and the Unscented Particle Filter (UPF). See Haykin (2001) for a detailed description of these algorithms.

1.3.2 The Algorithm
Given the above framework, the algorithm for an Extended or Unscented Particle Filter could be implemented in the following way:

(1) For time step $k = 0$ choose x_0 and $P_0 > 0$.
   For i such that $1 \le i \le N_{sims}$ take

$$x_0^{(i)} = x_0 + \sqrt{P_0} Z^{(i)}$$

where $Z^{(i)}$ is a standard Gaussian simulated number.
   Also take $P_0^{(i)} = P_0$ and $w_0^{(i)} = 1/N_{sims}$
   While $1 \le k \le N$

(2) For each simulation-index i

$$\hat{x}_k^{(i)} = \mathbf{KF}(x_{k-1}^{(i)})$$

with $P_k^{(i)}$ the associated *a posteriori* error covariance matrix.
   (KF could be either the EKF or the UKF)

(3) For each i between 1 and N_{sims}

$$\tilde{x}_k^{(i)} = \hat{x}_k^{(i)} + \sqrt{P_k^{(i)}} Z^{(i)}$$

where again $Z^{(i)}$ is a standard Gaussian simulated number.
(4) Calculate the associated weights for each i

$$w_k^{(i)} = w_{k-1}^{(i)} \frac{p(z_k|\tilde{x}_k^{(i)})p(\tilde{x}_k^{(i)}|x_{k-1}^{(i)})}{q(\tilde{x}_k^{(i)}|x_{k-1}^{(i)}, z_{1:k})}$$

with $q()$ the Normal density with mean $\hat{x}_k^{(i)}$ and variance $P_k^{(i)}$.
(5) Normalize the weights

$$\tilde{w}_k^{(i)} = \frac{w_k^{(i)}}{\sum_{i=1}^{N_{sims}} w_k^{(i)}}$$

(6) Resample the points $\tilde{x}_k^{(i)}$ and get $x_k^{(i)}$ and reset $w_k^{(i)} = \tilde{w}_k^{(i)} = 1/N_{sims}$.
(7) Increment k, Go back to step **(2)** and **Stop** at the end of the **While** loop.

1.3.3 Parameter Estimation

As in the previous section, in order to estimate the parameter-set Ψ we can use an ML Estimator. However since the particle filter does not necessarily assume Gaussian noise, the likelihood function to be maximized has a more general form than the one used in previous sections.

Given the likelihood at step k

$$l_k = p(z_k|z_{1:k-1}) = \int p(z_k|x_k)p(x_k|z_{1:k-1})dx_k$$

the total likelihood is the product of the l_k's above and therefore the log-likelihood to be maximized is

$$\ln(L_{1:N}) = \sum_{k=1}^{N} \ln(l_k) \tag{30}$$

Now l_k could be written as

$$l_k = \int p(z_k|x_k) \frac{p(x_k|z_{1:k-1})}{q(x_k|x_{k-1}, z_{1:k})} q(x_k|x_{k-1}, z_{1:k})dx_k$$

and given that by construction the $\tilde{x}_k^{(i)}$'s are distributed according to $q()$, considering the resetting of $w_k^{(i)}$ to a constant $1/N_{sims}$ during the resampling step, we can approximate l_k with

$$\tilde{l}_k = \sum_{i=1}^{N_{sims}} w_k^{(i)}$$

which will provide us with an interpretation of the likelihood as the *total weight*.

2 Term Structure Models of Commodity Prices

In this section, the Kalman filter is applied to a well known term structure model of commodity prices developed by Schwartz (1997). We first present the Schwartz model, and show how it can be transformed into a state-spaced model for a simple filter as well as an extended filter. We then explain how the iteration process can be initiated, and how it can be stabilized. Lastly, we compare the performance obtained with the simple and extended filters, and make a sensitivity analysis.

2.1 The Schwartz model

The Schwartz model supposes that two state variables, namely the spot price S and the convenience yield C, can explain the behavior of the futures prices F. The convenience yield can briefly be defined as the comfort associated with the possession of physical stocks. There are usually no empirical data for these two variables, because most of the time there are no reliable time series for the spot price, and the convenience yield is not a traded asset.

2.1.1 Presentation of the model

The dynamics of these state variables are the following:

$$\begin{cases} dS = (\mu - C)S dt + \sigma_S S dz_S \\ dC = [\kappa(\alpha - C)]dt + \sigma_C dz_C \end{cases}$$

with:

$$E[dz_S \times dz_C] = \rho dt$$

$$\kappa, \sigma_S, \sigma_C > 0$$

where:
- μ is the immediate expected return for the spot price S,
- σ_S is the spot price's volatility,
- dz_S is the increment of the Brownian motion associated with S,
- α is the long run mean of the convenience yield C,
- κ represents the convergence of the convenience yield towards α,
- σ_C is the convenience yield's volatility,
- dz_C is the increment of the Brownian motion associated with C.
- ρ is the correlation between the two Brownian motions associated with S and C,

The model's solution expresses the relationship between an observable futures price F for a delivery in T, and the state variables. This solution is:

$$F(S, C, t, T) = S(t) \times \exp\left[-C(t)\frac{1 - e^{-\kappa\tau}}{\kappa} + B(\tau)\right]$$

with:

$$B(\tau) = \left[\left(r - \hat{\alpha} + \frac{\sigma_C^2}{2\kappa^2} - \frac{\sigma_S\sigma_C\rho}{\kappa}\right) \times \tau\right] + \left[\frac{\sigma_C^2}{4} \times \frac{1 - e^{-2\kappa\tau}}{\kappa^3}\right] + \left[\left(\hat{\alpha}\kappa + \sigma_S\sigma_C\rho - \frac{\sigma_C^2}{\kappa}\right) \times \left(\frac{1 - e^{-\kappa\tau}}{\kappa^2}\right)\right],$$

$$\hat{\alpha} = \alpha - (\lambda/\kappa)$$

where:
- r is the risk-free interest rate[10],
- λ is the risk premium associated with the convenience yield,
- $\tau = T - t$ is the maturity of the futures contract.

The model risk-neutral parameter-set is therefore: $\psi = (\sigma_S, \kappa, \alpha, \sigma_C, \rho, \lambda)$

2.1.2 Applying the simple filter to the Schwartz model

To apply the simple filter, the Schwartz model must be expressed under a linear form:

$$\ln(F(S, C, t, T)) = \ln(S(t)) - C(t) \times \frac{1 - e^{-\kappa\tau}}{\kappa} + B(\tau)$$

Considering the relationship $G = \ln(S)$, we also have:

$$\begin{cases} dG = \left(\mu - C - \frac{1}{2}\sigma_S^2\right)dt + \sigma_S dz_S \\ dC = [k(\alpha - C)]dt + \sigma_C dz_C \end{cases}$$

Then, to apply the Kalman filter, the model must be expressed under its state-spaced form. The transition equation is:

$$\begin{bmatrix} G_t \\ C_t \end{bmatrix} = c + F \times \begin{bmatrix} G_{t-1} \\ C_{t-1} \end{bmatrix} + Tw_t, \quad t = 1, \dots NT$$

where:

$$-c = \begin{bmatrix} (\mu - \frac{1}{2}\sigma_S^2)\Delta t \\ \kappa\alpha\Delta t \end{bmatrix}$$ is a (2×1) vector, and Δt is the period separating 2 observation dates

$$-F = \begin{bmatrix} 1 & -\Delta t \\ 0 & 1 - \kappa\Delta t \end{bmatrix}$$ is a (2×2) matrix,

$- T$ is an identity matrix, (2×2),

$- w_t$ is a sequence of uncorrelated errors, with:

$$E[w_t] = 0, \text{ and} \qquad Q = Var[w_t] = \begin{bmatrix} \sigma_S^2\Delta t & \rho\sigma_S\sigma_C\Delta t \\ \rho\sigma_S\sigma_C\Delta t & \sigma_C^2\Delta t \end{bmatrix}$$

The measurement equation is directly written from the solution of the model:

$$z_t = d + H \times \begin{bmatrix} G_t \\ C_t \end{bmatrix} + u_t, t = 1, \dots \text{NT}$$

where:

— The i$^\text{th}$ component of the vector z_t of size N is $\ln(F(\tau_i))$. N is the number of maturities that were used for the estimation.
— The i$^\text{th}$ component of the vector d of size N is $B(\tau_i)$
— H is the $(N \times 2)$ matrix whose i$^\text{th}$ line is $\left[1, -\dfrac{1 - e^{-\kappa\tau_i}}{\kappa}\right]$
— u_t is a white noise vector, of size N, with no serial correlation:

$$E[u_t] = 0 \text{ and } R = Var[u_t]. R \text{ is } (N \times N)$$

2.1.3 Applying the extended filter to the Schwartz model
The transition equation is directly obtained from the dynamics of the state variables:

$$\begin{bmatrix} S_t \\ C_t \end{bmatrix} = f(S_{t-1}, C_{t-1}) + T(S_{t-1}, C_{t-1})w_t$$

where:

$-\begin{bmatrix} S_t \\ C_t \end{bmatrix}$ is the state vector, (2×1),

$-f(S_{t-1}, C_{t-1})$ is a (2×1) vector:

$$f(S_{t-1}, C_{t-1}) = \begin{bmatrix} S_{t-1}(1 + \mu\Delta t - C_{t-1}\Delta t) \\ \kappa\alpha\Delta t + C_{t-1}(1 - \kappa\Delta t) \end{bmatrix}$$

$-T(S_{t-1}, C_{t-1})$ is a (2×2) matrix: $T(S_{t-1}, C_{t-1}) = \begin{bmatrix} S_{t-1} & 0 \\ 0 & 1 \end{bmatrix}$

$- Q$ is a (2×2) matrix: $\quad Var(w_t) = \begin{bmatrix} \sigma_S^2 & \rho\sigma_S\sigma_C \\ \rho\sigma_S\sigma_C & \sigma_C^2 \end{bmatrix}$

The measurement equation becomes:

$$z_t = h(S_t, C_t) + u_t$$

where:
— The i$^\text{th}$ component of the vector z_t is $F(\tau_i)$
— $h(S_t, C_t)$ is a vector whose i$^\text{th}$ component is:

$$[S_t \times \exp(-z_iC_{t/t-1} + B(\tau_i))]$$

with: $Z_i = \dfrac{1 - e^{-\kappa\tau_i}}{\kappa}$

$$B(\tau_i) = \left[\left(r - \widehat{\alpha} + \frac{\sigma_C^2}{2\kappa^2} - \frac{\sigma_S\sigma_C\rho}{\kappa}\right) \times \tau_i\right] + \left[\frac{\sigma_C^2}{4} \times \frac{1 - e^{-2\kappa\tau_i}}{\kappa^3}\right]$$

$$+ \left[\left(\widehat{\alpha}\kappa + \sigma_S\sigma_C\rho - \frac{\sigma_C^2}{\kappa}\right) \times \left(\frac{1 - e^{-\kappa\tau_i}}{\kappa^2}\right)\right]$$

$$\widehat{\alpha} = \alpha - \lambda/\kappa$$

— u_t is a white noise vector, $(N \times 1)$, with no serial correlation:

$$E[u_t] = 0 \text{ and } R = Var[u_t]. R \text{ is } (N \times N)$$

Lastly A and H, the derivatives of the functions f and h with respect to the state variables, are the following:

$-A$ is a (2×2) matrix:

$$A(S_{t-1}, C_{t-1}) = \begin{bmatrix} 1 + \mu\Delta t - C_{t-1}\Delta t & -S_{t-1}\Delta t \\ 0 & (1 - \kappa\Delta t) \end{bmatrix}$$

$-H$ is a $(N \times 2)$ matrix whose i$^\text{th}$ line is:

$$\left[e^{(-Z_iC_{t-1} + B(\tau_i))} - Z_i \times e^{(-z_iC_{t-1} + B(\tau_i))}\right]$$

$$B(\tau_i) = \left[\left(r - \widehat{\alpha} + \frac{\sigma_C^2}{2\kappa^2} - \frac{\sigma_S\sigma_C\rho}{\kappa}\right) \times \tau_i\right] + \left[\frac{\sigma_C^2}{4} \times \frac{1 - e^{-2\kappa\tau_i}}{\kappa^3}\right]$$

$$+ \left[\left(\widehat{\alpha}\kappa + \sigma_S\sigma_C\rho - \frac{\sigma_C^2}{\kappa}\right) \times \left(\frac{1 - e^{-\kappa\tau_i}}{\kappa^2}\right)\right]$$

$$\widehat{\alpha} = \alpha - \lambda/\kappa$$

2.2 Practical difficulties associated with the empirical study
To perform the empirical study, some difficulties must be overcome. First, there are choices to make when the iterative process is started. Second, if the model has been expressed under the logarithmic form for the simple Kalman filter, some precautions must be taken when the performance is being considered. Third, the stability of the iteration process and the model's performance are extremely sensitive to the covariance matrix R.

2.2.1 Starting the iterative process
To start the iterative process, there is a need for the initial values of the non-observable variables and for their covariance matrix. In the case of

term structure models of commodity prices, the nearest futures price is generally used as the spot price S, and the convenience yield C is calculated as the solution of the Brennan and Schwartz (1985) model. This solution requires the use of two observed futures prices, for delivery in T_1 and in T_2:

$$c = r - \frac{\ln(F(S, t, T_1)) - \ln(F(S, t, T_2))}{T_1 - T_2}$$

where T_1 is the nearest delivery, and T_2 is the one immediately afterwards. We choose the first value of the estimation period for the non-observable variables.

The covariance matrix associated with the state variables must also be initialized. We choose a diagonal matrix, and calculate the variances with the first 30 points of the estimation period.

2.2.2 Measuring the performance
As we have transformed the equations taking the logarithms, in order to use the simple Kalman filter, care has to be taken not to introduce a bias when transforming back the results. Indeed in that case, the innovations are calculated with the logarithms of the futures prices.

$$z_t = \hat{z}_t^- + \sigma V$$

where σ is the standard deviation of the innovations and V is a standardized Gaussian residual independent to $\hat{z}_t^-$. The exponential of $\hat{z}_t^-$ is a biased estimator of the future prices, because $e^{z_t} = e^{\hat{z}_t^-} \times e^{\sigma V}$ and taking the expectation we find:

$$E[e_t^z] = E[e^{\hat{z}_t^-}] \times e^{\frac{\sigma^2}{2}}$$

We therefore have to correct it by the exponential factor.

2.2.3 Stabilizing the iteration process
Another important choice must be made before initiating the Kalman iteration process, concerning the estimation of the covariance matrix associated with the errors introduced in the measurement equation. This matrix R is crucial for the iteration's stability, because it is added to the innovations covariance matrix, during the innovation phase. Therefore, the updating of the non-observable variable is strongly affected by the matrix R, and if the terms of this matrix are too high, the iteration can become unstable.

Most of the time, the easiest way to estimate this matrix is to calculate the variances and the covariances of the estimation database. This method was chosen to measure the model's performance and is presented in paragraph 2.3. But it is important to know how much the empirical results are affected by this choice. In order to answer this question, a few simulations will be run and analyzed.

2.3 Comparison between the two filters
The comparison between the performance of the Schwartz model measured with the two filters allows us to appreciate the influence of the

linearization on the results. We first present the empirical data. Then the performance criteria are exposed. Finally, the results are delivered and commented upon.

2.3.1 The empirical data
The data used for the empirical study corresponds to daily crude oil prices for the settlement of the Nymex's WTI futures contracts, between the 18th of May 1998, and the 15th of October 2001. They have been arranged such that the first futures price's maturity τ_1 is actually the one month maturity, and the second futures price's corresponds to the two months maturity $\tau_2,\dots$ Keeping the first observation of each group of five, these daily data points were transformed into weekly data. For the parameters estimation, and for the measure of the model's performance, four series of futures prices[11] were retained, corresponding to the one, the three, the six and the nine months maturities. The interest rates are set to the three-month T-bill rates. Because interest rates are supposed to be constant in the model, we use the mean of all the observations between 1995 and 2002.

2.3.2 The performance criteria
To measure the model's performance, two criteria were retained: the mean pricing errors and the root mean squared errors.

The mean pricing errors (MPE) are defined in the following way:

$$MPE = \frac{1}{N} \sum_{n=1}^{N} (\hat{F}(n, \tau) - F(n, \tau))$$

where N is the number of observations, $\hat{F}(n, \tau)$ is the estimated futures price for a maturity τ at time-step n, and $F(n, \tau)$ is the observed futures price.

Retaining the same notations, the root mean squared error (RMSE), is defined in the following way, for one given maturity τ:

$$RMSE = \sqrt{\frac{1}{N} \sum_{n=1}^{N} (\hat{F}(n, \tau) - F(n, \tau))^2}$$

2.3.3 The empirical results
First, the optimal parameters obtained with the two filters are compared. Then the model's ability to represent the price curve and its dynamics is appreciated. Finally, the sensitivity of the results to the errors covariance matrix is presented.

2.3.3.1 Optimal parameters[12]
The differences between the optimal parameters obtained with the two filters show that the linearization has a significant influence. Nevertheless, the parameters have the same order size as those Schwartz obtained in 1997 on the crude oil market, on different periods.

2.3.3.2 The model's performance
The simple filter is always more precise than the extended one. This is true for all the maturities[14]. These measures also always decrease with maturity,

TABLE 1. OPTIMAL PARAMETERS, 1998–2001[13]

	Simple filter		Extended filter	
	Parameters	Gradients	Parameters	Gradients
Pull back force: κ	1.59171	−0.003631	1.258133	0.000628
Trend: μ	0.379926	0.000497	0.352014	−0.001178
Spot price's volatility: σ_s	0.263525	−0.000448	0.320235	−0.000338
Long run mean: α	0.252260	−0.012867	0.232547	0.004723
Convenience yield's volatility: σ_c	0.237071	−0.000602	0.288427	−0.001070
Correlation coefficient: ρ	0.938487	−0.001533	0.969985	0.000008
Risk premium: λ	0.177159	0.009272	0.181955	−0.002426

TABLE 2. THE MODEL'S PERFORMANCE WITH THE SIMPLE AND THE EXTENDED FILTERS, 1998–2001

	Simple filter		Extended filter	
Maturity	MPE	RMSE	MPE	RMSE
1 month	−0.060423	2.319730	0.09793	2.294503
3 months	−0.107783	1.989428	0.057327	2.120727
6 months	−0.054536	1.715223	0.109584	1.877654
9 months	−0.007316	1.567467	0.141204	1.695222
Average	−0.057514	1.897962	0.101511	1.997027

Unit: USD/b.

TABLE 3. THE COMPARISON BETWEEN THE MODEL'S PERFORMANCE ASSOCIATED WITH THE SIMPLE FILTER, WITH AND WITHOUT CORRECTIONS FOR THE LOGARITHM, 1998–2001

	Simple filter		Simple filter corrected	
Maturity	MPE	RMSE	MPE	RMSE
1 month	−0.060423	2.319730	0.065644	2.314178
3 months	−0.107783	1.989428	0.006419	1.981453
6 months	−0.054536	1.715223	0.026010	1.709931
9 months	−0.007316	1.567467	0.061301	1.564854
Average	−0.057514	1.897962	0.036637	1.892604

Unit: USD/b.

FIGURE 1: Innovations, 1998–2001

which is consistent with Schwartz's results on other periods. Nevertheless, Schwartz has worked with longer maturities, and shown that the root mean squared error increases again for deliveries after 15 months.

To be rigorous, the model's performance associated with the simple Kalman filter should be corrected when, as is the case here, the logarithms of the estimations are used to obtain the estimations themselves. The correction slightly improves the performance, as shown in table 3: the root mean squared errors and the mean pricing errors diminish a bit for almost all the maturities.

Finally, the innovations range diminishes with the futures contracts maturity. Figure 1 illustrates the innovations behavior for the one-month maturity case. It shows that they tend to return to zero for the two filters. Nevertheless as the figure illustrates it, even if the mean pricing errors are low for the two filters, the pricing errors can be rather important at certain specific dates. They reach USD 6 in absolute value for the extended filter, which represents 25% of the mean futures price for the one-month maturity case. It is a bit lower for the simple filter: USD 5.

Consequently we can say that there is clearly an impact of the linearization introduced in the extended filter: it can be observed from the optimal parameters, from the performance, and from the innovations.
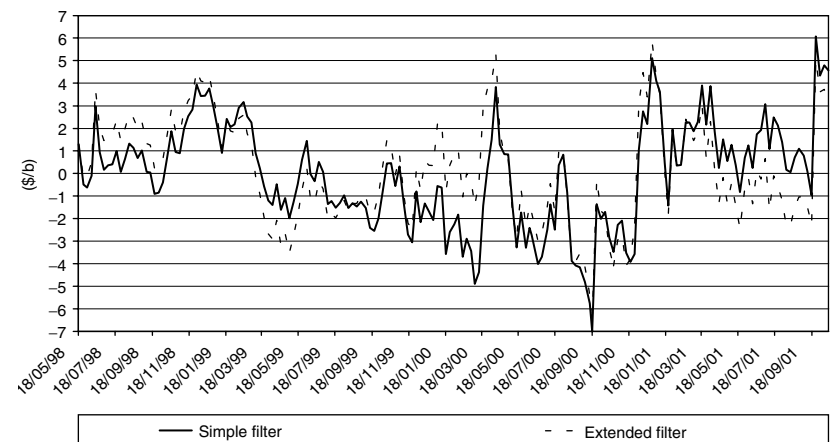
Nevertheless with an extended filter, the model's ability to represent the price curve is still quite good for this case.

Another way to assess a model's performance is to see if it is able to reproduce the price dynamics, which can be shown graphically.

Considering this, the first important conclusion is that the model is able to reproduce the price dynamics quite precisely, even if as in 1998—2001, there are very large fluctuations in the futures prices. Figure 2 shows the results obtained for the one-month maturity case. During that period, the crude oil price goes from USD 11 per barrel to USD 37 per barrel! Even if the Kalman filters are often suspected to be unstable, these results show that they can be used even with extremely volatile data. The graph also shows that the two filters attenuate the range of
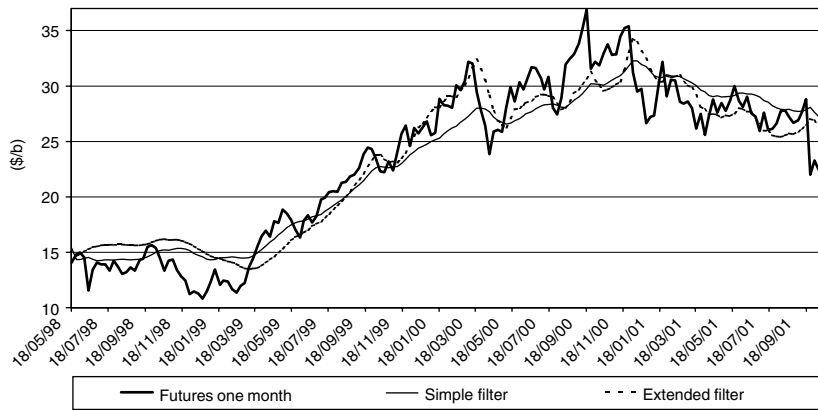
FIGURE 2: Estimated and observed futures prices for the one month maturity, 1998–2001

price fluctuations. This phenomenon can actually be observed for every maturity.

2.3.3.3 Simulations

The last results presented in this sesction are simulations. They show how the model's performance is affected by the choice of the system's matrix R. This matrix represents the errors in the measurement equation and the way it is estimated has a strong influence on the empirical results.

TABLE 4. SIMULATIONS WITH DIFFERENT SYSTEM'S MATRIX *R*

	1 month	3 months	6 months	9 months	Average
Observations					
MPE	0.0979	0.0573	0.1096	0.1412	**0.1015**
RMSE	2.2945	2.1207	1.8777	1.6952	**1.9970**
Simulation 1					
MPE	0.0013	0.0935	0.1501	1.6506	**0.4739**
RMSE	1.8356	1.5405	1.2478	2.6602	**1.8210**
Simulation 2					
MPE	0.0073	0.0152	0.0612	0.0137	**0.0244**
RMSE	1.4759	1.1686	0.9386	0.8317	**1.1037**
Simulation 3					
MPE	0.0035	−0.0003	0.0383	0.0005	**0.0105**
RMSE	1.3812	1.0950	0.8647	0.7499	**1.0227**
Simulation 4					
MPE	0.0131	0.0067	0.0415	0.0075	**0.0172**
RMSE	1.3602	1.0919	0.8697	0.7591	**1.0202**

Most of the time, the terms of this matrix correspond to the variances and the covariances of the estimation database, namely, in the case studied here, the variances and covariances between futures prices for different maturities. But one must know that the results obtained with the Kalman filter can be more precise if these terms are (artificially) lowered, as shown in table 4. This table exposes the different results obtained during 1998–2001 with the extended Kalman filter. This period is especially interesting because the data strongly fluctuates. The performance is achieved, first with the matrix based on the observations, then with the artificially lowered one.

Simulations 1 to 4 correspond respectively to the model's performance obtained by multiplying the system's matrix R by (1/2), (1/16), (1/160), and (1/1600). As the matrix elements are lowered, the model's performance strongly improves: from the initial simulation to the fourth one, the root-mean-squared-error is almost divided by two. The comparison between the third and the fourth simulations also illustrates the fact that there is a limit to the performance amelioration. Figure 3 portrays the main results of these simulations.
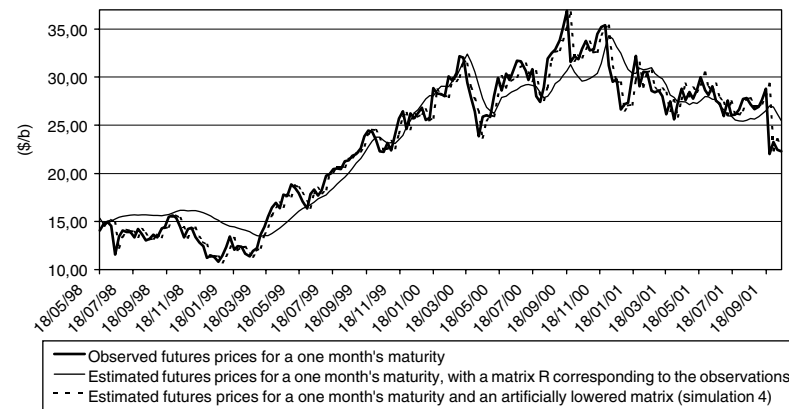

FIGURE 3: One month futures prices observed/estimated, 1998–2001

2.4 Conclusion

The main conclusions of this section are the following: Firstly, the approximation introduced in the extended Kalman filter has clearly an influence on the model's performance; the extended filter generally leads to less precise estimations than the simple one. Nevertheless, the difference between the two filters is quite low and the extended filter is still acceptable. Secondly, the estimation results are sensitive to the system's matrix containing the errors of the measurement equation, and this matrix can be used to obtain more precise results on the estimation database. Thirdly, the approximation made in the extended Kalman filter is not a real problem until the model starts to become highly nonlinear. The following synthetic example illustrates the situation.

Let x follow an Ornstein-Uhlenbeck or mean-reverting diffusion, like the convenience yield in the Schwartz model, and z be an exponential:

$$z = exp(ax) + 0.4\,N(0, 1)$$

The advantage of the example chosen is that we can easily simulate x and it is therefore easy to assess the performance of each version of the Kalman filter. It also allows us to make the nonlinearity vary with a.

We first illustrate the effect of an increasing nonlinearity by making the parameter a vary from 0.2 to 1. The statistics for the expectation and the variance of the errors have been obtained on a series of 2000 points.

Table 5 shows clearly that when a is small, the bias is small as well. The latter increases with a. The estimation of the variance is quite high when the values of a are smaller than or equal to 0.4, because the relative weight of the noise is large. These estimations of the variance also increase with the nonlinearity.

To illustrate the performance of the filters and for a better understanding of the bias, we show on Figure 4 the 200 first points of the

TABLE 5. PERFORMANCE OF THE TWO FILTERS

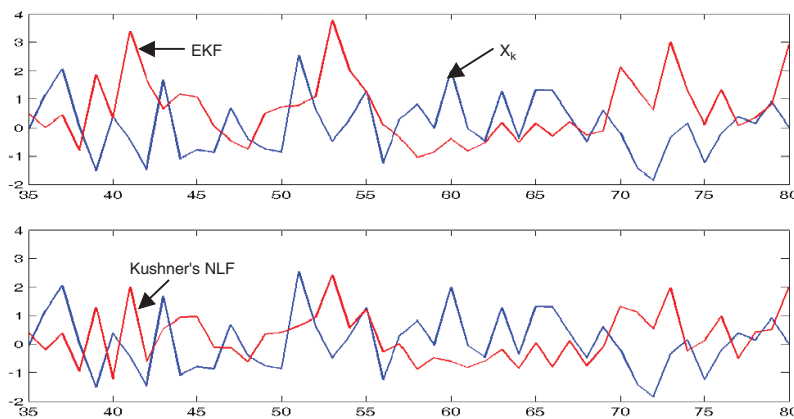|	EKF		Kushner's NLF	
a	E(X-X*)	Var(X-X*)	E(X-X*)	Var(X-X*)
0.2	0.01	0.94	0.02	0.92
0.4	−0.1	0.65	−0.06	0.60
0.6	−0.23	0.47	−0.11	0.41
0.8	0.4	0.62	0.14	0.31
1	0.68	1.67	−0.12	0.26

FIGURE 4: Real process versus its estimates: Top EKF, bottom Kushner's NLF

initial series and the estimations we obtain with the two versions of the Kalman filter when $a = 0.8$.

3 Stochastic Volatility Models

In this section, we apply the different filters to a few stochastic volatility models including the Heston, the GARCH and the 3/2 models. To test the performance of each filter, we use five years of S&P500 time-series.

The idea of applying the Kalman Filter to Stochastic Volatility models goes back to Harvey, Ruiz & Shephard (1994), where the authors attempt to determine the system parameters via a QML Estimator. This approach has the obvious advantage of simplicity, however it does not account for the nonlinearities and non-Gaussianities of the system.

More recently, Pitt & Shephard (Doucet, De Freitas and Gordon (2001)) suggested the use of *Auxiliary* Particle Filters to overcome some of these difficulties. An alternative method based upon the Fourier transform has been presented in Bates (2002).

3.1 The State Space Model
Let us first present the state-space form of the stochastic volatility models:

3.1.1 The Heston Model
Let us study the Euler-discretized Heston (1993) Stochastic Volatility model in a risk-neutral framework[15]

$$lnS_k = lnS_{k-1} + (r_{k-1} - \tfrac{1}{2}v_{k-1})\Delta t + \sqrt{v_{k-1}}\sqrt{\Delta t}\,B_{k-1} \tag{31}$$

$$v_k = v_{k-1} + (\omega - \theta v_{k-1})\Delta t + \xi\sqrt{v_{k-1}}\sqrt{\Delta t}\,Z_{k-1} \tag{32}$$

where S_k is the stock price at time-step k, Δt the time interval, r_k the risk-free rate of interest (possibly netted by a dividend-yield), v_k the stock variance and B_k, Z_k two sequences of temporally-uncorrelated Gaussian random-variables with a mutual correlation ρ. The model risk-neutral parameter-set is therefore $\Psi = (\omega, \theta, \xi, \rho)$.

Considering v_k as the hidden state and lnS_{k+1} as the observation, we can subtract from both sides of the transition equation $x_k = f(x_{k-1}, w_{k-1})$, a multiple of the quantity $h(x_{k-1}, u_{k-1}) - z_{k-1}$ which is equal to zero. This would allow us to eliminate the correlation between the system and the measurement noises.

Indeed, if the system equation is

$$v_k = v_{k-1} + (\omega - \theta v_{k-1})\Delta t + \xi\sqrt{v_{k-1}}\sqrt{\Delta t}\,Z_{k-1} - \rho\xi\,[lnS_{k-1}$$
$$+ (r_k - \tfrac{1}{2}v_{k-1})\Delta t + \sqrt{v_{k-1}}\sqrt{\Delta t}\,B_{k-1} - lnS_k]$$

posing for every k

$$\tilde{Z}_k = \frac{1}{\sqrt{1 - \rho^2}}(Z_k - \rho B_k)$$

we will have as expected $\tilde{Z}_k$ *uncorrelated* with B_k and

$$x_k = v_k = v_{k-1} + [\omega - \rho\xi r_k - \left(\theta - \frac{1}{2}\rho\xi\right)v_{k-1}]\Delta t$$
$$+ \rho\xi \ln\left(\frac{S_k}{S_{k-1}}\right) + \xi\sqrt{1-\rho^2}\sqrt{v_{k-1}}\sqrt{\Delta t}\,\tilde{Z}_{k-1} \qquad (33)$$

and the measurement equation would be

$$z_k = \ln S_{k+1} = \ln S_k + (r_k - \frac{1}{2}v_k)\Delta t + \sqrt{v_k}\sqrt{\Delta t}\,B_k \qquad (34)$$

3.1.2 Other Stochastic Volatility Models
It is easy to generalize the above state space model to other stochastic volatility approaches. Indeed we could replace (32) with

$$v_k = v_{k-1} + (\omega - \theta v_{k-1})\Delta t + \xi v_{k-1}^p \sqrt{\Delta t}\, z_{k-1} \qquad (35)$$

where $p = 1/2$ would naturally correspond to the Heston (Square-Root) model, $p = 1$ to the GARCH diffusion-limit model, and $p = 3/2$ to the 3/2 model. These models have all been described and analyzed in Lewis (2000).

The new state transition equation would therefore become

$$v_k = v_{k-1} + \left[\omega - \rho\xi r_k v_{k-1}^{p-\frac{1}{2}} - \left(\theta - \frac{1}{2}\rho\xi v_{k-1}^{p-\frac{1}{2}}\right)v_{k-1}\right]\Delta t$$
$$+ \rho\xi v_{k-1}^{p-\frac{1}{2}}\ln\left(\frac{S_k}{S_{k-1}}\right) + \xi\sqrt{1-\rho^2}v_{k-1}^p\sqrt{\Delta t}\,\tilde{Z}_{k-1} \qquad (36)$$

where the same choice of state space $x_k = v_k$ is made.

3.1.3 Robustness and Stability
In this state-space formulation, we only need to choose a value for v_0 which could be set to the historic-volatility over a period preceding our time-series. Ideally, the choice of v_0 should not affect the results enormously, *i.e.* we should have a robust system.

As we saw in the previous section, the system stability greatly depends on the measurement noise. However in this case the system noise is precisely $\sqrt{\Delta t}v_k$, and therefore we do not have a direct control on this issue. Nevertheless we could add an independent Normal measurement noise with a given variance R

$$z_k = \ln S_{k+1} = \ln S_k + (r_k - \frac{1}{2}v_k)\Delta t + \sqrt{v_k}\sqrt{\Delta t}B_k + RW_k$$

which would allow us to *tune* the filter.[16]

3.2 The Filters
We can now apply the Gaussian and the Particle Filters to our problem:

3.2.1 Gaussian Filters
For the EKF we will have

$$A_k = 1 - \left[\rho\xi r_k\left(p - \frac{1}{2}\right)v_{k-1}^{p-\frac{3}{2}} + \theta - \frac{1}{2}\rho\xi\left(p + \frac{1}{2}\right)v_{k-1}^{p-\frac{1}{2}}\right]\Delta t$$
$$+ \left(p - \frac{1}{2}\right)\rho\xi v_{k-1}^{p-\frac{3}{2}}\ln\left(\frac{S_k}{S_{k-1}}\right)$$

and

$$W_k = \xi\sqrt{1-\rho^2}v_{k-1}^p\sqrt{\Delta t}$$

as well as

$$H_k = -\frac{1}{2}\Delta t$$

and

$$U_k = \sqrt{v_k}\sqrt{\Delta t}$$

The same time update and measurement update equations could be used with the UKF or Kushner's NLF.

3.2.2 Particle Filters
We could also apply the Particle Filtering algorithm to our problem. Using the same notations as in section 1.3.2 and calling

$$\mathbf{n}(x, \mathbf{m}, \mathbf{s}) = \frac{1}{\sqrt{2\pi}\mathbf{s}}exp(-\frac{(x-\mathbf{m})^2}{2\mathbf{s}^2})$$

the Normal density with mean $\mathbf{m}$ and standard deviation $\mathbf{s}$, we will have

$$q(\tilde{x}_k^{(i)}|x_{k-1}^{(i)}, z_{1:k}) = \mathbf{n}\left(\tilde{x}_k^{(i)}, \mathbf{m} = \hat{x}_k^{(i)}, \mathbf{s} = \sqrt{P_k^{(i)}}\right)$$

as well as

$$p(z_k|\tilde{x}_k^{(i)}) = \mathbf{n}\left(z_k, \mathbf{m} = z_{k-1} + (r_k - \frac{1}{2}\tilde{x}_k^{(i)})\Delta t, \mathbf{s} = \sqrt{\tilde{x}_k^{(i)}}\sqrt{\Delta t}\right)$$

and

$$p(\tilde{x}_k^{(i)}|x_{k-1}^{(i)}) = \mathbf{n}\left(\tilde{x}_k^{(i)}, \mathbf{m_x}, \mathbf{s} = \xi\sqrt{1-\rho^2}(x_{k-1}^{(i)})^p\sqrt{\Delta t}\right)$$

with

$$\mathbf{m}_x = x_{k-1}^{(i)} + \left[\omega - \rho\xi r_k(x_{k-1}^{(i)})^{p-\frac{1}{2}} - \left(\theta - \frac{1}{2}\rho\xi(x_{k-1}^{(i)})^{p-\frac{1}{2}}\right)x_{k-1}^{(i)}\right]\Delta t$$
$$+ \rho\xi(x_{k-1}^{(i)})^{p-\frac{1}{2}}(z_{k-1} - z_{k-2})$$

and as before we have

$$w_k^{(i)} = w_{k-1}^{(i)}\frac{p(z_k|\tilde{x}_k^{(i)})p(\tilde{x}_k^{(i)}|x_{k-1}^{(i)})}{q(\tilde{x}_k^{(i)}|x_{k-1}^{(i)}, z_{1:k})}$$

which provides us with what we need for the filter implementation.

3.3 Parameter Estimation and Back-Testing

For the Gaussian MLE we will need to minimize

$$\phi(\omega, \theta, \xi, \rho) = \sum_{k=1}^{K} \left[ln(F_k) + \frac{v_k^2}{F_k} \right]$$

with $v_k = \mathbf{z}_k - \hat{\mathbf{z}}_k^-$ and $\mathbf{F}_k = \mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^t + \mathbf{U}_k \mathbf{U}_k^t$ for the EKF.

For the Particle MLE, as previously mentioned, we need to maximize

$$\sum_{k=1}^{N} ln \left(\sum_{i=1}^{N_{sims}} w_k^{(i)} \right)$$

By maximizing the above likelihood functions, we will find the optimal parameter-set $\hat{\Psi} = (\hat{\omega}, \hat{\theta}, \hat{\xi}, \hat{\rho})$. This calibration procedure could then be used for pricing of derivatives instruments or forecasting volatility.

We could perform a back-testing process in the following way. Choosing an original parameter-set

$$\Psi^* = (0.02, 0.5, 0.05, -0.5)$$

and using a Monte-Carlo simulation, we generate an artificial time-series. We take $S_0 = 1000$ USD, $v_0 = 0.04$, $r_k = 0.027$ and $\Delta t = 1$. Note that we are taking a large Δt in order to have meaningful errors. The time-series is generated via the above transition equation (32) and the usual log-normal measurement equation.

We find the following optimal[17] parameter-sets:

$$\hat{\Psi}_{EKF} = (0.036335, 0.928632, 0.036008, -1.000000)$$

$$\hat{\Psi}_{UKF} = (0.033104, 0.848984, 0.033263, -0.983985)$$

$$\hat{\Psi}_{EPF} = (0.019357, 0.500021, 0.033354, -0.581794)$$

$$\hat{\Psi}_{UPF} = (0.019480, 0.489375, 0.047030, -0.229242)$$

which show the better performance of the Particle Filters. However, it should be reminded that the Particle Filters are also more computation-intensive than the Gaussian ones.

3.4 Application to the S&P500 Index

The above filters were applied to five years of S&P500 time-series (1996 to 2001) in Javaheri (2002) and the filtering errors were considered for the Heston model, the GARCH model and the 3/2 model. Daily index close-prices were used for this purpose, and the time interval was set to $\Delta t = 1/252$. The appropriate risk-free rate was applied and was adjusted by the index dividend yield at the time of the measurement.

As in the previous section, the performance could be measured via the MPE and the RMSE. We could also refer to figures 5 to 11 for a visual interpretation of the performance measurements.

$$MPE_{EKF-Heston} = 3.58207e - 05 \quad RMSE_{EKF-Heston} = 1.83223e - 05$$

$$MPE_{EKF-GARCH} = 2.78438e - 05 \quad RMSE_{EKF-GARCH} = 1.42428e - 05$$

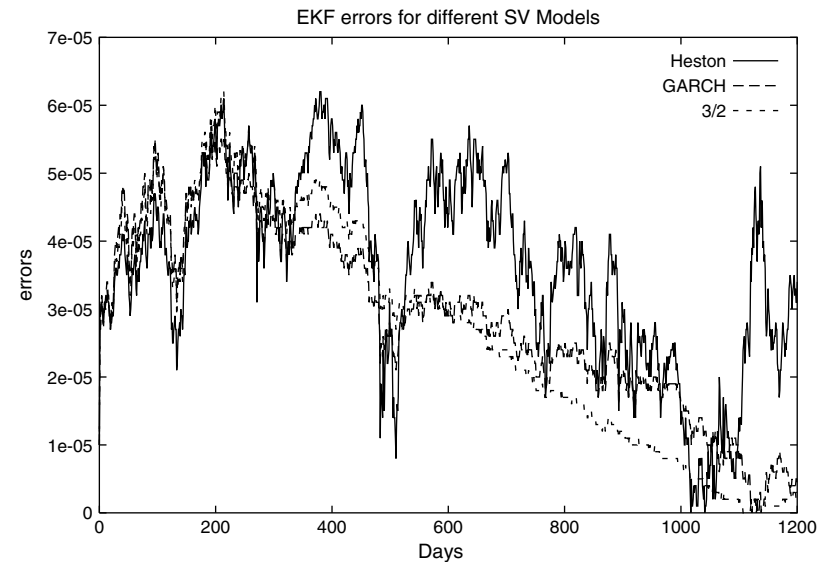$$MPE_{EKF-\frac{3}{2}} = 2.63227e - 05 \quad RMSE_{EKF-\frac{3}{2}} = 1.74760e - 05$$

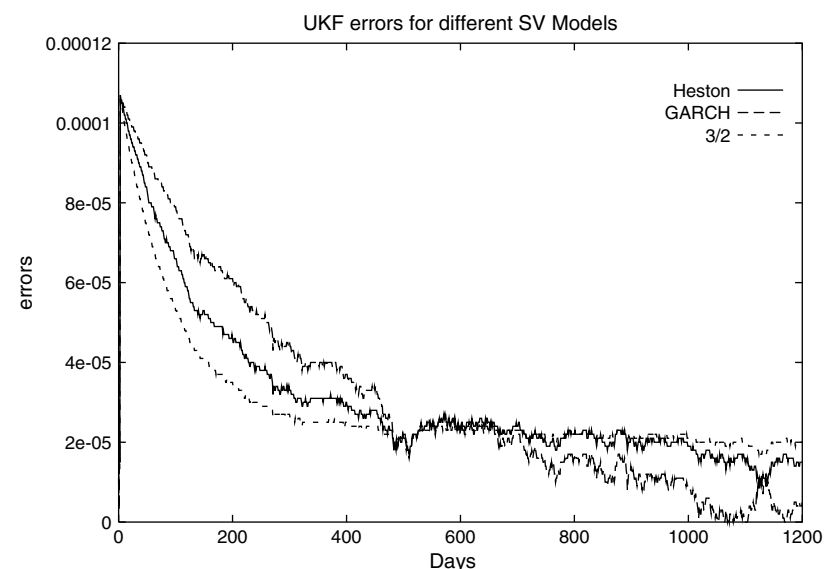**FIGURE 5: Comparison of EKF Filtering errors for Heston, GARCH and 3/2 Models**

FIGURE 6: Comparison of UKF Filtering errors for Heston, GARCH and 3/2 Models
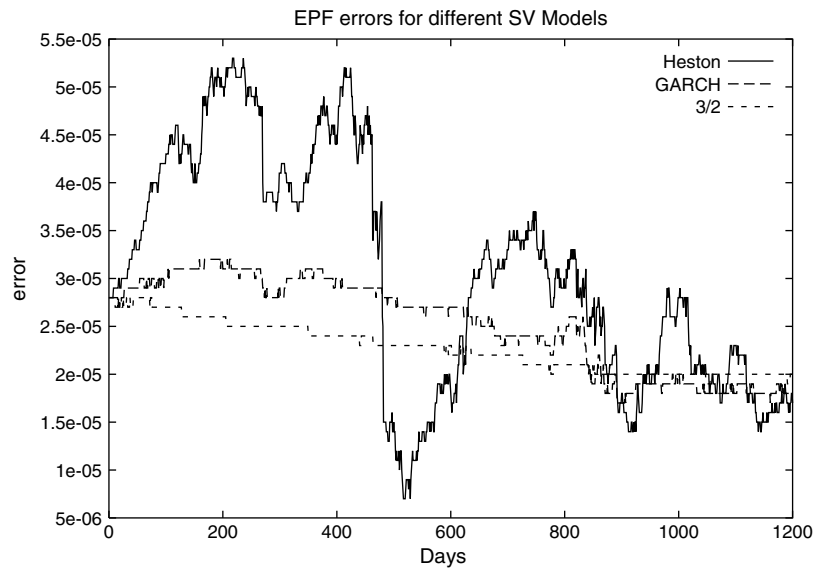
FIGURE 7: Comparison of EPF Filtering errors for Heston, GARCH and 3/2 Models
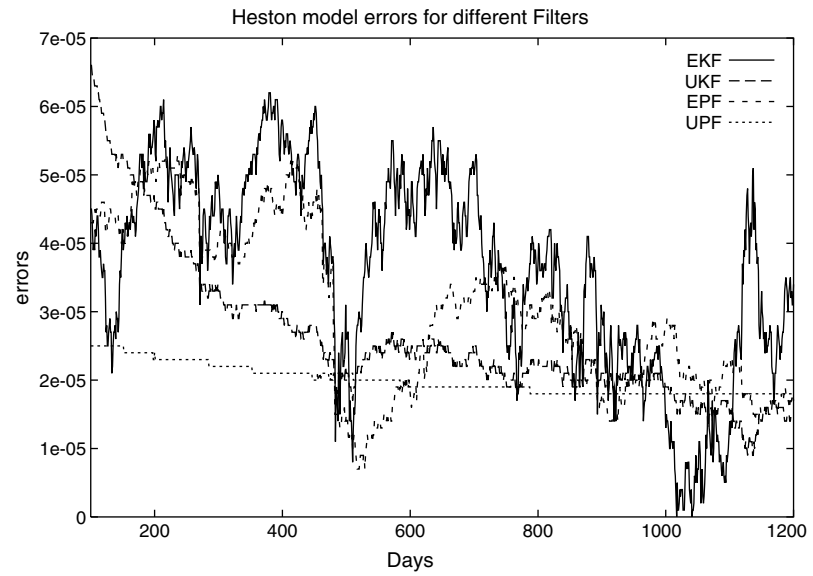

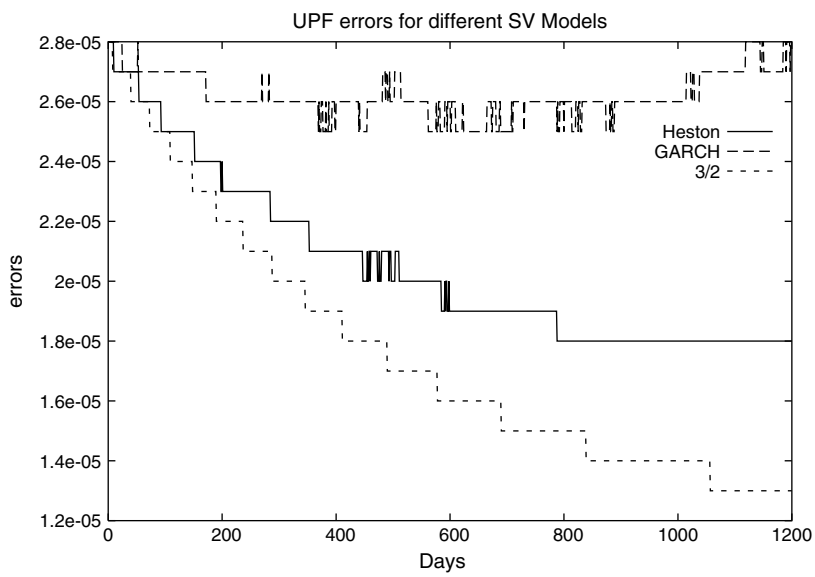FIGURE 9: Comparison of Filtering errors for the Heston Model

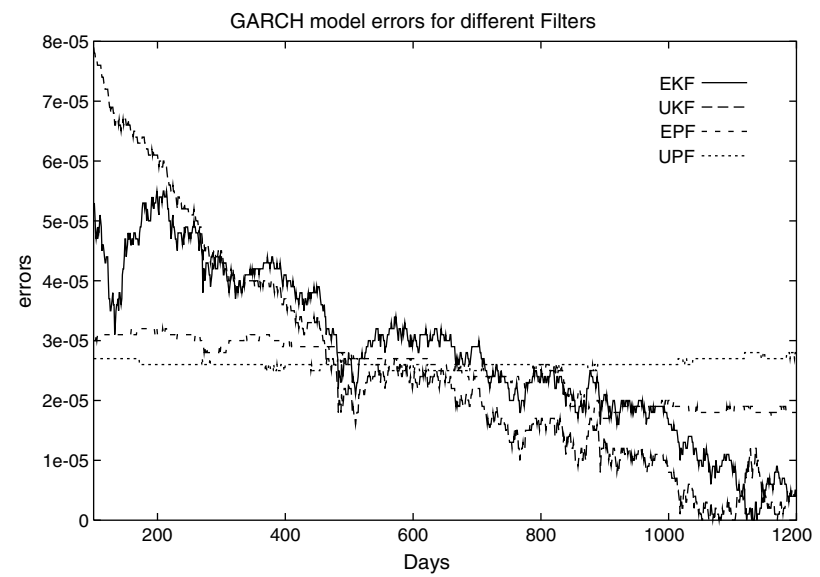FIGURE 8: Comparison of UPF Filtering errors for Heston, GARCH and 3/2 Models

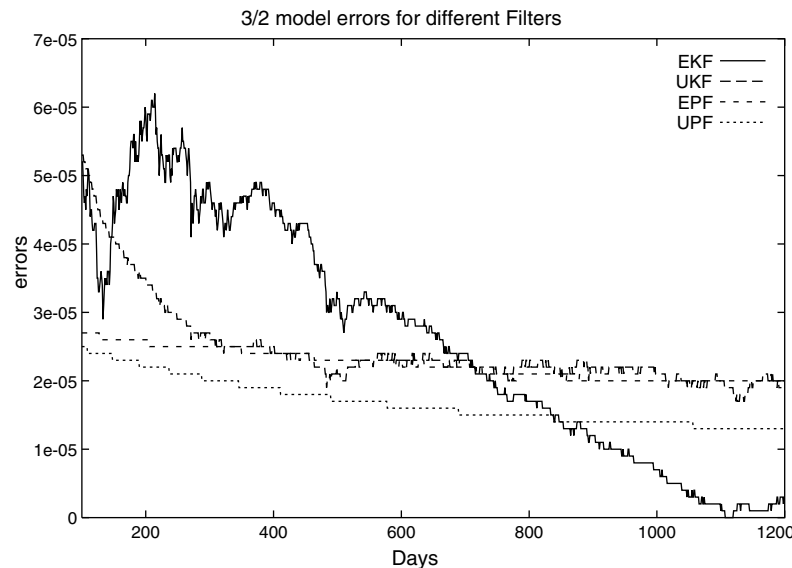FIGURE 10: Comparison of Filtering errors for the GARCH Model

3/2 model errors for different Filters

FIGURE 11: Comparison of Filtering errors for the 3/2 Model

$$MPE_{UKF-Heston} = 3.00000e - 05 \quad RMSE_{UKF-Heston} = 1.91280e - 05$$
$$MPE_{UKF-GARCH} = 2.99275e - 05 \quad RMSE_{UKF-GARCH} = 2.58131e - 05$$
$$MPE_{UKF-\frac{3}{2}} = 2.82279e - 05 \quad RMSE_{UKF-\frac{3}{2}} = 1.55777e - 05$$

$$MPE_{EPF-Heston} = 2.70104e - 05 \quad RMSE_{EPF-Heston} = 1.34534e - 05$$
$$MPE_{EPF-GARCH} = 2.48733e - 05 \quad RMSE_{EPF-GARCH} = 4.99337e - 06$$
$$MPE_{EPF-\frac{3}{2}} = 2.26462e - 05 \quad RMSE_{EPF-\frac{3}{2}} = 2.58645e - 06$$

$$MPE_{UPF-Heston} = 2.04000e - 05 \quad RMSE_{UPF-Heston} = 2.74818e - 06$$
$$MPE_{UPF-GARCH} = 2.63036e - 05 \quad RMSE_{UPF-GARCH} = 8.44030e - 07$$
$$MPE_{UPF-\frac{3}{2}} = 1.73857e - 05 \quad RMSE_{UPF-\frac{3}{2}} = 4.09918e - 06$$

Two immediate observations can be made: On the one hand the Particle Filters have a better performance than the Gaussian ones, which reconfirms what one would anticipate. On the other hand for most of the Filters, the 3/2 model seems to outperform the Heston model, which is in line with the findings of Engle & Ishida (2002).

3.5 Conclusion
Using the Gaussian or Particle Filtering techniques, it is possible to estimate the stochastic volatility parameters from the underlying asset time-series, in the risk-neutral or real-world context.

As expected, we observe an improvement when Particle Filters are used. What is more, given the tests carried out on the S&P500 data it seems that, despite its vast popularity, the Heston model does not perform as well as the 3/2 representation.

This suggests further research on other existing models such as Jump Diffusion (Merton (1976)), Variance Gamma (Madan, Carr and Chang (1998)) or CGMY (Carr, Geman, Madan and Yor (2002)). Clearly, because of the non-Gaussianity of these models, the Particle Filtering technique will need to be applied to them[18].

Finally it would be instructive to compare the risk-neutral parameter-set obtained from the above *time-series* based approaches, to the parameter-set resulting from a *cross-sectional* approach using options market prices at a given point in time[19]. Inconsistent parameters between the two approaches would signal either arbitrage opportunities in the market, or a misspecification in the tested model[20].

4 Summary
In this article, we present an introduction to various filtering algorithms: the Kalman filter, the Extended Filter (EKF), as well as the Unscented Kalman Filter (UKF) similar to Kushner's Nonlinear filter. We also tackle the subject of Non-Gaussian filters and describe the Particle Filtering (PF) algorithm.

We then apply the filters to a term structure model of commodity prices. Our main results are the following: Firstly, the approximation introduced in the Extended filter has an influence on the model performances. Secondly, the estimation results are sensitive to the system matrix containing the errors of the measurement equation. Thirdly, the approximation made in the extended filter is not a real issue until the model becomes highly nonlinear. In that case, other nonlinear filters such as those described in section 1.2 may be used.

Lastly, the application of the filters to stochastic volatility models shows that the Particle Filters perform better than the Gaussian ones, however they are also more expensive. What is more, given the tests carried out on the S&P500 data it seems that, despite its vast popularity, the Heston model does not perform as well as the 3/2 representation. This suggests further research on other existing models such as Jump Diffusion, Variance Gamma or CGMY. Clearly, because of the non-Gaussianity of these models, the Particle Filtering technique will need to be applied to them.

Appendix
The Measurement Update equation is

$$p(x_k|z_{1:k}) = \frac{p(z_k|x_k)p(x_k|z_{1:k-1})}{p(z_k|z_{1:k-1})}$$

where the denominator $p(z_k|z_{1:k-1})$ could be written as

$$p(z_k|z_{1:k-1}) = \int p(z_k|x_k)p(x_k|z_{1:k-1})dx_k$$

and corresponds to the Likelihood Function for the time-step k.

Indeed using the Bayes rule and the Markov property, we have

$$
\begin{aligned}
p(x_k|z_{1:k}) &= \frac{p(z_{1:k}|x_k)p(x_k)}{p(z_{1:k})} \\
&= \frac{p(z_k, z_{1:k-1}|x_k)p(x_k)}{p(z_k, z_{1:k-1})} \\
&= \frac{p(z_k|z_{1:k-1}, x_k)p(z_{1:k-1}|x_k)p(x_k)}{p(z_k|z_{1:k-1})p(z_{1:k-1})} \\
&= \frac{p(z_k|z_{1:k-1}, x_k)p(x_k|z_{1:k-1})p(z_{1:k-1})p(x_k)}{p(z_k|z_{1:k-1})p(z_{1:k-1})p(x_k)} \\
&= \frac{p(z_k|x_k)p(x_k|z_{1:k-1})}{p(z_k|z_{1:k-1})}
\end{aligned}
$$

Note that $p(x_k|z_k)$ is proportional to

$$exp\left(-\tfrac{1}{2}(z_k - h(x_k))^t R_k^{-1}(z_k - h(x_k))\right)$$

under the hypothesis of additive measurement noises.

FOOTNOTES & REFERENCES

1. Some prefer to write $x_k = f(x_{k-1}, w_{k-1})$. Needless to say, the two notations are equivalent.

2. The square-root can be computed via a Cholesky factorization combined with a Singular Value Decomposition. See Press et al. (1997) for details.

3. This analogy between Kushner's NLF and the UKF has been studied in Ito and Xiong (2000).

4. A description of the Gaussian Quadrature could be found in Press et al. (1997).

5. See Harvey (1989) for an explanation on Smoothing.

6. See for instance Arulampalam et al. (2002) or Van der Merwe, et al. (2000) for details.

7. The same exact methodology could be used in a non risk-neutral setting. We are supposing we have a small time-step Δt in order to be able to apply the Girsanov theorem.

8. In that model, interest rates are supposed to be constant.

9. This means that N = 4 in the case we study.

10. In the whole empirical study, optimizations have been made with a precision of 1e-5 on the gradients.

11. For the two filters, the parameters values retained to initiate the optimization are the same. These values are the following: $\kappa = 0.5$; $\mu = 0.1$; $\sigma_S = 0.3$; $\alpha = 0.1$; $\sigma_C = 0.4$; $\rho = 0.5$; $\lambda = 0.1$.

12. The MPE and the RMSE presented here can not be directly compared with those proposed by Schwartz in 1997 because his computations were done using the logarithm of the futures prices.

13. An alternative approach would consist in choosing a volatility *proxy* such as f(ln S_k, ln S_{k+1}) = ln S_{k+1} — ln S_k and ignoring the stock-drift term, given that Δt = o($\sqrt{\Delta t}$). We would therefore write z_k = ln(|f(ln S_k, ln S_{k+1})|) = E[ln(|f(ln$S^\star_k$, ln $S^\star_{k+1}$)|)] + $\tfrac{1}{2}$ ln v_k + ϵ_k with $S_t^\star$ the same process as S_t but with a volatility of *one*, and ϵ_k corresponding to the measurement noise. See Alizadeh, Brandt and Diebold (2002) for details.

14. In this section, all optimizations were made via the Direction-Set algorithm as described in Press et al. (1997). The precision was set to 1.0e-6.

15. A study on Filtering and Lévy processes has recently been done in Barndorff—Nielsen and Shephard N. (2002).

16. This idea is explored in Aït-Sahalia, Wang and Yared (2001) in a Non-parametric fashion.

17. This comparison supposes that the Girsanov theorem *is* applicable to the tested model.

■ Aït-Sahalia Y., Wang Y., Yared F. (2001) "Do Option Markets Correctly Price the Probabilities of Movement of the Underlying Asset?" *Journal of Econometrics*, 101

■ Alizadeh S., Brandt M.W., Diebold F.X. (2002) "Range-Based Estimation of Stochastic Volatility Models" *Journal of Finance*, Vol. 57, No. 3

■ Anderson B. D. O., Moore J. B. (1979) "Optimal Filtering" *Englewood Cliffs, Prentice Hall*

■ Arulampalam S., Maskell S., Gordon N., Clapp T. (2002) "A Tutorial on Particle Filters for On-line Nonlinear/Non-Gaussian Bayesian Tracking" *IEEE Transactions on Signal Processing*, Vol. 50, No. 2

■ Babbs S. H., Nowman K. B. (1999) "Kalman Filtering of Generalized Vasicek Term Structure Models" *Journal of Financial and Quantitative Analysis*, Vol. 34, No. 1

■ Barndorff-Nielsen O. E., Shephard N. (2002) "Econometric Analysis of Realized Volatility and its use in Estimating Stochastic Volatility Models" *Journal of the Royal Statistical Society*, Series B, Vol. 64

■ Bates D. S. (2002) "Maximum Likelihood Estimation of Latent Affine Processes" *University of Iowa & NBER*

■ Brennan M.J., Schwartz E.S. (1985) "Evaluating Natural Resource Investments" *The Journal of Business*, Vol. 58, No. 2

■ Carr P., Geman H., Madan D., Yor M. (2002) "The Fine Structure of Asset Returns" *Journal of Business*, Vol. 75, No. 2

■ Doucet A., De Freitas N., Gordon N. *(Editors)* (2001) "Sequential Monte-Carlo Methods in Practice" *Springer-Verlag*

■ Engle R. F., Ishida I. (2002) "The Square-Root, the Affine, and the CEV GARCH Models" *Working Paper, New York University and University of California San Diego*

■ Harvey A. C. (1989) "Forecasting, Structural Time Series Models and the Kalman Filter" *Cambridge University Press*

■ Harvey A. C., Ruiz E., Shephard N. (1994) "Multivariate Stochastic Variance Models" *Review of Economic Studies*, Vol. 61, No. 2

■ Haykin S. *(Editor)* (2001) "Kalman Filtering and Neural Networks" *Wiley Inter-Science*

■ Heston S. (1993) "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options" *Review of Financial Studies*, Vol. 6, No. 2

■ Ito K., Xiong K. (2000) "Gaussian Filters for Nonlinear Filtering Problems" *IEEE Transactions on Automatic Control*, Vol. 45, No. 5

■ Javaheri A. (2002) "The Volatility Process" *Ph.D. Thesis in progress, Ecole des Mines de Paris*

■ Julier S. J., Uhlmann J.K. (1997) "A New Extension of the Kalman Filter to Nonlinear Systems" *The University of Oxford, The Robotics Research Group*

■ Kushner H. J. (1967) "Approximations to Optimal Nonlinear Filters" *IEEE Transactions on Automatic Control*, Vol. 12

■ Kushner H. J., Budhiraja A. S. (2000) "A Nonlinear Filtering Algorithm based on an Approximation of the Conditional Distribution" *IEEE Transactions on Automatic Control*, Vol. 45, No. 3

■ Lautier D. (2000) "La Structure par Terme des Prix des Commodités : Analyse Théorique et Applications au Marché Pétrolier" *Thèse de Doctorat, Université Paris IX*

■ Lautier D., Galli A. (2000) "Un Modèle de Structure par Terme des Prix des Matières Premières avec Comportement Asymétrique du Rendement d'Opportunité" *Finéco*, Vol. 10

■ Lewis A. (2000) "Option Valuation under Stochastic Volatility" *Finance Press*

■ Madan D., Carr P., Chang E.C. (1998) "The Variance Gamma Process and Option Pricing" *European Finance Review*, Vol. 2, No. 1

■ Merton R. C. (1976) "Option Pricing when the Underlying Stock Returns are Discontinuous" *Journal of Financial Economics*, No. 3

■ Pennacchi G. G. (1991) "Identifying the Dynamics of Real Interest Rates and Inflation : Evidence Using Survey Data" *The Review of Financial Studies*, Vol. 4, No. 1

■ Press W. H., Teukolsky S.A., Vetterling W.T., Flannery B. P. (1997) "Numerical Recipes in C: The Art of Scientific Computing" *Cambridge University Press*, 2nd Edition

■ Schwartz E. S. (1997) "The Stochastic Behavior of Commodity Prices : Implications for Valuation and Hedging" *The Journal of Finance*, Vol. 52, No. 3

■ Van der Merwe R., Doucet A., de Freitas N., Wan E. (2000) "The Unscented Particle Filter" *Oregon Graduate Institute, Cambridge University and UC Berkeley*

■ Welch G., Bishop G. (2002) "An Introduction to the Kalman Filter" *Department of Computer Science, University of North Carolina at Chapel Hill*

■ Wells C. (1996) "The Kalman Filter in Finance" *Advanced Studies in Theoretical and Applied Econometrics, Kluwer Academic Publishers*, Vol. 32

ACKNOWLEDGEMENT