

Linear and nonlinear filtering in mathematical finance: a review

P. DATE AND K. PONOMAREVA

*Department of Mathematical Sciences, Brunel University,
Uxbridge, UB8 3PH, UK.*

This paper presents a review of time series filtering and its applications in mathematical finance. A summary of results of recent empirical studies with market data are presented for yield curve modelling and stochastic volatility modelling. The paper also outlines different approaches to filtering of nonlinear time series.

Keywords: Kalman filtering, volatility models, time series calibration

1. Introduction

The problem of estimating unobserved latent variables from observed market data arises frequently in mathematical finance. Kalman filter, first proposed in Kalman (1960), and its generalizations have been the main tools for estimating the unobserved variables from the observed ones in econometrics and in engineering for several decades and their use is now becoming common in finance. Kalman filter is a conditional moment estimator for linear Gaussian systems. It is used in calibration of time series models, forecasting of variables and also in data smoothing applications.

The purpose of this paper is to provide an introductory and accessible exposition of applications of filtering in finance to operational researchers. It gives a brief overview of Kalman filtering theory and presents recent empirical results on two applications in finance. Some recent developments in approximate nonlinear filtering are also reviewed. The rest of the paper is organized as follows. In section 2, basic linear Gaussian filtering methodology is described, along with its application to maximum likelihood-based calibration to time series models. Sections 3.1 and 3.2 present two case studies for applications of Kalman filtering in mathematical finance. Section 4 outlines some recent approaches for approximate filtering in nonlinear time series and also presents a brief overview of an empirical application on calibration and forecasting using a nonlinear interest rate model. Finally, section 5 summarises the contributions discussed in the paper and outlines promising directions for future research.

2. Basic linear filtering theory

2.1 The filtering problem

Consider a discrete stochastic system described by state space equations

$$d\mathbf{x}_t = \mathbf{a}(\mathbf{x}_t, t)dt + Q(\mathbf{x}_t, t)d\mathbf{w}_t, \quad (2.1)$$

$$\mathbf{y}_t = \mathbf{c}(\mathbf{x}_t, t) + R(\mathbf{x}_t, t)\mathbf{z}_t, \quad (2.2)$$

where \mathbf{y}_t is a variable observed in noise, \mathbf{x}_t is an unobservable or latent variable of interest, \mathbf{a}, \mathbf{c} are deterministic vector-valued functions, Q, R are deterministic matrix valued functions and $\mathbf{w}_t, \mathbf{z}_t$ are standard (vector-valued) Wiener processes. Boldface lowercase letters are used to denote vectors and

\mathbf{m}^\top denotes transpose of a vector \mathbf{m} . Uppercase letters are used to denote matrix valued functions or constant matrices. \mathbf{a}, Q are assumed to be sufficiently regular to allow a unique t-continuous solution to the stochastic differential equation (2.1) (see, e.g., Øksendal (2003), chapter 5). The interval between successive sampling times $\Delta = t_{i+1} - t_i$ is assumed to be constant. The *filtering problem* is the problem of finding the conditional distribution of $\mathbb{E}(\mathbf{x}_{t_i+\Delta} | \mathbf{x}_{t_i}, \mathbf{y}_{t_i})$ for $\Delta > 0$, or finding the samples of conditional moments of this random variable corresponding to the observed values of $\mathbf{y}(t_i)$ and estimated \mathbf{x}_{t_i} . In the rest of the paper, the value of a variable \mathbf{z} at time $t_{i+k\Delta}$ is denoted as \mathbf{z}_{i+k} , $k \geq 0$. The equation (2.1) is often referred to as the *transition equation* and (2.2) is referred to as the *measurement equation*. This generic set-up covers a large variety of applications in signal processing, econometrics, weather sciences and mathematical finance. For simplicity of exposition, we set aside the question of appropriate discretization of (2.1) for the moment and we will start with a very special case of a discrete, linear state space model in the next subsection. We will return to (2.1)-(2.2) when we discuss nonlinear filtering in section 4.

2.2 The linear Kalman filter

Consider a discrete time, linear state space system:

$$\mathbf{x}_{n+1} = A\mathbf{x}_n + \mathbf{b} + Q\mathbf{w}_{n+1}, \quad (2.3a)$$

$$\mathbf{y}_n = C\mathbf{x}_n + \mathbf{d} + R\mathbf{z}_n, \quad (2.3b)$$

where, $\mathbf{w}_n, \mathbf{z}_n$ are zero mean, unit variance, Gaussian and uncorrelated random variables at each time t_n . $A, \mathbf{b}, C, \mathbf{d}, Q > 0$ and $R > 0$ are constants or are known functions of time. Only the real valued variable \mathbf{y}_n is measured or is observable; the variable \mathbf{x}_n is of interest and needs to be estimated. One may arrive at (2.3a) by Euler or Milnstein discretisation of (2.1), with appropriate choices of functions \mathbf{a} and Q . We denote the estimate of \mathbf{x}_n based on information up to time t_{n-i} as $\hat{\mathbf{x}}_{n|n-i}$ for $i \geq 0$ and we assume that the initial estimate $\hat{\mathbf{x}}_{0|0}$ is known. The conditional variance of the estimate is correspondingly denoted by $P_{n|n-i}$ and $P_{0|0} > 0$ is assumed to be known. With this notation, the following set of recursive equations is conventionally referred to as the *Kalman filter*:

$$\mathbf{v}_n = \mathbf{y}_n - (C\hat{\mathbf{x}}_{n|n-1} + \mathbf{d}), \quad (2.4a)$$

$$\Sigma_n = CP_{n|n-1}C^T + RR^\top, \quad (2.4b)$$

$$K_n = P_{n|n-1}C^T\Sigma_n^{-1}, \quad (2.4c)$$

$$\hat{\mathbf{x}}_{n|n} = \hat{\mathbf{x}}_{n|n-1} + K_n\mathbf{v}_n, \quad (2.4d)$$

$$\hat{\mathbf{x}}_{n+1|n} = A\hat{\mathbf{x}}_{n|n} + \mathbf{b}, \quad (2.4e)$$

$$P_{n+1|n} = AP_{n|n-1}A^T + QQ^\top - AP_{n|n-1}C^T\Sigma_n^{-1}CP_{n|n-1}A^T. \quad (2.4f)$$

Here, \mathbf{v}_n in (2.4a) represent information which could not have been derived from data up to time t_{n-1} and are called *innovations*. Σ_n represents the covariance matrix of innovations. $\hat{\mathbf{x}}_{n+1|n}$ is in fact the sample of conditional expectation of \mathbf{x}_{n+1} based on information up to time t_n , determined by the realized value of

y_n . This set of equations can be derived from the following relationship between conditional moments of jointly Gaussian variables (see, e.g. Grimmer & Storzaker (2004)): if x, y are jointly Gaussian,

$$\mathbb{E}(\mathbf{x}|\mathbf{y}) = \mathbb{E}(\mathbf{x}) + \Sigma_{xy}\Sigma_{yy}^{-1}(\mathbf{y} - \mathbb{E}(\mathbf{y})), \quad (2.5)$$

$$\mathbb{E}(\mathbf{x} - \mathbb{E}(\mathbf{x}|\mathbf{y}))(\mathbf{x} - \mathbb{E}(\mathbf{x}|\mathbf{y}))^\top = \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}, \quad (2.6)$$

where Σ_{xy} etc. are variance terms. This has to be interpreted with care since $\hat{\mathbf{x}}_{n|n-1}$ is not *unconditional* mean; the reader is referred to specialist textbooks such as Durbin & Koopman (2001) for more details. Given $\mathbf{y}_n, \hat{\mathbf{x}}_{n+1|n}$ and $(C\hat{\mathbf{x}}_{n+1|n} + \mathbf{d})$ serve as one-step ahead forecasts of \mathbf{x}_{n+1} and \mathbf{y}_{n+1} respectively.

2.3 Time series calibration using maximum likelihood

Given (possibly vector-valued) measurements $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N$, one may use Kalman filter to calibrate a time series model with latent variables such as the one described by (2.3a)-(2.3b) as follows. Let \mathcal{F}_n denote all the measurements available until and including time t_n . The probability density $p(\mathbf{y}_{n+1}|\mathcal{F}_n)$ is Gaussian with

$$\begin{aligned} \mathbb{E}(\mathbf{y}_{n+1}|\mathcal{F}_n) &= C\hat{\mathbf{x}}_{n+1|n} + \mathbf{d} \quad \text{and} \\ \text{Var}(\mathbf{y}_{n+1}|\mathcal{F}_n) &= \Sigma_{n+1}. \end{aligned}$$

We can write the likelihood function (i.e., the joint probability function) for the set of observations $Y = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N\}$ as

$$L(Y) = p(\mathbf{y}_1) \prod_{i=2}^N p(\mathbf{y}_i|\mathcal{F}_{i-1}).$$

It is usually simpler to work with the logarithm of likelihood, which is given by

$$\log L(Y) = \sum_{i=1}^N \log p(\mathbf{y}_i|\mathcal{F}_{i-1}) = -\frac{1}{2} \sum_{i=1}^N (\log |\Sigma_i| + \mathbf{v}_i^\top \Sigma_i^{-1} \mathbf{v}_i),$$

when the constant terms are ignored. Given time series data $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N$, the quantities \mathbf{v}_i and Σ_i are found through Kalman filter recursions outlined in the previous section. The above function can then be maximized to find the parameter vectors \mathbf{b}, \mathbf{d} and matrices A, C, Q and R using an off-the-shelf nonlinear solver such as `fminsearch` in MATLAB. The initial state $\hat{\mathbf{x}}_{0|0}$ and its covariance $P_{0|0}$ may be independently parameterized or it can be expressed in terms of other parameters (as in section 3.2 below). The readers are referred to Harvey (1989) and Durbin & Koopman (2001) for more details on maximum likelihood-based calibration, including parameter initialisation issues in Kalman filtering framework.

3. Applications of linear filtering in finance

In mathematical finance, filters are useful in many applications, including estimation of underlying volatility from observed intra-day stock prices as discussed in Barndorf-Nielsen & Shephard (2002), estimation of underlying instantaneous interest rate (or short rate) from observed bond yields as explored in several papers including Babbs & Nowman (1999) and estimation of spot prices from futures prices for commodities, as reported in Schwartz (1997), Manoliu & Tompaidis (2002) and Lautier & Galli (2004). Besides the estimation of latent factors which affect pricing described in these papers, linear

filtering is used in Monoyios (2007) for updating the estimates of uncertain drift parameters in the price process, in the context of hedging in incomplete markets.

We look at two specific application case studies of linear filtering as used in mathematical finance next.

3.1 Linear Gaussian interest rate models

Exponential affine term structure models is one of the oldest and the most widely studied class of dynamic interest rate models. The main advantage of these models is the fact that the yields can be expressed as affine functions of the short rate (i.e. the instantaneously compounded interest rate). One factor linear Gaussian model was introduced in Vasicek (1977). Multi-factor linear Gaussian models under filtering framework are discussed in Jegadeesh & Pennachhi (1996), Babbs & Nowman (1999) and De Rossi (2004), among others. Extensive treatment to theoretical properties and applications of these models can be found in advanced graduate textbooks such as James & Webber (2000) and Brigo & Mercurio (2006).

In the work presented here, we consider a n -factor, linear Gaussian short rate model of the form

$$i_t = \mu - \sum_{j=1}^n x_{j,t}, \quad (3.7a)$$

$$dx_{j,t} = -\alpha_j x_{j,t} + \sum_{l=1}^n \sigma_{jl} dw_{l,t}, \quad (3.7b)$$

where i_t is the (unobservable) short rate, α_j , σ_{jl} , μ etc are constants and $w_{l,t}$ are independent standard Wiener processes. The price of a zero coupon bond at time t with time to maturity τ_k is given by

$$P(t, \tau_k) = \mathbb{E} \left[\exp \left(- \int_t^{t+\tau_k} i_s ds \right) \right],$$

where the expectation is assumed to be taken under appropriate risk neutral measure. The equivalent continuously compounded interest rate for the same time to maturity τ_k is called spot rate and is denoted by $i_t(\tau_k|\theta)$. For the chosen dynamics, theoretical spot rates are affine functions of the instantaneous interest rate r_t :

$$i_t(\tau_k|\theta) = A_0(\tau_k|\theta) + \sum_{j=1}^n A_j(\tau_k|\theta) x_{j,t}, \quad (3.8)$$

where $A_j(\cdot|\cdot)$ are known functions and θ is the vector of parameters of (3.7); see Babbs & Nowman (1999) or Date & Wang (2009) for the exact form of A_j .

From a commercial data provider such as Datastream in the UK, one may obtain the actual spot rates for different times to maturities, each day. We assume that our model is imperfect and the actual spot rates equal the corresponding theoretical spot rates plus a zero mean noise term:

$$y_{t,\tau_k} = i_t(\tau_k|\theta) + z_{t,k}, \quad (3.9)$$

with $\mathbb{E}(z_{t,k}) = 0$, $\mathbb{E}(z_{t,k}^2) = h_k^2$ and h_k are constants. A discretised version of (3.7a)-(3.7b) gives us a transition equation similar to (2.1) while the equations (3.8)-(3.9) give a measurement equation similar to (2.2). This forms a linear state space system to which the methodology explained in section 2 can be applied. In particular, given observed y_{t,τ_k} for various maturities at each t , we can use Kalman filtering to calibrate this model and forecast spot rates.

We provide a selective summary of numerical experiments reported in Date & Wang (2009). Two different data sets were used in this paper; weekly data on 7 different UK government bond (i.e. gilt) yields from January 2001 to June 2005 formed the first data set and weekly data on 7 different US government bond (i.e. treasury) yields during a more volatile period from December 1997 to August 2001 was the second data set. Out of 232 observations for UK data, 180 were used for calibration and 42 were used for validation. Through principal component analysis, it was found that over 97.8% variation in the interest rates is explained by two factors and over 99.4% variation was explained by three factors. This corroborates similar conclusions in Babbs & Nowman (1999). Accordingly, one, two and three factor linear models were calibrated using Kalman filter and maximum likelihood method described earlier. The parameter values after calibration of a 15 parameter, two factor model are reported in table 1.

Table 1. Parameter values for two factor model (UK data)

α_1	0.7036	α_2	0.2807
σ_{11}	0.0080	σ_{22}	0.1806
λ_1	0.4249	λ_2	-0.0414
μ	0.2768	h_1	0.0006
h_2	0.0007	h_3	0.0014
h_4	0.0017	h_5	0.0011
h_6	0.0007	h_7	0.0010
σ_{12}	-0.0003		

Here λ_i are prices of risk which relate the drifts implied by the time series data and those implied by the cross sectional spot rate data at a particular point in time. These were assumed to be constant over time, similar to the earlier empirical work in Babbs & Nowman (1999).

The intended use of the model after calibration was short and medium term forecasting. Hence we consider the sample mean of the relative absolute error (MRAE) as our measure of error for each time to maturity τ_k :

$$MRAE = \frac{1}{N} \sum_{j=1}^N \frac{|\text{observed yield} - \text{predicted yield}|}{\text{observed yield}}.$$

This was computed over the relevant set of N observations (either in-sample or out-of sample), for one step ahead prediction of yields. It was found that the maximum out-of-sample error (over 7 yields) is only 2.31%, over a period of almost two years from calibration. The results are summarised in table 2, where the first column indicates the time to maturity for which the MRAE is computed. Similar, if less spectacular, results were obtained for US T-bill data for a more turbulent economic period of 1999-2001. The maximum out-of-sample error for a two factor model calibrated with US data was 3.5%. The findings in Date & Wang (2009) may be summarized as follows.

1. When comparing out-of-sample, one step ahead forecasting with latent variable linear Gaussian models, a two factor model performs better than a single factor or a three factor model.
2. Over a restricted yield spectrum (i.e., only short dated or only the long dated yields are considered), a single factor model leads to a similar magnitude of errors as multi-factor models.
3. The performance of multi-step ahead forecasting of one factor and multi-factor models is comparable, or in other words, additional factors do not make it easier to look further into the future.

The reader is referred to Date & Wang (2009) for full numerical results.

Table 2. Mean relative absolute errors (UK data)

τ_k	in-sample	out-of-sample
3m	0.0111	0.0118
6m	0.0178	0.0081
1y	0.0331	0.0216
2y	0.0337	0.0229
4y	0.0262	0.0152
8y	0.0209	0.0231
10y	0.0161	0.0213

3.2 Stochastic volatility models

The second application reviewed here is that of modelling volatility of stock prices as a stochastic process. For pricing complex financial derivatives, assuming that volatility of asset price is constant is inadequate and, in the last two decades, a lot of effort has been directed towards modelling it as a stochastic process which is both tractable and leads to consistent derivative prices. The landmark work in this area was the closed-form European option pricing result in Heston (1993) for mean reverting stochastic volatility. This was followed by research into multi-factor volatility models and models with jump processes; the reader is referred to Javaheri (2005) and Gatheral (2006) for extensive reviews of this topic. From a somewhat different point of view, the research in this area was also pursued by risk practitioners for forecasting the future volatility and hence the risk of loss.

In the work reported here, we focus on the latter application and use intra-day stock prices to forecast 1-day ahead volatility using a Kalman filter. The return equation for the log-stock price s^l is given by

$$ds^l(t) = \mu dt + \sigma(t)d\tilde{w}(t),$$

where μ is the instantaneous drift, $\tilde{w}(t)$ is the standard Wiener process and $\sigma^2(t)$ is the *spot* volatility or instantaneous volatility. μ is assumed to be constant while $\sigma^2(t)$ is assumed to be driven by a Wiener process uncorrelated with $\tilde{w}(t)$. $\sigma^2(t)$ is obviously unobservable. An integral of spot volatility over a given period Δ is called the *actual* volatility, σ_n^2 :

$$\sigma_n^2 = \int_0^{n\Delta} \sigma(u)^2 du - \int_0^{(n-1)\Delta} \sigma(u)^2 du.$$

σ_n^2 is unobservable as well. The *realised* volatility ζ_n , given by

$$\begin{aligned} \zeta_n &= (1+c) \sum_{j=1}^M \left[s^l \left\{ (n-1)\Delta + \frac{\Delta j}{M} \right\} - s^l \left\{ (n-1)\Delta + \frac{\Delta(j-1)}{M} \right\} \right]^2, \\ &= \sigma_n^2 + u_n, \end{aligned}$$

is a noisy estimate of σ_n^2 , where u_n is a zero mean i.i.d. sequence. This is a consistent estimate of σ_n^2 as $M \rightarrow \infty$ if $c = 0$ and is an unbiased estimate if $c = \mu = 0$. μ is usually small for a small period Δ , e.g. a one day stock price return is dominated by volatility rather than by drift. c is a scaling factor

to account for overnight returns being much larger than the intra-day returns, see, e.g. Martens (2002). Euler discretisation of a constant coefficient, linear, Gaussian and mean reverting stochastic process

$$dx(t) = a(b - x(t))dt + \beta dw(t), \quad (3.10)$$

with $x(t) = \sigma^2(t)$ and $w(t)$ a standard Wiener process uncorrelated to $\tilde{w}(t)$, leads to a linear discrete time state space system similar to the one in (2.3a)-(2.3b):

$$\begin{aligned} \sigma_{n+1}^2 &= \phi \sigma_n^2 + \gamma + q w_{n+1}, \\ \zeta_n &= \sigma_n^2 + r z_n. \end{aligned} \quad (3.11)$$

Here, $\phi = 1 - a\Delta$, $\gamma = ab\Delta$, $q = \beta\sqrt{\Delta}$, a, b, β, r are constants and w_n, z_n are standard normal i.i.d. random variables. Δ is the time step and the subscript n indicates the value of variable at time t_n , as before. Similar models are used in earlier, seminal work on realized volatility in Barndorff-Nielsen & Shephard (2002). A more complex model where a, β above are functions of a finite state Markov chain is described in Frey & Runggaldier (2001), although we focus on the exposition for the linear state space model in this review. From this point onwards, we can use the calibration and forecasting procedures mentioned in the earlier sections, along with the realised volatility ζ_n computed from intra-day asset prices, to calibrate the model and to forecast future volatility. Extensive numerical experiments for calibration and forecasting with stochastic volatility models were reported in Hawkes (2007) and Hawkes & Date (2007). We provide a selective summary of this work here.

For numerical experiments, we used 5 minutes intra-day data for IBM and Citigroup stocks from August 1997 to January 2005 from commercial data provider *PriceData*. Δ in both cases was one day and M was 78, i.e. daily volatility is estimated based on 78 intra-day stock price readings. Both the stocks were very liquid in this period and did not suffer from price discreteness (i.e. from intra-day price changes being few and far between). Days which contained no data were ignored, as is a common practice within econometrics. Any intra-day missing data was dealt with using standard methods for stochastic interpolation. Log-price data was further scaled by a factor of 100 so that the returns are expressed as percentage. We used maximum likelihood along with the output of the Kalman filter for estimation of parameters, as outlined in section 2.3. For initialization, we take steady state values

$$\hat{\sigma}_{0|0}^2 = \lim_{n \rightarrow \infty} \mathbb{E}(\sigma_n^2|0) = \frac{\gamma}{1 - \phi} \text{ and } P_{0|0} = \lim_{n \rightarrow \infty} \text{Var}(\sigma_n^2|0) = \frac{q^2}{1 - \phi^2}.$$

In-built routine `fminsearch` from MATLAB were used for non-convex optimisation in calibration.

Single factor as well as two-factor state space models were calibrated using Kalman filter. The two factor model has γ in (3.11) driven by a Gaussian stochastic process. The discussion of two-factor model is omitted for brevity; see Hawkes & Date (2007) details. The forecasting performance of these models was compared against GARCH type models calibrated from same data. GARCH (Generalized Auto Regressive Conditional Heteroskedasticity) type models (first introduced in Bollerslev (1986)) are used commonly in the finance industry for short term volatility forecasting. As a benchmark for comparison with the latent state model, we use GJR GARCH model as proposed in Glosten *et al* (1993), but with Student-t distributed returns as proposed in Bollerslev (1987). The volatility under this model can be expressed as

$$\sigma_{n+1}^2 = \beta \sigma_n^2 + \rho + \alpha \zeta_n^2 + \lambda \zeta_n^2 \max(0, \zeta_n), \quad (3.12)$$

where ζ_n is the daily log-return which is assumed to have a Student-t distribution with ν degrees of freedom and $\beta, \alpha, \rho, \lambda$ are constants with $\alpha + \beta < 1$. For more details on GARCH and modified GARCH models, the reader is referred to Hansen & Lunde (2005) and references therein.

Table 3 below shows the parameter values of single factor stochastic volatility models for the two stocks.

Table 3. Parameter values

	IBM	Citigroup
ϕ	0.93	0.89
γ	0.31	0.71
q	1.38	3.32
r	3.68	5.7

The comparison of forecasting performance was based on the variance of k -step ahead forecasting errors, i.e. on the sample mean of $(\hat{\sigma}_{n+k|n}^2 - \zeta_{n+k})^2$. For a given $k > 1$, the forecast $\sigma_{n+k|n}^2$ can be constructed in various ways. One ad-hoc, but industry-standard approach is scaling $\hat{\sigma}_{n+1|n}^2$ by k . Table 4 below illustrates the variance of errors obtained by a single factor stochastic volatility (SV) model and the GJR GARCH-t model (3.12) using this approach for forecasting, with $k = 10$. It can be seen that Kalman filter based stochastic volatility model outperforms the modified GARCH model. The experiments reported in Hawkes & Date (2007) indicate that the latent state based volatility models usually outperform GARCH type models and provide a simple alternative to the established practice of using GARCH models in short term forecasting.

Table 4. Variance of 10-step ahead forecasting errors for IBM and Citigroup stock.

	IBM	Citigroup
SV model	724.51	2101.75
GARCH-t	1142.25	3402.81

The reader is referred to Hawkes (2007) for statistical properties of data, calibration of multi-factor models and detailed numerical results on forecasting.

4. Approximate nonlinear filtering

We now return to equations (2.1)-(2.2) and consider the filtering problem for more general nonlinear systems. A typical example of a practically useful nonlinear latent variable model in finance is the square root affine model first introduced in Cox *et al.* (1985) and extended to multi-factor case in Beaglehole & Tenney (1991):

$$i_t = \sum_{j=1}^n x_{j,t}, \quad (4.13a)$$

$$dx_{j,t} = \alpha_j(\mu_j - x_{j,t}) + \sum_{l=1}^n \sigma_{jl} \sqrt{x_{l,t}} dw_{l,t}, \quad (4.13b)$$

$$i_t(\tau_k | \theta) = A_0(\tau_k | \theta) + \sum_{j=1}^n A_j(\tau_k | \theta) x_{j,t}, \quad (4.13c)$$

$$y_{t,\tau_k} = i_t(\tau_k | \theta) + z_{t,k}, \quad k = 1, 2, \dots, s. \quad (4.13d)$$

Here, θ denotes the vector of parameters of the transition equation (4.13b). Detailed expressions for the functions $A_0(\tau_k|\theta)$ and $A_j(\tau_k|\theta)$ will be given in section 4.5, where we discuss an empirical filtering application with the above model. A similar, but more extensive empirical study was reported earlier in Geyer & Pichler (1999). Typically, the measurements y_{t,τ_k} represent s different bond yields observable in the market at time t , corresponding to times to maturity $\tau_k, k = 1, 2, \dots, s$ while i_t represents the unobservable short rate at time t . The other parameters have the same meaning as in (3.7)-(3.9) in the previous section, although the formulae for $A_j(\tau_k|\theta)$ are different from those in the linear Gaussian case.

In cases similar to (4.13) where \mathbf{a}, Q in (2.1)-(2.2) are not affine, closed-form formulae are not available for computing the conditional moments $\mathbb{E}(\mathbf{x}_n|\mathbf{y}_n)$ and $\text{Var}(\mathbf{x}_n|\mathbf{y}_n)$. In continuous time, the evolution of conditional distribution is described by a stochastic partial differential equation (SPDE); interested readers are referred to Kallianpur (1990) and Kushner (2000) for the exact solution to the continuous time filtering problem. In Brigo *et al.* (1998), a finite dimensional approximation to exact nonlinear filter is proposed using a differential geometric approach to statistics, which leads to an ordinary stochastic differential equation instead of the SPDE mentioned above. This approach is rigorous, but is still computationally very demanding; especially for multi-variable models. A similar geometric approach was followed earlier in Kulhavý (1990) for stochastic difference equations (i.e. for variables evolving in discrete time). In this review, we focus our attention to commonly employed approximate solutions to nonlinear filtering problems. We will consider scalar systems in what follows, although most of the results can easily be generalized to deal with multi-variable time series.

4.1 Extended Kalman filter

The first and obvious approach to filter x_n is to use *extended* Kalman filters. For scalar versions of (2.1)-(2.2), one can expand the dynamics in Taylor series about $\hat{x}_{n|n-1}$ as

$$a(x_n, t_n) \approx a(\hat{x}_{n|n-1}, t_n) + \left(\frac{\partial a}{\partial x}\right) (x_n - \hat{x}_{n|n-1}) + \left(\frac{\partial a}{\partial t}\right) \Delta, \quad (4.14a)$$

$$c(x_n, t_n) \approx c(\hat{x}_{n|n-1}, t_n) + \left(\frac{\partial c}{\partial x}\right) (x_n - \hat{x}_{n|n-1}) + \left(\frac{\partial c}{\partial t}\right) \Delta, \quad (4.14b)$$

$$Q(x_n, t_n) \approx Q(\hat{x}_{n|n-1}, t_n), \quad (4.14c)$$

$$R(x_n, t_n) \approx R(\hat{x}_{n|n-1}, t_n), \quad (4.14d)$$

where the partial derivatives are evaluated at $(\hat{x}_{n|n-1}, t_{n-1})$ and $\Delta = t_n - t_{n-1}$, as before. (4.14) gives a linear approximation to the original nonlinear state space system, which, in turn allows us to use the techniques from sections 2.2 and 2.3 for model calibration and forecasting. This method works reasonably well in systems with smooth nonlinearities. This filter is referred to as a *local linearization filter* in Jimenez & Ozaki (2003) and a similar formulation has been used in parameter estimation for forward rate models in Chiarella *et al.* (2009).

4.2 Particle filters

In most filtering applications, the aim is to compute either

$$\mu = \int f(x_n) p(x_n|y_n) dx$$

or
$$\mu = \int f(x_n) p(x_n|x_{n-1}, y_{n-1}) dx,$$

where $f(x)$ is such that μ represents conditional mean or conditional variance of x_n . Apart from the linear Gaussian case, a closed-form expression for μ is not available in general. In fact, it may not be easy to sample from the conditional density $p(x_n|y_n)$. One may instead choose to sample from another density, say g , which is close in appropriate sense to the posterior density $p(x_n|y_n)$ and is easier to draw samples from. g is then called the *importance density*. Given samples $x_{n-1}^{(i)}$ drawn from $g(x_{n-1}|x_{n-2}, y_{n-1})$, choosing the corresponding probability weights $w_{n-1}^{(i)}$, $i = 1, 2, \dots, M$ such that

$$w_{n-1}^{(i)} \propto \frac{p(x_{n-1}^{(i)}|y_{n-1})}{g(x_{n-1}^{(i)}|y_{n-1})}$$

ensures that

$$\lim_{M \rightarrow \infty} \sum_{i=1}^M w_{n-1}^{(i)} h(x_{n-1}^{(i)}) = \mathbb{E}_p(h(x_{n-1}))$$

holds for any measurable function h for which $\mathbb{E}_p(h(x))$ exists, where $\mathbb{E}_p(\cdot)$ is expectation with respect to probability measure $p(x_{n-1}|y_{n-1})$. \propto indicates equality up to a constant of proportionality. The idea of *particle filter* (also called sequential Monte Carlo filter or sequential importance sampling filter) is to generate a discrete distribution with support points $x_n^{(i)}$ and probability weights $w_n^{(i)}$ which approximate the posterior density $p(x_n|y_n)$ and update these recursively as new measurements become available. To derive a recursive update formula, assume that the factorization

$$g(x_n|y_n) = g(x_n|x_{n-1}, y_n)g(x_{n-1}|y_{n-1}) \quad (4.15)$$

holds for the importance density. Let $p_o(y_n|x_n)$ and $p_t(x_n|x_{n-1})$ denote the observation density and the state transition density respectively and note that

$$p(x_n|y_n) \propto p_o(y_n|x_n)p_t(x_n|x_{n-1})p(x_{n-1}|y_{n-1}) \quad (4.16)$$

holds due to Bayes rule, assuming that the state x_n evolves according to a Markov process. Given $x_{n-1}^{(i)}$ and $p_{n-1}^{(i)}$, $i = 1, 2, \dots, M$, $x_n^{(i)}$ and $w_n^{(i)}$ are generated as follows.

1. Draw samples $x_n^{(i)}$ from the importance density $g(x_n|x_{n-1}, y_n)$.
2. Choose new weights as

$$w_n^{(i)} \propto \frac{p(x_n^{(i)}|y_n)}{g(x_n^{(i)}|y_n)}, \quad (4.17)$$

which, after substituting from (4.15)-(4.16) yields the necessary recursive update formula:

$$w_n^{(i)} \propto w_{n-1}^{(i)} \frac{p_o(y_n|x_n^{(i)})p_t(x_n^{(i)}|x_{n-1}^{(i)})}{g(x_n^{(i)}|x_{n-1}^{(i)}, y_n)} \quad (4.18)$$

The choice of density $g(x_n|x_{n-1}, y_n)$ is a crucial step in the design of particle filter. A common choice is the Gaussian posterior density obtained by using the extended Kalman filter for the same system.

Various heuristics exist to improve the performance of the filter in relation to the variance of state estimate and numerical efficiency.

Methods of this type have been used in a variety of areas including speech recognition, image processing, target tracking and financial modelling; see Doucet *et al.* (2001) for a review of applications while Arulampalam *et al.* (2002) and Daum (2005) provide comprehensive tutorials on various types of particle filters. One expects that the posterior density of the particle filter converges to the optimal conditional density as the number of particles becomes large. Crisan & Doucet (2002) bring together different asymptotic convergence results related to particle filters. Applications of particle filters in financial time series have been reported in Pitt & Shepherd (1999), Javaheri *et al.* (2003) and Fearnhead (2005), among others.

These filters are difficult to calibrate due to the computation involved in computing the likelihood, since a closed-form expression is rarely available. Some computationally efficient alternatives to particle filters are discussed below.

4.3 *Sigma point filters*

An increasing popular alternative to particle filters for signal processing in real time is using *unscented* or *sigma point* filters. Methods of this type have been developed independently in engineering (see, *e.g.* Julier & Uhlmann (2004) and references therein for examples) and in weather sciences, where they are referred to as *ensemble Kalman filters* (see, *e.g.* Evensen (1994) and Mitchell & Hotekamer (1998) for examples). In contrast with the extended Kalman filters described earlier, sigma point filters do not involve computation of derivatives. The basic methodology behind these filtering methods can be explained as follows. As in the previous two subsections, scalar systems are considered here for notational simplicity.

1. Given $P_{n|n-1}$, $\hat{x}_{n|n-1}$, generate *sigma points* $x_n^{(i)}$ and probability weights $p_n^{(i)}$ via a deterministic algorithm or random sampling, such that $\sum_{i=1}^M p_n^{(i)} x_n^{(i)} = \hat{x}_{n|n-1}$ and $\sum_{i=1}^M p_n^{(i)} (x_n^{(i)} - \hat{x}_{n|n-1})^2 = P_{n|n-1}$.
2. Then use

$$\hat{x}_{n+1|n} = \sum_{i=1}^M p_n^{(i)} g(x_n^{(i)}) + \Sigma_{xy} \Sigma_{yy}^{-1} \left(y_n - \sum_{i=1}^M p_n^{(i)} h(x_n^{(i)}) \right),$$

with covariance matrices Σ_{xy} , Σ_{yy} defined using the probability weights $p_n^{(i)}$. $P_{n+1|n}$ is similarly computed using (2.6).

It can be seen that the methods are based on constructing covariance matrices from the samples of distribution with correct first two moments, and then using the closed-form formulae (2.5)-(2.6) for the state and the covariance update. The number of samples used tends to be significantly smaller than in particle filters. Ensemble filters use sampling from a Gaussian distribution and match sample mean and sample covariance, while sigma point filters use a deterministic algorithm to generate a discrete distribution (support points as well as possibly unequal probability weights) matching the specified two moments. A nonlinear filtering heuristic which combines the desirable features of both the ensemble filters and sigma point filters has been reported in Date *et al.* (2008). This heuristic is based on a sampling algorithm reported in a companion paper by the same authors (Date *et al.* (2008)). This sampling algorithm leads to exact matching of specified moments while still randomly generating support points

with unequal probability weights. The idea of this sampling method can be briefly explained as follows. Let $p_n^{(i)}, i = 1, 2, \dots, m$ be any set of points such that $\sum_{i=1}^m p_i < \frac{1}{2}$ and let

$$p_n^{(m+1)} = 1 - 2 \sum_{i=1}^m p_n^{(i)}.$$

Let μ and σ^2 be the target mean and target variance respectively. Define a discrete random variable x_n on a support of $M = 2m + 1$ points by

$$\mathbb{P}\left(x_n = \mu \pm \frac{\sigma}{\sqrt{2mp_n^{(i)}}}\right) = p_n^{(i)}, i = 1, 2, \dots, m,$$

$$\mathbb{P}(x_n = \mu) = p_n^{(m+1)}.$$

Then it can easily be shown that

$$\mathbb{E}(x_n) = \mu, \mathbb{E}(x_n - \mu)^2 = \sigma^2$$

holds irrespective of exact values of $p_n^{(i)}$. $p_n^{(i)}$ can either be generated by sampling a distribution on non-negative support or they can be used for matching higher order moments. This can be generalized to multi-variable case, as shown in Date *et al.* (2008).

Finally, sigma point filters are also frequently used to generate posterior importance densities in particle filters. Their potential for use in nonlinear multi-factor models in finance is yet not fully explored.

4.4 *Optimisation based filters*

Kalman filter and related methodologies are based on the premise that the optimal estimate is a conditional mean of the unobserved variable, given the measurements. Different heuristic methods for nonlinear filtering represent different ways of approximating this conditional mean. An entirely different approach is followed in Cortazar & Schwartz (2003). To put the approach in Cortazar & Schwartz (2003) in a slightly general setting, consider a set of discretised equations

$$\begin{aligned} x_{n+1} &= a(x_n, t_n) + q(x_n, t_n)w_{n+1}, \\ y_n &= c(x, t_n) + r(x_n, t_n)z_n. \end{aligned} \quad (4.19)$$

Given y_n , x_n is found as a solution to the optimization problem of the following form:

$$\min_x J(y_n - c(x, t_n)), \quad (4.20)$$

where $J(\cdot)$ is a suitable non-negative cost function which is zero only at the origin. If \hat{x}_n is the argument minimising J in (4.20), the prediction \hat{x}_{n+1} is found using

$$\hat{x}_{n+1} = a(\hat{x}_n, t_n).$$

This does away entirely with the need for knowing the statistical properties of the ‘noise’ terms $q(x_n, t_n)w_n$ and $r(x_n, t_n)z_n$ and can work well whenever the noise variances are not too large relative to the magnitude of the hidden states. A somewhat similar approach is followed in the *partially linearised sigma*

point filter proposed in Date *et al.* (2009), where a set of linear programming problems are solved at each t_n to generate sigma points for a linearized measurement equation.

However, analysis of accuracy and convergence properties of optimisation-based filters is more difficult and needs to be explored further. The unknown-but-bounded noise framework used in system identification may be useful for this purpose; see, e.g. Bravo *et al.* (2006) and references therein.

4.5 Empirical application of nonlinear filters: CIR model

Here, we outline a small empirical study of nonlinear filtering algorithms applied to exponential affine term structure model with a nonlinear transition equation. The system under consideration is a discretisation of the multi-factor CIR model with the state evolution given by (4.13b):

$$x_{j,n+1} = (1 - \exp(-\alpha_j \Delta)) \mu_j + \exp(-\alpha_j \Delta) x_{j,n} + Q_{j,n} w_{j,n}, \quad (4.21)$$

for $j = 1, 2$, where $w_{j,n}$ are zero mean, unit variance and uncorrelated Gaussian random variables. The standard deviation Q depends on the latent state and is given by

$$Q_{j,n} = \sigma_j \sqrt{\frac{(1 - \exp(-\alpha_j \Delta))}{\alpha_j} (0.5 \mu_j (1 - \exp(-\alpha_j \Delta)) + \exp(-\alpha_j \Delta) x_{j,n-1})}, \quad (4.22)$$

It can be shown that this discretisation preserves the first two conditional moments, *i.e.*,

$$\begin{aligned} \mathbb{E}(x_{j,t_n+\Delta} | x_{j,t_n}) &= (1 - \exp(-\alpha_j \Delta)) \mu_j + \exp(-\alpha_j \Delta) x_{j,t_n} \text{ and} \\ \text{Var}(x_{j,t_n+\Delta} | x_{j,t_n}) &= Q_{j,n}^2, \end{aligned}$$

see, *e.g.* Brigo & Mercurio (2006) for details. The observable variables are exponential in the latent states and are given by

$$Y_{i,n} = \prod_{j=1}^2 \left(\tilde{A}_{i,j} \exp\left(-\sum_{j=1}^2 (B_{i,j} x_{j,n})\right) \right) + z_{i,n}, \quad (4.23)$$

where

$$\tilde{A}_{i,j} = \left(\frac{2\gamma_j \exp((\alpha_j + \gamma_j + \lambda_j) T_i / 2)}{2\gamma_j + (\alpha_j + \lambda_j + \gamma_j)(\exp(T_i \gamma_j) - 1)} \right)^{\frac{2\alpha_j \mu_j}{\sigma_j^2}}, \quad (4.24)$$

$$B_{i,j} = \frac{2(\exp(T_i \gamma_j) - 1)}{2\gamma_j + (\alpha_j + \lambda_j + \gamma_j)(\exp(T_i \gamma_j) - 1)}, \quad (4.25)$$

$$\gamma_j = \sqrt{(\alpha_j + \lambda_j)^2 + 2\sigma_j^2}, \quad (4.26)$$

$z_{i,n}$ is observational noise with zero mean and a constant variance h^2 for each i . The terms $\tilde{A}_{i,j}, B_{i,j}$ in (4.24)-(4.25) are related to $A_0(\tau_i | \theta), A_j(\tau_i | \theta)$ in (4.13c) as

$$\begin{aligned} A_0 &= -\sum_{j=1}^n \frac{\log(\tilde{A}_{i,j})}{T_i}, \\ A_j &= \frac{B_{i,j}}{T_i}, \end{aligned}$$

and $n = 2$. In practice, T_i represents time to maturity and $Y_{i,n}$ represents the price of a zero coupon bond with maturity $T_i + t_n$, at time t_n . Here we use three maturities, $T_1 = 1$, $T_2 = 2$ and $T_3 = 4$. This gives a two-state, three measurement state space system. Note that the logarithm of $Y_{i,n}$ is often available as a measurement and is affine in the state variables; this logarithm is nothing but the *spot rate* employed earlier in section 3.1. In this section, we choose to use nonlinear measurement equation with $Y_{i,n}$ as measurements to illustrate performance of sigma point filter in comparison to extended Kalman filter.

For numerical experiments we use weekly data from January 2001 to June 2005 for 3 different UK government bond yields. As in section 3.1, 180 observations were used for calibration and 42 were used for validation. A 2-factor model was calibrated using the extended Kalman filter described in section 4.1 and the maximum likelihood method described in section 2.3. As before, in-built optimization routines from MATLAB were used for calibration. Table 5 reports the parameter values obtained as a result of calibration.

Table 5. Parameter values

μ_1	0.0254
μ_2	0.0175
σ_1	0.0710
σ_2	0.1870
α_1	0.0978
α_2	0.8035
λ_1	-0.0350
λ_2	-0.0490
h	0.001

Here λ_1 and λ_2 have the same interpretation as the corresponding parameters in table 1. After calibration, we use the sigma point generation method described in Julier & Uhlmann (2004) to generate sigma points at each t_n , with initial values for mean $x_{j,0} = \mu_j$ and diagonal elements of covariance as $\frac{\mu_j \sigma_j^2}{2\alpha_j}$. Five sigma points are generated at each t_n . Bearing in mind the nonnegativity restriction on state variables $x_{j,n} \geq 0$ we replace any negative element of state estimate $x_{n|n-1}$ with zero. These points are then used to construct $\hat{x}_{j,n+1|n}$, $j = 1, 2$ and the corresponding predictions of $Y_{i,n+1}$, $i = 1, 2, 3$. As a benchmark for comparison, we use the predictions made using the extended Kalman filter.

To compare the performance of sigma point filters and extended Kalman filters, we consider the sample mean of the relative absolute error (MRAE) for each time to maturity, in the same way as in section 3.1. MRAE is computed over in-sample and out-of-sample data separately. Table 6 lists the errors computed for one step ahead prediction of yields, for both the extended Kalman filter (EKF) and the sigma point filter (SPF). It can be seen that the sigma point filter outperforms the extended Kalman filter, both in-sample and out-of-sample and for all yields. This complements a huge body of empirical experience within engineering about the performance improvement obtainable with sigma point filters. Given the ease of implementation (relative to particle filters) and improved performance (relative to the extended Kalman filters), sigma point filter provides an attractive approximate solution for nonlinear time series filtering within econometrics and finance.

Table 6. Relative absolute errors of 1-step ahead prediction for 2-factor model

τ_k	EKF in-sample	EKF out-of-sample	SPF in-sample	SPF out-of-sample
1Y	0.00237	0.00411	0.00160	0.00144
2Y	0.00388	0.00683	0.00283	0.00265
4Y	0.00606	0.01132	0.00483	0.00412

5. Summary

A brief and introductory review of linear and nonlinear filtering methodology has been presented in this paper, including results of three numerical studies on filtering of financial time series data. Filtering in nonlinear financial time series is a very active area of research worldwide. Several problems of theoretical and practical interest remain to be explored. Two specific challenges for operations researchers are worth mentioning.

1. In multi-stage stochastic optimisation of fixed income or commodity portfolios, the use of latent state based models can provide arbitrage-free evolution of prices while providing huge savings in terms of number of nodes reached and number of decision variables (*e.g.*, price evolution for a typical 20 or 30 government bond portfolio can be modelled with a reasonable accuracy by a 2 factor linear Gaussian model and can hence be described by a two dimensional re-combining lattice). Empirical studies for stochastic portfolio optimisation using latent state based models will be valuable to both industrial and academic researchers in the OR community.
2. While moment-matching based methods have been used in a static stochastic optimisation set-up in Høyland *et al* (2003) and in subsequent work, the potential for dynamic moment matching for latent factors in a multi-stage optimisation, using the same scenario (or sigma point) generation methodology as in sigma point filters, has not yet been exploited.

It is hoped that this review will stimulate more interest in research on theory and applications of filtering in mathematical finance.

Acknowledgements

The authors are grateful to the editor and the anonymous referee for their constructive comments which helped to improve the quality of this paper.

References

- ARULAMPALAM, M.S., MASKELL, S., GORDON, N. & CLAPP, T. (2002) A tutorial on particle filters for online nonlinear /non-Gaussian Bayesian tracking, *IEEE T. Signal Proces.*, **50**, 174–188.
- BABBS, S.H. & NOWMAN, K.B. (1999) Kalman filtering of generalized Vasicek term structure models, *J. Financ. Quant. Anal.*, **34**, 115–130.
- BARNDORFF-NIELSEN, O.E. & SHEPHARD, N. (2002) Econometric analysis of realized volatility and its use in estimating stochastic volatility models, *J. Roy. Stat. Soc. B. Stat. Meth.* **64**, 253–280.
- BEAGLEHOLE, D.R. & TENNEY M.S. (1991) General solutions of some interest rate-contingent claim pricing equations, *J. Fixed Income*, **1**, 69–83.
- BRAVO, J.M., ALAMO, T. & CAMANCHO, E.F. (2006) Bounded error identification of systems with time varying parameters, *IEEE T. Automat. Contr.*, **51**, 1144–1150.

- BRIGO, D., HANZON, B. & LEGLAND, F. (1998) A Differential geometric approach to nonlinear filtering: the projection filter, *IEEE T. Automat. Contr.*, **43**, 247–252.
- BRIGO, D. & MERCURIO, F. (2006) *Interest Rate Models - Theory and Practice*, Springer.
- BOLLERSLEV T. (1987) A conditionally heteroskedastic time series model for speculative prices and rates of return, *Rev. Econ. Statist.*, 69,542–547.
- BOLLERSLEV T. (1986) Generalized autoregressive conditional heteroskedasticity *J. Econometrics* **31**, 307–327.
- CHIARELLA, C., HUNG, H. & THUY-DUONG, TÔ (2009) The volatility structure of the fixed income market under the HJM framework: A nonlinear filtering approach, *Comput Stat Data An.*, **53**, 2075–2088.
- CORTAZAR G. & SCHWARTZ E.S. (2003) Implementing a stochastic model for oil futures prices, *Energ. Econ.*, 25, 215–238.
- COX, J.C., INGERSOLL, J.E. & ROSS, S.A. (1985) A theory of the term structure of interest rates, *Econometrica*, **53**, 385–407.
- CRISAN, D. & DOUCET, A. (2002) A survey of convergence results on particle filtering methods for practitioners, *IEEE T. Signal Proces.*, **50**, 736–746.
- DATE, P., JALEN L. & MAMON, R. (2008a) A new algorithm for latent state estimation in nonlinear time series models, *Appl. Math. Comput.* **203**, 224–232.
- DATE, P., MAMON, R. & JALEN L. (2009) A partially linearized sigma point filter for latent state estimation in nonlinear time series models, *J. Comput. Appl. Math.*, In press.
- DATE, P., MAMON, R. & JALEN L. (2008b) A new moment matching algorithm for sampling from partially specified symmetric distributions, *Oper. Res. Lett.* **36**, 669–672.
- DATE P. & WANG, I.C. (2009), Linear Gaussian affine term structure models with unobservable factors: calibration and yield forecasting, *Eur. J. Oper. Res.*, **195**, 156–166.
- DAUM, F. (2005) Nonlinear filters: beyond the Kalman filter, *IEEE Aero. El. Sys. Mag.*, **20**, 57–69.
- DE ROSSI, G. (2004) Kalman filtering of consistent forward rate curves: a tool to estimate and model dynamically the term structure, *J. Empir. Finance* **11**, 277–308.
- DOUCET A., DE FRIETAS, A. & GORDON N. (EDITORS) (2001), *Sequential Monte Carlo Methods in Practice*, Springer.
- DURBIN, J. & KOOPMAN, S.J. (2001), *Time Series Analysis by State Space Methods*, Oxford University Press.
- EVENSEN, G. (1994) Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophy. Res.*, **10**, 143–162.
- FEARNHEAD, F. (2005), Using random quasi-Monte-Carlo within particle filters, with application to financial time series, *J. Comput. Graph. Stat.*, **14**, 751–769.

- FREY, R. & RUNGALDIER, W. (2001) A non-linear filtering approach to volatility estimation with a view towards high frequency data, *Internat. J. Theor. Appl. Finance*, **4**, 199–210.
- GATHERAL, J. (2006) *The Volatility Surface, a Practitioner's Guide*, Wiley Finance.
- GEYER, A. & PICHLER, S. (1993) A state-space approach to estimate and test multifactor cox-ingersoll-ross models of the term structure, *J. Financ. Res.*, **22**, 107–130.
- GLOSTEN, L.R., JAGANATHAN, R. & RUNKLE, D.E. (1993) On the relation between the expected value and the volatility of the nominal excess return on stocks, *J. Finan.*, **48**, 1779–1801.
- GRIMMETT, G. & STIRZAKER, D. (2004) *Probability and Random Processes*, Oxford University Press.
- HANSEN, P.R. & LUNDE, A. (2005), A forecast comparison of volatility models: does anything beat GARCH(1,1)? *J. Appl. Econometrics*, **20**, 873–889.
- HARVEY, A.C. (1989) *Forecasting Structural Time Series Models and the Kalman Filter*, Cambridge University Press.
- HAWKES, R. (2007), *Linear state models for volatility estimation and prediction*, PhD thesis, Brunel University, UK.
- HAWKES, R. & DATE, P. (2007) Medium term horizon volatility forecasting: A comparative study, *Appl. Stoch. Model. Bus. Ind.*, **23**, 465–481.
- HESTON, S. (1993) A closed-form solution for options with stochastic volatility with applications to bond and currency options, *Rev. Financ. Stud.*, **6**, 327–343.
- HØYLAND, K., KAUT, M. & WALLACE, S.W. (2003) a heuristic for moment-matching scenario generation, *Comput. Optim. Appl.*, **24**, 169–185.
- JAMES, J. & WEBBER, N. (2004), *Interest Rate Modelling*, John Wiley.
- JAVAHERI, A., LAUTIER, D. & GALLI, A. (2003) Filtering in finance, *Wilmott*, **5**, 2–18.
- JAVAHERI, A. (2005) *Inside Volatility Arbitrage: the Secrets of Skewness*, Wiley Finance.
- JEGADEESH, N. & PENNACHHI, G. 1996 The behaviour of interest rates implied by the term structure of Eurodollar futures, *Journal of Money, Credit, Banking*, **28**, 426–446.
- JIMENEZ, J.C. & OZAKI, T. (2003) Local linearization filters for nonlinear continuous-discrete state space models with multiplicative noise, *Int. J. Control*, **76**, 1159–1170.
- JULIER, S.J. & UHLMANN, J.K. (2004), Unscented filtering and nonlinear estimation, *Proc. IEEE*, **92**, 401–422.
- KALLIANPUR, G. 1990, *Stochastic Filtering Theory*, Springer Verlag.
- KALMAN, R.E. (1960) A new approach to linear filtering and prediction problems, *J. Basic Eng.*, **82**, 35–45.
- KULHAVÝ, R. (1990) Recursive nonlinear estimation: a geometric approach, *IEEE T. Automat. Contr.*, **26**, 545–555.

- KUSHNER, H. (2000) *Numerical Methods for Stochastic Control Problems in Continuous Time*, Springer Verlag.
- LAUTIER, D. & GALLI, A. (2004) Simple and extended Kalman filters: an application to term structures of commodity prices, *Applied Financial Economics*, **14**, 963–973.
- MARTENS, M. (2002), Measuring and forecasting S&P 500 index-futures volatility using high frequency data, *J. Futures Markets*, **22**, 497–518.
- MANOLIU, M. & TOMPAIDIS, S. (2002) Energy futures prices: term structure models with Kalman filter estimation, *Appl. Math. Finance*, **9**, 21–43.
- MITCHELL H. & HOTEKAMER, P. (1998) Data assimilation using an ensemble Kalman filter technique, *Mon. Weather. Rev.*, **126**, 796–811.
- MONOYIOS, M. (2007) Optimal hedging and parameter uncertainty, *IMA Journal of Management Mathematics*, **18**, 331–351.
- ØKSENDAL, B. (2003) *Stochastic Differential Equations, an Introduction with Applications*, Springer.
- PITT, M.K. & SHEPHERD, N. (1999) Filtering via simulation: auxiliary particle filters, *J. Am. Stat. Assoc.*, **94**, 590–599.
- SCHWARTZ, E.S. (1997) The stochastic behavior of commodity prices: implications for valuation and hedging, *J. Finan.*, **52**, 923–973.
- VASICEK, O. (1977) An equilibrium characterisation of the term structure, *J. Financ. Econ.*, **5**, 177–188.