International Journal of Theoretical and Applied Finance
Vol. 3, No. 2 (2000) 279–308

ESTIMATION IN CONTINUOUS-TIME STOCHASTIC VOLATILITY MODELS USING NONLINEAR FILTERS

JAN NYGAARD NIELSEN, MARTIN VESTERGAARD*

Department of Mathematical Modelling, Technical University of Denmark,
Build. 321, DK-2800 Lyngby, Denmark
E-mail: *{jnn,hm}@imm.dtu.dk*

Received 15 May 1999

The stylized facts of stock prices, interest and exchange rates have led econometricians to propose stochastic volatility models in both discrete and continuous time. However, the volatility as a measure of economic uncertainty is not directly observable in the financial markets. The objective of the continuous-discrete filtering problem considered here is to obtain estimates of the stock price and, in particular, the volatility using discrete-time observations of the stock price. Furthermore, the nonlinear filter acts as an important part of a proposed method for maximum likelihood for estimating embedded parameters in stochastic differential equations. In general, only approximate solutions to the continuous-discrete filtering problem exist in the form of a set of ordinary differential equations for the mean and covariance of the state variables. In the present paper the small-sample properties of a second order filter is examined for some bivariate stochastic volatility models and the new combined parameter and state estimation method is applied to US stock market data.

Keywords: Stochastic volatility, volatility estimation, nonlinear filtering, Monte Carlo simulation.

1. Introduction

Volatility modelling and estimation play an important role in the valuation of financial derivatives and the application of risk management systems (for Value-at-risk computations [47] or more coherent risk measures, see e.g. [4, 16]).

The Black–Scholes model is routinely used to evaluate the price of European type options, even though it is known to produce systematic pricing biases, e.g. it does not accommodate volatility smiles. It is now well-documented that stock returns exhibit leptokurtosis, skewness and pronounced conditional heteroscedasticity in the form of volatility clustering, all features that are at odds with the Black–Scholes assumptions.

A number of continuous-time stochastic volatility models have been proposed in the literature under the natural assumption that the failure of the Black–Scholes

*Current affiliation: The Hull Group, L.L.C., A Goldman Sachs company, 311 S. Wacker Drive, Suite 1400, Chicago, Illinois 60606. Email: mvesterg@htc.com

model is attributable to the assumed geometric Brownian motion model for the stock price dynamics. Valuation of options under stochastic volatility is covered in [3, 7, 36, 39, 66, 77], whereas [31] treats the valuation of futures, and [38] considers hedging. Market completeness is, in general, not attainable in stochastic volatility models, yet [37, 78] suggest a methodology that provides a complete market. Discrete-time stochastic volatility models have been proposed by [17, 30, 34, 35] among others. In discrete-time, a plethora of related autoregressive conditional heteroskedastic (ARCH) models have been proposed, see e.g. [10, 68, 73] and the collection of papers in [22, 63].

Until recent years, there has been a dichotomy between the discrete-time models favored for empirical work on economic and financial time series and the continuous-time models typically used in theoretical work on asset pricing. There is a growing literature devoted to closing this gap by considering the continuous-time models obtained when the sampling time tends to zero with obvious applications for high-frequency data, see e.g. [63]. As opposed to ARMA-models, GARCH models are not, in general, closed under temporal aggregation, which implies that there does not exist a continuous-time counterpart of any GARCH model. It has been shown that jump-diffusion models may arise as limits of a limited class of ARCH models [55] when the time gap between observations falls. Similar results have been obtained [18] for the widely used GARCH(1,1) model [9] and for a very large class of GARCH models [19]. However, for the econometrician, there is no particular reason for restricting the class of continuous-time stochastic volatility models to those that are attainable as limits of GARCH models.

Parameter estimation in discretely observed diffusion processes with unobserved states is an inherently difficult problem to which a number of solutions have been proposed in the literature. The fundamental problem is that the exact transition density functions, and hence the likelihood function, cannot generally be expressed in closed form. For univariate models [1] has proposed a method, where the Kolmogorov forward equation is used to extract a semi-nonparametric estimator of the diffusion function when the drift function is given. The small sample properties of this method has been studied in [61]. Using the same basic framework [70] has proposed a semi-nonparametric method for estimating discrete-time approximations of the drift and diffusion functions, and [43] suggested a slightly different approach.[a] The work by [1, 70] has been studied extensively by [12], where it is shown that these kernel based methods may yield spurious non-linearities. Unfortunately, it is difficult to extend these methods to cope with multivariate diffusion processes, in particular processes with unobserved states.

Another branch of the literature has extended the [60] approach of considering the estimation problem as a missing value problem by using Markov Chain Monte Carlo Methods [21, 25, 46].

A third branch of the literature deals with simulation-based methods and moment matching. The simulated method of moments [20] obtains moment conditions

[a]In [44], the small sample properties of a number of nonparametric estimators have been compared.

by matching the sample moments with simulated moments from the proposed model. The basic idea behind the indirect inference method [30, 69] is to match the moments of a discrete-time auxiliary model identified and estimated using the real data with the moments of a discretized version of the proposed continuous time model. This approach is taken for a stochastic volatility model by [24]. The efficient method of moments [26] is a very similar, but slightly refined method.[b] This method has been applied to stochastic volatility models in [27, 28]. The drawback of the very general EMM method is that it requires both an auxiliary model and a continuous time model that captures all the features of the data and a one-to-one mapping between the parameters of these models.

The main contribution of this paper is an estimation and filtering method that allows the estimation of embedded parameters in multivariate continuous-time stochastic volatility models using discrete-time observations.

A growing interest for filtering methods in the financial engineering literature is easily detected: Filtering methods for structural models in discrete-time is covered in [33] and applied to discrete-time stochastic volatility models in [35]. Filtering of volatility from stock prices in an ARCH-framework is considered in e.g. [56, 57], and volatility estimation in a linear structural model is treated in [74] using an ordinary Kalman filter.

For nonlinear systems the extended Kalman filter provides an approximate solution to the filtering problem [42], but for SDEs with a state-dependent diffusion function higher order filters[c] are needed [52]. A maximum likelihood method for direct estimation of embedded parameters in SDEs is proposed in [58], which also covers a generalization of the transformation proposed in [5] such that the extended Kalman filter may be applied to a class of SDEs with a state-dependent diffusion term. Stochastic volatility models do not, however, belong to this class such that higher order filters must be used. Furthermore, nonlinear filtering methods make it possible to estimate unobservable states in a large class of continuous-time models. In the present paper this will be demonstrated using stochastic volatility models. Hence the proposed method allows for a simultaneous estimation of the parameters and the unobserved states such that actual estimates of the stochastic volatility are provided.

The paper is organized as follows: Sec. 2 presents the continuous-time bivariate stochastic volatility models to be considered. Section 3 describes the second order filters that provides an approximate solution to the continuous-discrete filtering problem. Section 4 provides simulation studies in order to validate the proposed filtering and estimation method. Section 5 contains empirical work on US stock market data, and Sec. 6 concludes.

[b] An examination of the relative efficiency of the EMM method is reported in [29].
[c] Higher order filters and some extension for discrete-time state space models have been compared by [72] using Monte Carlo simulation.

2. Bivariate Stochastic Volatility Models

This section provides an overview of previously proposed diffusion processes for modelling of stochastic volatility. It is customary to extend the Black–Scholes model by letting the volatility itself be modelled as a diffusion process, i.e.

$$dS_t = \alpha S_t dt + \sigma_t S_t dW_t^1 , \tag{1}$$

$$d\psi(\sigma_t) = a(\sigma_t)dt + b(\sigma_t)dW_t^2 , \tag{2}$$

where W_t^1, W_t^2 are correlated Wiener processes with correlation coefficient ρ. The instantaneous rate-of-return is α, $\psi(\sigma_t)$ is some mapping of σ_t, $a(\sigma_t)$ and $b^2(\sigma_t)$ account for the instantaneous mean and variance, respectively, of the stochastic volatility $\{\sigma_t\}$ and ρ accounts for the socalled leverage effect, i.e. the fact that large upward moves in equity markets typically have smaller volatility impacts than large downwards moves of the same magnitude. Table 1 lists a number of the specifications given in the literature.

The real-valued discrete-time observations $\{Y_{t_i}\}$ are obtained at the sampling instants $t_1 < \cdots < t_i < \cdots < t_N$, where N denotes the number of observations. The observation equation is

$$Y_{t_i} = S_{t_i} + e_{t_i} , \quad i = 1, \ldots, N , \tag{3}$$

where $\{e_{t_i}\}$ is a Gaussian white noise process with mean zero and variance Σ_{t_i}. The stochastic entities $\mathbf{W}_t = (W_t^1, W_t^2)^T$ and e_{t_i} are assumed to be mutually independent for all t and t_i.

As shown in Table 1, a plethora of continuous-time stochastic volatility models have been proposed in the literature. The modelling framework (1)–(3) makes it possible to specify and estimate very general stochastic volatility models. In particular estimates of the stochastic volatility σ_t and the parameters in the underlying SDE (2) may be obtained. However the unique identification of such models is inherently

Table 1. An overview of a number of bivariate stochastic volatility models on the form (1)–(2).

$\psi(\sigma_t)$	$a(\sigma_t)$	$b(\sigma_t)$	Reference(s)
σ_t	$\kappa(\beta - \sigma_t)$	ξ	[71]
σ_t	$\kappa(\beta - \sigma_t)$	$\xi\sigma_t$	[14]
σ_t	$\kappa\sigma_t(\beta - \sigma_t)$	$\xi\sigma_t$	[66]
σ_t	$\kappa\sigma_t$	$\xi\sigma_t$	[39, 45, 78]
σ_t	$(\beta - \kappa\sigma_t^2)/\sigma_t$	ξ	[6, 36, 40]
σ_t	$\kappa(\beta - \sigma_t^2)$	$\xi\sigma_t$	[54]
$\ln \sigma_t$	$\kappa(\beta - \ln \sigma_t)$	ξ	[67, 77]

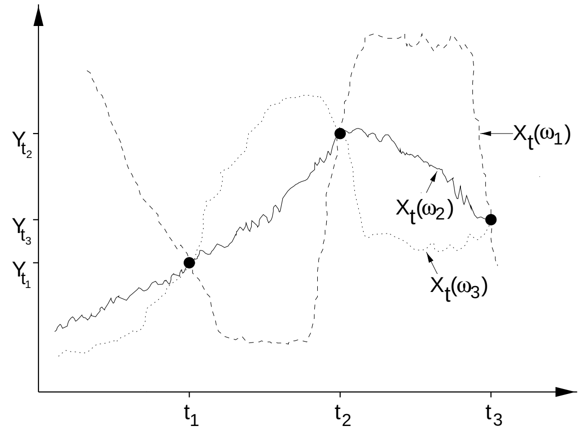


Fig. 1. The aliasing problem: An infinite number of sample paths may give rise to the same observed data set. Only three sample paths have been sketched here.

limited due to the aliasing problem.[d] To be precise, let $\mathcal{F}_t = \sigma\{Y_{t_i}\}$ denote the information set generated by the observations $\{Y_{t_i}\}$ and let $\mathcal{G}_t = \sigma\{X_t, Y_{t_i}\}$ denote the information set generated by the states $\{\mathbf{X}_s\}_{t_0 \leq s \leq t}$ and the observations $\{Y_{t_i}\}$, $i = 1, \ldots, N$. Obviously $\mathcal{F}_t \subseteq \mathcal{G}_t$, which implies that infinitely many sample paths of (1)–(2) observed through (3) may yield the same observations, as illustrated in Fig. 1, such that (1)–(2) cannot be uniquely identified from $\mathcal{F}_t$.

3. Nonlinear Filtering Techniques

In this section the continuous-discrete nonlinear filtering problem will be described for a general stochastic state space model and the approximations made to obtain the second order filter will be discussed in detail. The presentation follows [52].

Assume that a general model for the state variables $\mathbf{X}_t \in \mathbb{R}^n$ is given by

$$d\mathbf{X}_t = \mathbf{f}(\mathbf{X}_t; \boldsymbol{\theta})dt + \mathbf{G}(\mathbf{X}_t; \boldsymbol{\theta})d\mathbf{W}_t\,; \quad \mathbf{X}_{t_0} = \mathbf{X}_0\,, \tag{4}$$

where $\mathbf{X}_0$ is a stochastic initial condition satisfying $E[|\mathbf{X}_0|^2] < \infty$, the drift function $\mathbf{f} : \mathbb{R}^n \times \mathbb{R}^p \mapsto \mathbb{R}^n$ and the diffusion function $\mathbf{G}: \mathbb{R}^n \times \mathbb{R}^p \mapsto \mathbb{R}^{n \times d}$ are assumed to be known up to the unknown parameter vector $\boldsymbol{\theta} \subseteq \boldsymbol{\Theta} \in \mathbb{R}^p$ and $\mathbf{W}_t = (W_t^1, \ldots, W_t^d)^T$ is a d-dimensional Wiener process with incremental covariance $\mathbf{Q}_t$ defined on the usual probability space $(\Omega, \mathcal{F}, P)$, see [62] for the technical details. Assume that $\mathbf{f}$ and $\mathbf{G}$ satisfy necessary and sufficient conditions to ensure the existence of unique solutions to (4), and that they are twice continuously differentiable with respect to $\mathbf{X}_t$.

[d]For first order scalar SDEs some progress has been made, see [2, 32], where the latter shows that the aliasing problem does not exist for scalar time reversible SDEs.

Further assume that observations are made available at discrete time instants $t_1 < \cdots < t_i < \cdots t_N$, where N denotes the number of observations. The relation between the state variables and the observations is given by the observation equation. The observation equation is given by

$$\mathbf{Y}_{t_i} = \mathbf{h}(\mathbf{X}_{t_i}; \boldsymbol{\theta}) + \mathbf{e}_{t_i}\,, \tag{5}$$

where $\mathbf{h}: \mathbb{R}^n \times \mathbb{R}^p \mapsto \mathbb{R}^m$ is a known function, which is assumed to be twice continuously differentiable with respect to $\mathbf{X}_t$. Finally $\{\mathbf{e}_{t_i}\}$ is a m-dimensional zero mean Gaussian white noise process with covariance $\boldsymbol{\Sigma}_{t_i}$. The stochastic entities $\mathbf{X}_0$, $\mathbf{W}_t$ and $\mathbf{e}_{t_i}$ are assumed to be mutually independent for all t and t_i.

Remark 3.1. Clearly the discretely observed stochastic volatility model (1)–(3) fits into the general specification (4)–(5) by defining the state vector as $\mathbf{X}_t = (S_t, \psi(\sigma_t))^T$, and the observation equation as $h(\mathbf{X}_{t_i}; \boldsymbol{\theta}) = S_{t_i}$. The proposed methodology can also be applied to the augmented state space vector $\mathbf{X}_t = (S_t, \psi(\sigma_t), \alpha_t)^T$ if a SDE is specified for the stochastic rate-of-return α_t, see [75].

The filtering problem and the filtering equations will be derived only in the univariate case ($n = m = d = 1$) with $h(X_{t_i}; \boldsymbol{\theta}) = X_{t_i}$ in Sec. 3.1, but the general results will be stated for the multivariate case in Sec. 3.2.

3.1. *The univariate case*

Let $p_x(x_t, t_i|\xi, t_{i-1})$ denote the conditional probability density function (cpdf) of being in state $X_t = x_t$ at time t_i given that the process was in state $X_{t_{i-1}} = \xi$ at time t_{i-1}. It is well known that the solution to (4) is a Markov process, so the process is completely characterized by the sequence of transition densities $p_x(x_t, t_i|\xi, t_{i-1})$. The time evolution of $p_x(x_t, t|\xi, t_{i-1})$ for $t \in [t_{i-1}, t_i)$, i.e. between sampling instants t_{i-1} and t_i, is given as the solution to the Kolmogorov forward equation

$$\frac{\partial p_x(x_t, t|\xi, t_{i-1})}{\partial t} = -\frac{\partial[p_x(x_t, t|\xi, t_{i-1}) f(x_t; \boldsymbol{\theta})]}{\partial x_t} + \frac{1}{2} Q_t \frac{\partial^2 [p_x(x_t, t|\xi, t_{i-1}) G^2(x_t; \boldsymbol{\theta})]}{\partial x_t^2} \tag{6}$$

for $t \in [t_{i-1}, t_i)$, where it is assumed that the continuous partial derivatives exist.

Let $p_{x|y}(x_t, t_i|\mathcal{F}_i)$ denote cpdf of being in state $X_{t_i} = x_t$ at time t_i given observations up to and including time t_i, where $\mathcal{F}_i$ is short for $\mathcal{F}_{t_i}$, i.e. the σ-algebra generated by the observations up to time t_i. This cpdf can be obtained as follows: First consider the time propagation of $p_{x|y_-}(x_t, t_i|\mathcal{F}_{i-1})$ from sample time t_{i-1} to t_i. Knowing $p_x(x_t, t|\xi, t_{i-1})$, $t \in [t_{i-1}, t_i)$, makes it possible to compute $p_{x|y_-}(x_t, t_i|\mathcal{F}_{i-1})$ using the Chapman–Kolmogorov equation:

$$p_{x|y_-}(x_t, t_i|\mathcal{F}_{i-1}) = \int_{-\infty}^{\infty} p_x(x_t, t|\xi, t_{i-1}) p_{x_-|y_-}(\xi, t_{i-1}|\mathcal{F}_{i-1}) d\xi\,, \tag{7}$$

where $p_x(x_t, t|\xi, t_{i-1})$ is given as the solution to (6) and $p_{x_-|y_-}(\xi, t_{i-1}|\mathcal{F}_{i-1})$ is obtained from the previous observation update. Indeed, it can be shown that $p_{x|y_-}(x, t_i|\mathcal{F}_{i-1})$ itself satisfies (6) with an initial condition given by the previous observation update $p_{x_-|y_-}(\xi, t_{i-1}|\mathcal{F}_{i-1})$.

When a new observation $Y_{t_i} = y_{t_i}$ becomes available at time t_i, an update of the cpdf of the state $p_{x|y}$ follows from Bayes' formula, i.e.

$$p_{x|y}(x_t, t_i|\mathcal{F}_i) = \frac{p_{y|x,y_-}(\eta_{t_i}|x_t, \mathcal{F}_{i-1})p_{x|y_-}(x_t, t_{i-1}|\mathcal{F}_{i-1})}{p_{y|y_-}(\eta_{t_i}|\mathcal{F}_{i-1})}, \tag{8}$$

where the second numerator term has just been derived. For the first numerator term it is easily seen that

$$p_{y|x,y_-}(\eta_{t_i}|x_t, \mathcal{F}_{i-1}) = p_{y|x}(\eta_{t_i}|x_t) = \frac{1}{\sqrt{2\pi\Sigma_{t_i}}} \exp\left(-\frac{(\eta_{t_i} - x_{t_i})^2}{2\Sigma_{t_i}}\right). \tag{9}$$

The denominator in (8) may be computed using

$$\begin{aligned} p_{y|y_-}(\eta_{t_i}|\mathcal{F}_{i-1}) &= \int_{-\infty}^{\infty} p_{y,x|y_-}(x_t, \xi|\mathcal{F}_{i-1})d\xi \\ &= \int_{-\infty}^{\infty} p_{y|x}(x_t|\xi)p_{x|y_-}(\xi|\mathcal{F}_{i-1})d\xi\,. \end{aligned} \tag{10}$$

Knowing $p_{y|y_-}(\eta_{t_i}|\mathcal{F}_{i-1})$ from the above propagation and $p_{y|x}(x_t|\xi)$ from (9) provide sufficient information to compute $p_{x|y}(x_t, t_i|\mathcal{F}_i)$ using (8). Equations (6)–(10) constitute the general continuous-discrete time filtering problem. Unfortunately, except for a few special cases (e.g. narrow-sense linear systems), closed form solutions to these equations are not available. The computation of the entire density function $p_{x|y}(x_t, t_i|\mathcal{F}_i)$, which provides the connection between the evolution of the state variable and the observations, requires the solution of partial integro-differential equations (derived by means of the Kolmogorov forward equation) and observation updates involve solving functional integral difference equations (derived by means of the Bayes' formula). This implies that the general optimal nonlinear filter will be infinite dimensional. For practical purposes expansions truncated to some low order are required both in the time propagation and observation update of the nonlinear filter. One possible approach is to consider expansions of some of the conditional moments, and this will be pursued in the following. Other approaches are described in [52].

3.1.1. *Conditional moments estimator*

Let $\hat{X}_{t|t_{i-1}} = E[X_t|\mathcal{F}_{i-1}] = E_{i-1}[X_t]$ and $P_{t|t_{i-1}} = E[(X_t - \hat{X}_{t|t_{i-1}})^2|\mathcal{F}_{i-1}]$, $t \in [t_{i-1}, t_i)$, denote the conditional mean and variance, respectively. Using the Kolmogorov forward Eq. (6) the propagation of these moments between sampling instants t_{i-1} and t_i may be shown to satisfy

$$\frac{d\hat{X}_{t|t_{i-1}}}{dt} = E_{i-1}[f(X_t; \boldsymbol{\theta})]\,, \tag{11}$$

$$\frac{dP_{t|t_{i-1}}}{dt} = 2E_{i-1}[f(X_t;\boldsymbol{\theta})X_t] + Q_t E_{i-1}[G^2(X_t;\boldsymbol{\theta})] - 2E_{i-1}[f(X_t;\boldsymbol{\theta})]E_{i-1}[X_t]\,, \tag{12}$$

for $t \in [t_{i-1}, t_i)$. Note that these are not ordinary differential equations (ODEs), because $E_{i-1}[\cdot]$ depends on the cpdfs $p_x(x_t, t|\mathcal{F}_{i-1})$, i.e. all the moments of the conditional density. However, by disregarding or restricting the moments of higher order than two it is possible to derive an approximate set of prediction and updating equations that have a structure similar to the ordinary Kalman filter [33].

Hence by performing Taylor expansions of $f(\cdot)$ and $G(\cdot)$ about the current estimate $\hat{X}_{t|t_{i-1}}$, truncating after the second order terms and taking expectations, the following prediction equations are obtained:

$$\frac{d\hat{X}_{t|t_{i-1}}}{dt} = f(\hat{X}_{t|t_{i-1}};\boldsymbol{\theta}) + \frac{P_{t|t_{i-1}}}{2}\frac{\partial^2 f(\hat{X}_{t|t_{i-1}};\boldsymbol{\theta})}{\partial X_t^2} \tag{13}$$

$$\begin{aligned}\frac{dP_{t|t_{i-1}}}{dt} &= 2\frac{\partial f(\hat{X}_{t|t_{i-1}})}{\partial X_t}P_{t|t_{i-1}} + Q_t\Bigg[G^2(\hat{X}_{t|t_{i-1}};\boldsymbol{\theta}) \\ &\quad + \left(\frac{\partial G(\hat{X}_{t|t_{i-1}};\boldsymbol{\theta})}{\partial X_t}\right)^2 P_{t|t_{i-1}} + G(\hat{X}_{t|t_{i-1}};\boldsymbol{\theta})\frac{\partial^2 G(\hat{X}_{t|t_{i-1}};\boldsymbol{\theta})}{\partial X_t^2}P_{t|t_{i-1}} \\ &\quad + \frac{3}{4}\frac{\partial^2 G(\hat{X}_{t|t_{i-1}};\boldsymbol{\theta})}{\partial X_t^2}^2 P_{t|t_{i-1}}^2\Bigg]\end{aligned} \tag{14}$$

for $t \in [t_{i-1}, t_i)$, where it has further been assumed that the transition density is sufficiently close to the Gaussian density to ensure that the third and higher order odd central moments are essentially zero, i.e. $E[(X_t - \hat{X}_{t|t_{i-1}})^{2j+1}] \approx 0$, $j = 1, 2, \ldots$, that the fourth central moment may be expressed in terms of the variance, i.e. $E[(X_t - \hat{X}_{t|t_{i-1}})^4] = 3P_{t|t_{i-1}}^2$, and that the sixth and higher order even central moments are negligible, i.e. $E[(X_t - \hat{X}_{t|t_{i-1}})^{2j}] \approx 0$, $j = 3, 4 \ldots$. This filter is called the Gaussian truncated second order filter. Without the last term in (14), the filter is called the truncated second order filter. This filter ignores all central moments of X_t higher than second order. If $G(X_t;\boldsymbol{\theta})$ does not depend on X_t, then the extended Kalman filter is obtained. For illustration, Eq. (13) may be obtained as follows: A Taylor series expansion of $f(X_t;\boldsymbol{\theta})$ about the conditional mean, i.e. the current estimate $\hat{X}_{t|t_{i-1}}$, yields

$$\begin{aligned}f(X_t;\boldsymbol{\theta}) &= f(\hat{X}_{t|t_{i-1}};\boldsymbol{\theta}) + \left.\frac{\partial f(x;\boldsymbol{\theta})}{\partial x}\right|_{x=\hat{X}_{t|t_{i-1}}}\left(X_t - \hat{X}_{t|t_{i-1}}\right) \\ &\quad + \frac{1}{2}\left.\frac{\partial^2 f(x;\boldsymbol{\theta})}{\partial x^2}\right|_{x=\hat{X}_{t|t_{i-1}}}\left(X_t - \hat{X}_{t|t_{i-1}}\right)^2 + \cdots\,.\end{aligned} \tag{15}$$

By taking the conditional expectation $E_{i-1}[\cdot]$ on both sides of (15), the second term on the right-hand side clearly drops out such that (13) is obtained using the definition of $P_{t|t_{i-1}}$. Equation (14) is obtained in a similar manner, see [52] for more technical details.

Remark 3.2. The last term in (13) is often called a bias-correction term, because it gives rise to less biased state estimates than the extended Kalman filter [52].

After having obtained the observation Y_{t_i} at time t_i the conditional mean and variance can be improved or updated. The approximative updating equations are

$$\hat{X}_{t_i|t_i} = \hat{X}_{t_i|t_{i-1}} + K_{t_i} \left\{ Y_{t_i} - \hat{X}_{t_i|t_{i-1}} \right\} , \tag{16}$$

$$P_{t_i|t_i} = (1 - K_{t_i}) P_{t_i|t_{i-1}} , \tag{17}$$

where the Kalman gain K_{t_i} is

$$K_{t_i} = \frac{P_{t_i|t_{i-1}}}{P_{t_i|t_{i-1}} + \Sigma_{t_i}} . \tag{18}$$

The Kalman gain K_{t_i} describes the weight of the information provided by a new observation depending upon the variance associated with the state estimate, $P_{t_i|t_{i-1}}$, and the variance of the observation Σ_{t_i}. Equations (13)–(18) constitute the modified Gaussian second order filter [52]. This filter may be applied as follows: Assume that the functions $f(\cdot)$ and $G(\cdot)$, and the true parameters $\boldsymbol{\theta}$ (including Σ_{t_i} and Q_t) are known. Assume further that initial guesses of the state estimate $\hat{X}_{t_1|t_0}$, the associated variance $P_{t_1|t_0}$ and henceforth the Kalman gain K_{t_1} are provided. When the first observation Y_{t_1} is obtained, an update of the state estimate (16)–(17) may be computed. These serve as initial conditions for the ODEs (13)–(14) such that $\hat{X}_{t|t_1}$ and $P_{t|t_1}$, $t \in [t_1, t_2)$, may be computed using a numerical ODE solver, e.g. a Runge–Kutta method. When the next observation Y_{t_2} is obtained the new state estimate $\hat{X}_{t_2|t_2}$ and $P_{t_2|t_2}$ is obtained using (16)–(17), and so forth.

In [59] this filter is applied to the constant elasticity of variance model [15]. This filter and the extended Kalman filter (using a transformation) is compared for the interest model proposed by [11] in [5]. In the following attention will be concentrated on the multivariate case. Typically, a numerical method is needed to solve the ordinary differential Eqs. (13)–(14).

3.2. *The multivariate case*

In this section the modified truncated second order filter and the modified Gaussian second order filter will be stated for the general modelling framework (4)–(5) as immediate generalizations of the filters presented in the previous section. A multivariate version of the derivation of the exact filtering problem given in Sec. 3.1 does not offer any additional insight, so it will be skipped here.

3.2.1. *The truncated second order filter*

The time propagation equations are

$$\frac{d\hat{\mathbf{X}}_{t|t_{i-1}}}{dt} = \mathbf{f}(\hat{\mathbf{X}}_{t|t_{i-1}};\boldsymbol{\theta}) + E_{i-1}[\mathbf{B}_{t|t_{i-1}}]\,, \tag{19}$$

$$\begin{aligned}\frac{d\mathbf{P}_{t|t_{i-1}}}{dt} &= \mathbf{F}(\hat{\mathbf{X}}_{t|t_{i-1}};\boldsymbol{\theta})\mathbf{P}_{t|t_{i-1}} + \mathbf{P}_{t|t_{i-1}}\mathbf{F}^T(\hat{\mathbf{X}}_{t|t_{i-1}};\boldsymbol{\theta}) \\ &\quad + E_{i-1}\left[\mathbf{G}(\hat{\mathbf{X}}_{t|t_{i-1}};\boldsymbol{\theta})\mathbf{Q}_t\mathbf{G}^T(\hat{\mathbf{X}}_{t|t_{i-1}};\boldsymbol{\theta})\right]\,,\end{aligned} \tag{20}$$

with the initial conditions $\hat{\mathbf{X}}_{t_{i-1}|t_{i-1}}$ and $\mathbf{P}_{t_{i-1}|t_{i-1}}$.

The bias-correction term $E_{i-1}[\mathbf{B}_{t|t_{i-1}}]$ is a n-dimensional vector with the kth component

$$E^k_{i-1}[\mathbf{B}_{t|t_{i-1}}] = \frac{1}{2}\mathrm{tr}\left\{\frac{\partial^2 \mathbf{f}^k(\mathbf{x};\boldsymbol{\theta})}{\partial \mathbf{x}^2}\mathbf{P}_{t|t_{i-1}}\right\}\Bigg|_{\mathbf{x}=\hat{\mathbf{X}}_{t|t_{i-1}}} \tag{21}$$

and $\mathbf{F}(\hat{\mathbf{X}}_{t|t_{i-1}};\boldsymbol{\theta})$ is given by the $n\times n$ matrix

$$\mathbf{F}(\hat{\mathbf{X}}_{t|t_{i-1}};\boldsymbol{\theta}) = \frac{\partial \mathbf{f}(\mathbf{x};\boldsymbol{\theta})}{\partial \mathbf{x}}\Bigg|_{\mathbf{x}=\hat{\mathbf{X}}_{t|t_{i-1}}}\,. \tag{22}$$

The last term in (20) is a $n\times n$ symmetric matrix with element ij given by (where the dependence on $\hat{\mathbf{X}}_{t|t_{i-1}}$, $t|t_{i-1}$, and $\boldsymbol{\theta}$ have been dropped for convenience)

$$\begin{aligned}E^{ij}_{i-1}[\mathbf{G}\mathbf{Q}_t\mathbf{G}^T] &= \sum_{k=1}^{d}\sum_{l=1}^{d}\mathbf{G}^{ik}\mathbf{Q}^{kl}_t(\mathbf{G}^T)^{lj} + \mathrm{tr}\left\{\left(\frac{\partial \mathbf{G}^{ik}}{\partial \mathbf{x}}^T\mathbf{Q}^{kl}_t\frac{\partial(\mathbf{G}^T)^{lj}}{\partial \mathbf{x}}\right)\mathbf{P}\right\} \\ &\quad + \frac{1}{2}\mathbf{G}^{ik}\mathbf{Q}^{kl}_t\mathrm{tr}\left\{\frac{\partial^2(\mathbf{G}^T)^{lj}}{\partial \mathbf{x}^2}\mathbf{P}\right\} + \frac{1}{2}\mathrm{tr}\left\{\mathbf{P}\frac{\partial^2\mathbf{G}^{ik}}{\partial \mathbf{x}^2}\right\}\mathbf{Q}^{kl}_t(\mathbf{G}^T)^{lj}\,.\end{aligned} \tag{23}$$

Remark 3.3. Notice that $\mathbf{G}^{ik}$ denotes element ik of $\mathbf{G}$, whereas $(\mathbf{G}^T)^{lj}$ denotes element lj of the transpose of $\mathbf{G}$. Also notice that the partial derivative of a scalar with respect to a vector yields a row vector such that, say, $\frac{\partial(\mathbf{G}^T)^{lj}}{\partial \mathbf{x}}$ is a row vector, and $\frac{\partial \mathbf{G}^{ik}}{\partial \mathbf{x}}^T$ is a column vector.

The updating equations are given by

$$\begin{aligned}\mathbf{A}_{t_i} &= \mathbf{H}(\hat{\mathbf{X}}_{t_i|t_{i-1}};\boldsymbol{\theta})\mathbf{P}_{t_i|t_{i-1}}\mathbf{H}^T(\hat{\mathbf{X}}_{t_i|t_{i-1}};\boldsymbol{\theta}) - E_{i-1}[\tilde{\mathbf{B}}_{t_i|t_{i-1}}]E^T_{i-1}[\tilde{\mathbf{B}}_{t_i|t_{i-1}}] \\ &\quad + \boldsymbol{\Sigma}_{t_i}\,,\end{aligned} \tag{24}$$

$$\mathbf{K}_{t_i} = \mathbf{P}_{t_i|t_{i-1}}\mathbf{H}^T(\hat{\mathbf{X}}_{t_i|t_{i-1}};\boldsymbol{\theta})\mathbf{A}^{-1}_{t_i}\,, \tag{25}$$

$$\hat{\mathbf{X}}_{t_i|t_i} = \hat{\mathbf{X}}_{t_i|t_{i-1}} + \mathbf{K}_{t_i}\left\{\mathbf{y}_{t_i} - \mathbf{h}(\hat{\mathbf{X}}_{t_i|t_{i-1}};\boldsymbol{\theta}) - E_{i-1}[\tilde{\mathbf{B}}_{t_i|t_{i-1}}]\right\}\,, \tag{26}$$

$$\mathbf{P}_{t_i|t_i} = \mathbf{P}_{t_i|t_{i-1}} - \mathbf{K}_{t_i}\mathbf{H}(\hat{\mathbf{X}}_{t_i|t_{i-1}} : \boldsymbol{\theta})\mathbf{P}_{t_i|t_{i-1}}\,, \tag{27}$$

where $\mathbf{H}(\hat{\mathbf{X}}_{t_i|t_{i-1}};\boldsymbol{\theta})$ is defined as the $m \times n$ matrix

$$\mathbf{H}(\hat{\mathbf{X}}_{t_i|t_{i-1}};\boldsymbol{\theta}) = \left.\frac{\partial \mathbf{h}(\mathbf{x};\boldsymbol{\theta})}{\partial \mathbf{x}}\right|_{\mathbf{x}=\hat{\mathbf{X}}_{t_i|t_{i-1}}} \tag{28}$$

and the bias-correction term $E_{i-1}[\tilde{\mathbf{B}}_{t_i|t_{i-1}}]$ is a $m\times 1$-vector with the kth component given by

$$E^k_{i-1}[\tilde{\mathbf{B}}_{t_i|t_{i-1}}] = \frac{1}{2}\mathrm{tr}\left.\left\{\frac{\partial^2 \mathbf{h}^k(\mathbf{x};\boldsymbol{\theta})}{\partial \mathbf{x}^2}\mathbf{P}_{t_i|t_{i-1}}\right\}\right|_{\mathbf{x}=\hat{\mathbf{X}}_{t_i|t_{i-1}}}. \tag{29}$$

Remark 3.4. For a nonlinear observation function, Eq. (24) shows that the observation update equations also includes a bias-correction term as opposed to the linear case (16). Notice that the bias-correction term (29) drops out if the observation Eq. (5) is linear in the state variables.

Higher order filters can be obtained by including higher order terms from the Taylor series expansions of $\mathbf{f}$ and $\mathbf{G}$. However, the severe computational disadvantages makes such filters infeasible, and it is generally recommended to use the first or second order filters on better models. The numerical work is considerably more demanding for the multivariate case, i.e. it involves the numerical solution of $n + \frac{n}{2}(n+1) = \frac{n}{2}(n+3)$ ODEs for the conditional first and second order central moments given by (19)–(20) between each sampling instant.

3.2.2. *The Gaussian second order filter*

The prediction equations for the modified Gaussian second order filter are very similar to those for the truncated version. The only difference is the computation of (23), where the results may conveniently be expressed in terms of

$$\tilde{\mathbf{G}}(\hat{\mathbf{X}}_{t|t_{i-1}};\boldsymbol{\theta}) = \mathbf{G}(\hat{\mathbf{X}}_{t|t_{i-1}};\boldsymbol{\theta})\mathbf{Q}_t^{1/2}$$

so that $E_{i-1}[\mathbf{G}\mathbf{Q}_t\mathbf{G}^T] = E_{i-1}[\tilde{\mathbf{G}}\tilde{\mathbf{G}}^T]$ is a $n \times n$ symmetric matrix with element ij given by

$$\begin{aligned} E^{ij}_{i-1}[\mathbf{G}\mathbf{Q}_t\mathbf{G}^T] = \sum_{k=1}^{d}\Bigg[&\tilde{\mathbf{G}}^{ik}(\tilde{\mathbf{G}}^T)^{kj} + \mathrm{tr}\left\{\left(\frac{\partial \tilde{\mathbf{G}}^{ik}}{\partial \mathbf{x}}^T \frac{\partial (\tilde{\mathbf{G}}^T)^{kj}}{\partial \mathbf{x}}\right)\mathbf{P}\right\} \\ &+ \frac{1}{2}\tilde{\mathbf{G}}^{ik}\mathrm{tr}\left\{\frac{\partial^2 (\tilde{\mathbf{G}}^T)^{kj}}{\partial \mathbf{x}^2}\mathbf{P}\right\} + \frac{1}{2}\mathrm{tr}\left\{\mathbf{P}\frac{\partial^2 \tilde{\mathbf{G}}^{ik}}{\partial \mathbf{x}^2}\right\}(\tilde{\mathbf{G}}^T)^{kj} \\ &+ \frac{1}{4}\mathrm{tr}\left\{\frac{\partial^2 \tilde{\mathbf{G}}^{ik}}{\partial \mathbf{x}^2}\mathbf{P}\right\}\mathrm{tr}\left\{\frac{\partial^2 (\tilde{\mathbf{G}}^T)^{kj}}{\partial \mathbf{x}^2}\mathbf{P}\right\} \\ &+ \frac{1}{2}\mathrm{tr}\left\{\frac{\partial^2 \tilde{\mathbf{G}}^{ik}}{\partial \mathbf{x}^2}\mathbf{P}\frac{\partial^2 (\tilde{\mathbf{G}}^T)^{kj}}{\partial \mathbf{x}^2}\mathbf{P}\right\}\Bigg]. \end{aligned} \tag{30}$$

Remark 3.5. Notice that the first four terms in (30) replicate (23) and the last two terms correspond to the last term for the scalar case (14).

The Gaussian version of the second order filter is obtained by approximating the fourth central moments with the values they would assume if the density were in fact Gaussian, i.e.

$$E_{i-1}[(X^i - \hat{X}^i)(X^j - \hat{X}^j)(X^k - \hat{X}^k)(X^l - \hat{X}^l)]$$
$$= P^{ij}P^{kl} + P^{ik}P^{jl} + P^{il}P^{jk} \tag{31}$$

for $i, j, k, l = 1, \ldots, n$, where the time argument of $\hat{\mathbf{X}}_{t|t_{i-1}} = (\hat{X}^1_{t|t_{i-1}}, \ldots, \hat{X}^n_{t|t_{i-1}})^T$ and $\mathbf{P}_{t|t_{i-1}}$ has been left out for brevity. The updating equations for the modified Gaussian second order filter are almost similar to the truncated version. The only difference is the equations for the bias-correction term.

$$\mathbf{A}_{t_i} = \mathbf{H}(\hat{\mathbf{X}}_{t_i|t_{i-1}}; \boldsymbol{\theta})\mathbf{P}_{t_i|t_{i-1}}\mathbf{H}^T(\hat{\mathbf{X}}_{t_i|t_{i-1}}; \boldsymbol{\theta}) - E_{i-1}[\bar{\mathbf{B}}_{t_i|t_{i-1}}] + \boldsymbol{\Sigma}_{t_i}\,, \tag{32}$$

$$\mathbf{K}_{t_i} = \mathbf{P}_{t_i|t_{i-1}}\mathbf{H}^T(\hat{\mathbf{X}}_{t_i|t_{i-1}}; \boldsymbol{\theta})\mathbf{A}_{t_i}^{-1}\,, \tag{33}$$

$$\hat{\mathbf{X}}_{t_i|t_i} = \hat{\mathbf{X}}_{t_i|t_{i-1}} + \mathbf{K}_{t_i}\left\{\mathbf{Y}_{t_i} - \mathbf{h}(\hat{\mathbf{X}}_{t_i|t_{i-1}}; \boldsymbol{\theta}) - E_{i-1}[\tilde{\mathbf{B}}_{t_i|t_{i-1}}]\right\}\,, \tag{34}$$

$$\mathbf{P}_{t_i|t_i} = \mathbf{P}_{t_i|t_{i-1}} - \mathbf{K}_{t_i}\mathbf{H}(\hat{\mathbf{X}}_{t|t_{i-1}}; \boldsymbol{\theta})\mathbf{P}_{t_i|t_{i-1}}\,, \tag{35}$$

where $\mathbf{H}(\hat{\mathbf{X}}_{t|t_{i-1}}; \boldsymbol{\theta})$ is defined in (28) and $E_{i-1}[\tilde{\mathbf{B}}_{t_i|t_{i-1}}]$ is given by (29), and the bias-correction term $E_{i-1}[\bar{\mathbf{B}}_{t_i|t_{i-1}}]$ is a $m \times m$-matrix with the klth component given by

$$E^{kl}_{i-1}[\bar{\mathbf{B}}_{t_i|t_{i-1}}] = \frac{1}{2}\mathrm{tr}\left\{\frac{\partial^2 \mathbf{h}^k(\mathbf{x}; \boldsymbol{\theta})}{\partial \mathbf{x}^2}\mathbf{P}_{t_i|t_{i-1}}\frac{\partial^2 \mathbf{h}^l(\mathbf{x}; \boldsymbol{\theta})}{\partial \mathbf{x}^2}\mathbf{P}_{t_i|t_{i-1}}\right\}\Bigg|_{\mathbf{x}=\hat{\mathbf{X}}_{t_i|t_{i-1}}}. \tag{36}$$

Remark 3.6. Notice again that the bias-correction term (36) drops out if the observation Eq. (3) is linear in the state variables.

3.3. *Maximum likelihood estimation*

In this section a maximum likelihood method for estimation of the parameters in the continuous-discrete state space model (4)–(5) based on an assumption of Gaussianity for the one-step prediction errors given by the expressions in the curly brackets in either (16), (26) or (34) is presented.[e] This assumption may be tested using standard statistical tests for Gaussian white noise residuals. If these tests are rejected at all reasonable levels of significance, then the method can be considered as a prediction error method or a quasi-maximum likelihood method [48, 51].

Assume that $\mathbf{X}_t$, $t \geq t_0$, solves (4) and that the initial condition $\mathbf{X}_0$ is Gaussian, i.e. $\mathbf{X}_0 \in \mathcal{N}_n(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$. The likelihood function $\tilde{L}(\boldsymbol{\theta}; \mathcal{X}_{t_N})$, where $\mathcal{X}_{t_N} =$

[e]The idea follows the prediction error decomposition method proposed by [65].

$(\mathbf{X}_{t_N}, \ldots, \mathbf{X}_{t_0})$ denotes the sample path of state variable at the sampling instants t_i, $i = 1, \ldots, N$, is

$$\tilde{L}(\boldsymbol{\theta}; \mathcal{X}_{t_N}) = \left[\prod_{i=1}^{N} p_{\mathbf{x}}(\mathbf{x}_{t_i}, t_i | \mathbf{x}_{t_{i-1}}, t_{i-1}, \boldsymbol{\theta})\right] p(\mathbf{x}_{t_0} | \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \tag{37}$$

where the cpdf $p_{\mathbf{x}}(\mathbf{x}_{t_i}, t_i | \mathbf{x}_{t_{i-1}}, t_{i-1}, \boldsymbol{\theta})$ solves the multivariate extension of (6), i.e.

$$\begin{aligned} &\frac{\partial p_{\mathbf{x}}(\mathbf{x}_t, t | \mathbf{x}_{t_{i-1}}, t_{i-1}, \boldsymbol{\theta})}{\partial t} \\ &\quad = -\sum_{i=1}^{N} \frac{\partial [p_{\mathbf{x}}(\mathbf{x}_t, t | \mathbf{x}_{t_{i-1}}, t_{i-1}, \boldsymbol{\theta}) f^i(\mathbf{x}_t; \boldsymbol{\theta})]}{\partial \mathbf{x}_t} \\ &\qquad + \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\partial^2 [p_{\mathbf{x}}(\mathbf{x}_t, t | \mathbf{x}_{t_{i-1}}, t_{i-1}, \boldsymbol{\theta}) (\mathbf{G}(\mathbf{x}_t; \boldsymbol{\theta}) \mathbf{Q}_t \mathbf{G}^T(\mathbf{x}_t; \boldsymbol{\theta}))^{ij}]}{\partial \mathbf{x}_t^2} \end{aligned} \tag{38}$$

and that $p(\mathbf{x}_{t_0} | \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ denote the multivariate Gaussian pdf with mean $\boldsymbol{\mu}_0$ and covariance $\boldsymbol{\Sigma}_0$. An explicit solution to (38) is not available (as argued in Sec. 3.1), and furthermore only a function of $\mathbf{X}_t$ encumbered with noise is observed as described by (5). However, a similar construction can be made for the observations $\mathbf{Y}_{t_i}$, $i = 1, \ldots, N$, i.e.

$$\bar{L}(\boldsymbol{\theta}; \mathcal{F}_{t_N}) = \left(\prod_{i=1}^{N} p_{\mathbf{y}}(\mathbf{y}_{t_i} | \mathcal{F}_{i-1}, \boldsymbol{\theta})\right) (p(\mathbf{y}_{t_0} | \boldsymbol{\mu}_0, \boldsymbol{\varepsilon}_0, \boldsymbol{\Sigma}_{t_0}) \ , \tag{39}$$

where it is necessary, due to the incomplete observation of the state vector, to condition on all previous observations and not only the previous observation as in (37). Most frequently, however, the conditional likelihood function

$$L(\boldsymbol{\theta}; \mathcal{F}_{t_N}) = \left(\prod_{i=1}^{N} p_{\mathbf{y}}(\mathbf{y}_{t_i} | \mathbf{y}_{t_{i-1}}, \boldsymbol{\theta})\right) (p(\mathbf{y}_{t_0} | \boldsymbol{\mu}_0, \boldsymbol{\varepsilon}_0, \boldsymbol{\Sigma}_{t_0}) \tag{40}$$

is considered.

The cpdf $p_{\mathbf{y}}$ is rarely available (as argued in Sec. 3.1). However, progress may be made by considering the density of the one-step prediction errors

$$\boldsymbol{\varepsilon}_{t_i}(\boldsymbol{\theta}) \equiv \mathbf{Y}_{t_i} - \mathbf{h}(\hat{\mathbf{X}}_{t_i|t_{i-1}}; \boldsymbol{\theta}) \,. \tag{41}$$

Assuming Gaussianity of $\boldsymbol{\varepsilon}_{t_i}(\boldsymbol{\theta})$, the likelihood function is completely characterized by the conditional first and second order central moments

$$\hat{\mathbf{Y}}_{t_i|t_{i-1}} = E[\mathbf{Y}_{t_i} | \mathcal{F}_{i-1}, \boldsymbol{\theta}] = \mathbf{h}(\hat{\mathbf{X}}_{t_i|t_{i-1}}; \boldsymbol{\theta}) \tag{42}$$

$$\mathbf{R}_{t_i|t_{i-1}} = V[\mathbf{Y}_{t_i} | \mathcal{F}_{i-1}, \boldsymbol{\theta}] \tag{43}$$

i.e. the conditional likelihood function is given by

$$L(\boldsymbol{\theta}; \mathcal{F}_{t_N}) = \prod_{i=1}^{N} \left((2\pi)^m \det(\mathbf{R}_{t_i|t_{i-1}})\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \boldsymbol{\varepsilon}_{t_i}^T(\boldsymbol{\theta}) \mathbf{R}_{t_i|t_{i-1}}^{-1} \boldsymbol{\varepsilon}_{t_i}(\boldsymbol{\theta})\right) \tag{44}$$

and the log-likelihood function is

$$\begin{aligned}\ln L(\boldsymbol{\theta};\mathcal{F}_{t_N}) &= -\frac{1}{2}\sum_{i=1}^{N}(\ln \det(\mathbf{R}_{t_i|t_{i-1}}) \\ &\quad + \boldsymbol{\varepsilon}_{t_i}^T(\boldsymbol{\theta})\mathbf{R}_{t_i|t_{i-1}}^{-1}\boldsymbol{\varepsilon}_{t_i}(\boldsymbol{\theta})) + \text{constant}\,. \end{aligned} \tag{45}$$

An estimate of the uncertainty of the parameters is obtained using the fact that the ML-estimator is asymptotically Gaussian distributed with mean $\mathbf{V}$ and covariance $\boldsymbol{\Sigma}$ given by the lower bound of the Cramer–Rao inequality, i.e. $\mathbf{V} = \mathbf{H}^{-1}$, where the elements of the Hessian matrix are given by

$$h^{ij} = -E\left\{\frac{\partial^2 \ln L(\boldsymbol{\theta};\mathcal{F}_{t_N})}{\partial\theta^i\partial\theta^j}\right\} \tag{46}$$

such that the covariance matrix of the estimated parameter vector is readily available.

Remark 3.7. The one-step prediction errors given in the curly brackets of (16) are structurally in accordance with the innovations suggested by the observation Eq. (5) for $h(X_{t_i};\boldsymbol{\theta}) = X_{t_i}$, so the interpretation of (16) is clear. However, in the general multivariate case, the expressions in the curly brackets in (26) and (34) contain the additional bias-correction terms given by (29) and (36), respectively. This is due to the approximative nature of the second order filters, and it suggests that the one-step prediction errors (residuals) obtained from (41) may be confounded with some of the deficiencies of the filter in the general case.[f] Notice, however, that the bias-correction terms (29) and (36) drop out for the stochastic volatility models considered here, because the observation Eq. (5) is linear.

4. Monte Carlo Studies

In this section the estimation method proposed in Sec. 3.3 will be evaluated for two stochastic volatility models belonging to the class (1)–(3), see also Table 1. Explicit solutions to this class of stochastic volatility models either in terms of a stochastic process or the cpdf given by (38) do not exist for non-trivial choices of $\psi(\sigma_t)$, $a(\sigma_t)$ or $b(\sigma_t)$. An efficient and widely applicable approach to solving SDEs is to simulate sample paths of a time discrete approximation to the continuous-time model [49]. Without loss of generality, assume that $t_0 = 0$ and divide the time interval $[0, t_N]$ into M ($M \gg N$) small time intervals $0 = \tau_0 < \cdots < \tau_n < \cdots < \tau_M = t_N$, where the time intervals are assumed to be equally spaced $\Delta = \tau_{n+1} - \tau_n = t_N/M$. Approximate the Wiener processes W_t^k, $k = 1, 2$, by the increments

$$\Delta W_n^k = W_{\tau_{n+1}}^k - W_{\tau_n}^k$$

[f] See [72] for a discussion of this in discrete-time structural models.

with mean $E[\Delta W_n^k] = 0$ and variance $E[(\Delta W_n^k)^2] = \Delta$. The Euler approximation of (1)–(2) is given by the bivariate stochastic difference equation

$$S_{n+1} = S_n + \alpha S_n \Delta + \sigma_n S_n [\rho \Delta \tilde{W}_n^1 + \sqrt{1-\rho^2} \Delta \tilde{W}_n^2] \,, \tag{47}$$

$$\sigma_{n+1} = \sigma_n + a(\sigma_n)\Delta + b(\sigma_n)\Delta \tilde{W}_n^1 \,, \tag{48}$$

where it has been assumed that $\psi(\sigma_t) = \sigma_t$, and the correlated Wiener processes ΔW_n^1 and ΔW_n^2 have been replaced by the uncorrelated Wiener processes $\Delta \tilde{W}_n^1$ and $\Delta \tilde{W}_n^2$ with mean $E[\Delta \tilde{W}_n^k] = 0$ and variance $V[\Delta \tilde{W}_n^k] = \Delta$ for $k = 1, 2$.

4.1. *The Black–Scholes–Courtadon model — Full information*

Consider the following model as an example of the general class (1)–(3), where the stochastic volatility is modelled as a mean-reverting process proposed by [14], i.e.

$$\begin{pmatrix} dS_t \\ d\sigma_t \end{pmatrix} = \begin{pmatrix} \alpha S_t \\ \kappa(\beta - \sigma_t) \end{pmatrix} dt + \begin{pmatrix} \sigma_t S_t & 0 \\ 0 & \xi \sigma_t \end{pmatrix} \begin{pmatrix} dW_t^1 \\ dW_t^2 \end{pmatrix} , \tag{49}$$

where (W_t^1, W_t^2) is a bivariate Wiener processes with correlation coefficient ρ.

Instead of operating with both a $\mathbf{G}$ matrix and a $\mathbf{Q}$ matrix, the model may be written on a form where $\mathbf{Q}$ is the identity matrix, i.e.

$$\begin{pmatrix} dS_t \\ d\sigma_t \end{pmatrix} = \begin{pmatrix} \alpha S_t \\ \kappa(\beta - \sigma_t) \end{pmatrix} dt + \begin{pmatrix} \sigma_t S_t & 0 \\ \rho \xi \sigma_t & \sqrt{1-\rho^2} \xi \sigma_t \end{pmatrix} \begin{pmatrix} d\tilde{W}_t^1 \\ d\tilde{W}_t^2 \end{pmatrix} , \tag{50}$$

where $(\tilde{W}_t^1, \tilde{W}_t^2)$ is a bivariate Wiener process with uncorrelated elements. In this case of full information the observation equation is

$$\mathbf{Y}_{t_k} = \begin{pmatrix} S_{t_k} \\ \sigma_{t_k} \end{pmatrix} + \begin{pmatrix} e_{t_k} \\ 0 \end{pmatrix} , \; e_{t_k} \in \mathcal{N}(0, \Sigma) \,. \tag{51}$$

Let $\mathbf{X}_t = (X_t^1, X_t^2)^T = (S_t, \sigma_t)^T$ denote the state vector. The time propagation equations of the modified Gaussian second order filter are

$$\frac{d\hat{\mathbf{X}}_{t|t_{i-1}}}{dt} = \begin{pmatrix} \alpha \hat{X}^1_{t|t_{i-1}} \\ \kappa(\beta - \hat{X}^2_{t|t_{i-1}}) \end{pmatrix} , \tag{52}$$

$$\frac{d\mathbf{P}_{t|t_{i-1}}}{dt} = \begin{pmatrix} 2\alpha P^{11}_{t|t_{i-1}} & P^{12}_{t|t_{i-1}}(\alpha - \kappa) \\ P^{21}_{t|t_{i-1}}(\alpha - \kappa) & -2\kappa P^{22}_{t|t_{i-1}} \end{pmatrix} + \begin{pmatrix} \tilde{P}^{11}_{t|t_{i-1}} & \tilde{P}^{12}_{t|t_{i-1}} \\ \tilde{P}^{21}_{t|t_{i-1}} & \tilde{P}^{22}_{t|t_{i-1}} \end{pmatrix} , \tag{53}$$

where $\tilde{P}^{jk}_{t|t_{i-1}}$, with k, $j = 1, 2$ is defined as

$$\tilde{P}^{11}_{t|t_{i-1}} = (\hat{X}^1_{t|t_{i-1}} \hat{X}^2_{t|t_{i-1}})^2 + P^{11}_{t|t_{i-1}} (\hat{X}^2_{t|t_{i-1}})^2 + 4 \hat{X}^1_{t|t_{i-1}} \hat{X}^2_{t|t_{i-1}} P^{12}_{t|t_{i-1}} + (P^{12}_{t|t_{i-1}})^2$$
$$+ P^{22}_{t|t_{i-1}} (\hat{X}^1_{t|t_{i-1}})^2 + \frac{1}{2}((P^{12}_{t|t_{i-1}})^2 + (P^{21}_{t|t_{i-1}})^2 + 2 P^{11}_{t|t_{i-1}} P^{22}_{t|t_{i-1}}) \,,$$

$$\begin{aligned}\tilde{P}^{21}_{t|t_{i-1}} &= \tilde{P}^{12}_{t|t_{i-1}} \\ &= \xi\rho\left\{\hat{X}^1_{t|t_{i-1}}(\hat{X}^2_{t|t_{i-1}})^2 + P^{21}_{t|t_{i-1}}\hat{X}^2_{t|t_{i-1}} + P^{22}_{t|t_{i-1}}\hat{X}^1_{t|t_{i-1}} + \hat{X}^2_{t|t_{i-1}}P^{12}_{t|t_{i-1}}\right\}, \\ \tilde{P}^{22}_{t|t_{i-1}} &= \xi^2((\hat{X}^2_{t|t_{i-1}})^2 + P^{22}_{t|t_{i-1}}).\end{aligned}$$

The update equations are given by

$$\begin{aligned}\hat{\mathbf{X}}_{t_i|t_i} &= \begin{pmatrix}\hat{X}^1_{t_i|t_{i-1}} \\ \hat{X}^2_{t_i|t_{i-1}}\end{pmatrix} + \mathbf{P}_{t_i|t_{i-1}}\begin{pmatrix}P^{11}_{t_i|t_{i-1}} + \Sigma & P^{12}_{t_i|t_{i-1}} \\ P^{21}_{t_i|t_{i-1}} & P^{22}_{t_i|t_{i-1}}\end{pmatrix}^{-1} \\ &\quad\times\left\{\begin{pmatrix}Y^1_{t_i} \\ Y^2_{t_i}\end{pmatrix} - \begin{pmatrix}\hat{X}^1_{t_i|t_{i-1}} \\ \hat{X}^2_{t_i|t_{i-1}}\end{pmatrix}\right\}, \\ \mathbf{P}_{t_i|t_i} &= \mathbf{P}_{t_i|t_i} - \mathbf{P}_{t_i|t_{i-1}}\begin{pmatrix}P^{11}_{t_i|t_{i-1}} + \Sigma & P^{12}_{t_i|t_{i-1}} \\ P^{21}_{t_i|t_{i-1}} & P^{22}_{t_i|t_{i-1}}\end{pmatrix}^{-1}\mathbf{P}_{t_i|t_{i-1}}.\end{aligned}$$

The model (50) has been simulated using the Euler scheme (47)–(48) with $\Delta = 10^{-3}$ to obtain 1 million observations. Every 1000th observation has been sampled with the sampling time $t_i - t_{i-1} = 1$ such that $N = 1000$ observations are obtained. The parameter values

$$\boldsymbol{\theta} = (\alpha, \kappa, \beta, \xi, \Sigma, \rho)^T = (0.00198, 0.0071, 0.01, 0.016, 0.0195, 0.5)^T$$

have been chosen to mimic the properties of real data.

Table 2. Estimation results for the correlated Black–Scholes–Courtadon model (50)–(51), where $\bar{\theta}$ and s_θ denote respectively the mean and standard deviation of the parameter estimates obtained from the 10 independent simulations, and $|t| = (\theta_j - \bar{\theta}_j)/(s_{\theta_j}/\sqrt{10})$ is a t-test statistic under the null hypothesis that the estimated parameters are unbiased.

Parameter	α	κ	β	ξ	Σ	ρ
Simulation no.	0.0019800	0.0071000	0.0100000	0.0160000	0.0195000	0.5000000
1	0.0025093	0.0090798	0.0098299	0.0159729	0.0206965	0.5304104
2	0.0022716	0.0128750	0.0109020	0.0160406	0.0175810	0.5071539
3	0.0017152	0.0103744	0.0092906	0.0164556	0.0194734	0.4753382
4	0.0017658	0.0073760	0.0103590	0.0168831	0.0175455	0.5036983
5	0.0017914	0.0204004	0.0102107	0.0161234	0.0212303	0.5144773
6	0.0017008	0.0067208	0.0089242	0.0163276	0.0194376	0.4212158
7	0.0020048	0.0117637	0.0097847	0.0159738	0.0194874	0.5448961
8	0.0022034	0.0107719	0.0109357	0.0162803	0.0209394	0.5331270
9	0.0019004	0.0035937	0.0102286	0.0160098	0.0205281	0.5870837
10	0.0021469	0.0108092	0.0119591	0.0161880	0.0175823	0.5153257
mean ($\bar{\theta}$)	0.0020010	0.0103765	0.0102424	0.0162255	0.0194502	0.5132726
std. dev. (s_θ)	2.7387e-4	0.0044602	8.7563e-4	2.8255e-4	0.0014420	0.0436963
$\mid t \mid$-stat	0.2421484	2.3230069	0.8756128	2.5239735	0.1092857	0.9605341

Using the filter described above and the maximum likelihood method described in Sec. 3.3 the estimation results in Table 2 for each of 10 independent simulations are obtained. For each parameter, a t-test is provided to verify if the estimation results are unbiased. It is seen that the t-tests show critical test statistics for both κ and ξ. The problem with the estimation of κ originates from the chosen parametrization of the drift, which makes it difficult for the method to make a proper separation of both κ and β. The test statistics for ξ is significant due to the accuracy of the estimate of ξ thus producing a very small standard deviation.

To verify if the residuals have Gaussian white noise properties the model verification tests in Appendix A are carried out on the standardized residuals for each observed process individually.[g] All these tests are accepted on a 5% level indicating Gaussian white noise residuals.

4.2. *The Black–Scholes–Courtadon model — Partial information*

Consider again the model (50) and replace (51) by

$$Y_{t_i} = S_{t_i} + e_{t_i} \,. \tag{54}$$

The time propagation equations for the modified Gaussian second order filter are given by (52)–(53). The updating equations are given by

$$\hat{\mathbf{X}}_{t_i|t_i} = \begin{pmatrix} \hat{X}^1_{t_i|t_{i-1}} \\ \hat{X}^2_{t_i|t_{i-1}} \end{pmatrix} + \frac{1}{P^{11}_{t_i|t_{i-1}} + \Sigma} \begin{pmatrix} P^{11}_{t_i|t_{i-1}} \\ P^{21}_{t_i|t_{i-1}} \end{pmatrix} \left\{ Y_{t_i} - \hat{X}_{t_i|t_{i-1}} \right\} \tag{55}$$

$$\begin{aligned} \mathbf{P}_{t_i|t_i} &= \begin{pmatrix} P^{11}_{t_i|t_{i-1}} & P^{12}_{t_i|t_{i-1}} \\ P^{21}_{t_i|t_{i-1}} & P^{22}_{t_i|t_{i-1}} \end{pmatrix} \\ &\quad - \frac{1}{P^{11}_{t_i|t_{i-1}} + \Sigma} \begin{pmatrix} (P^{11}_{t_i|t_{i-1}})^2 & P^{12}_{t_i|t_{i-1}} P^{11}_{t_i|t_{i-1}} \\ P^{21}_{t_i|t_{i-1}} P^{11}_{t_i|t_{i-1}} & P^{12}_{t_i|t_{i-1}} P^{21}_{t_i|t_{i-1}} \end{pmatrix} . \end{aligned} \tag{56}$$

4.2.1. *Special case $\kappa = 1$*

The model (50)+(54) has been used for simulation of 10 independent realizations using the Euler scheme (47)–(48) with $\Delta = 10^{-3}$ and the sampling time $t_i - t_{i-1} = 10^{-1}$ such that $N = 1000$ observations are obtained for each time series. The parameter values $\boldsymbol{\theta} = (\alpha, \beta, \xi, \Sigma, \rho)^T = (0.035, 0.13, 0.5, 0.12, -0.5)^T$ have been used. The choice of a positive α implies that the model is non-stationary, but consistent with observed stock prices.

Using the filter described above the estimation results in Table 3 are obtained with the parameter restriction $\kappa = 1$. For each parameter, a t-test is provided to verify if the estimation results are unbiased. All the t-tests are accepted on

[g]The residuals of the stock price process have been normalized by the factor $(P_{t_i|t_{i-1}} + \Sigma)^{-1/2}$. A similar result holds for the residuals of the stochastic volatility process.

Table 3. Estimation results for the model (50) + (54) with the restriction $\kappa = 1$ imposed, where $\bar{\theta}$ and s_θ denote respectively the mean and standard deviation of the parameter estimates obtained from the 10 independent simulations, and $|t| = (\theta_j - \bar{\theta}_j)/(s_{\theta_j}/\sqrt{10})$ is a t-test statistic under the null hypothesis that the estimated parameters are unbiased.

Parameter	α	β	ξ	Σ	ρ
Simulation no.	0.0350000	0.1300000	0.5000000	0.1200000	−0.5000000
1	0.0293474	0.1310356	0.6401490	0.0880917	−0.3535361
2	0.0391841	0.1232693	0.6097279	0.1145222	−0.4564107
3	0.0292487	0.1385571	0.3364275	0.1141113	−0.6220152
4	0.0394285	0.1333746	0.3699841	0.1312704	−0.7702001
5	0.0221745	0.1173381	0.6941756	0.1349321	−0.4016965
6	0.0293484	0.1310349	0.6401513	0.0880908	−0.3535343
7	0.0391843	0.1232701	0.6097097	0.1145213	−0.4564247
8	0.0225359	0.1173069	0.6970154	0.1246237	−0.4882357
9	0.0546890	0.1305796	0.1552850	0.1496985	−0.8602473
10	0.0264529	0.1344481	0.6414826	0.1222545	−0.4648418
mean ($\bar{\theta}$)	0.0331594	0.1280214	0.5394108	0.1182116	−0.5227142
std. dev. (s_θ)	9.9906e-3	7.3012e-3	0.1846479	1.9315e-2	0.1732870
$\mid t \mid$-stat	0.5825891	0.8569505	0.6749489	0.2927969	0.4145073

Table 4. Test statistics for the standardized residuals from (50) + (54). JB is the Jarque–Bera test statistic for normality (A.1) and the critical value is $\chi^2_{95\%}(2) = 5.991$. BL is the Box–Ljung test statistic for autocorrelation in the residuals (A.2) and the critical value is $\chi^2_{95\%}(15) = 24.996$. BL2 is the Box–Ljung test statistic autocorrelation in the squared residuals and the critical value is $\chi^2_{95\%}(20) = 34.410$. NL is a test statistic (A.3) for heteroscedasticity, and the critical set is $C = \{H < 0.81 \wedge H > 1.24\}$ on a 5% level.

Simulation	1	2	3	4	5	6	7	8	9	10
JB	45.99	25.32	29.95	11.57	4.11	46.00	25.32	15.39	4.42	53.76
BL	10.91	27.23	18.45	14.48	21.47	10.91	27.23	18.20	17.53	13.96
BL2	60.98	95.60	40.67	84.22	33.84	60.98	95.60	15.59	64.42	56.58
NL	1.31	0.96	1.27	1.06	0.96	1.31	0.96	1.10	0.96	1.31

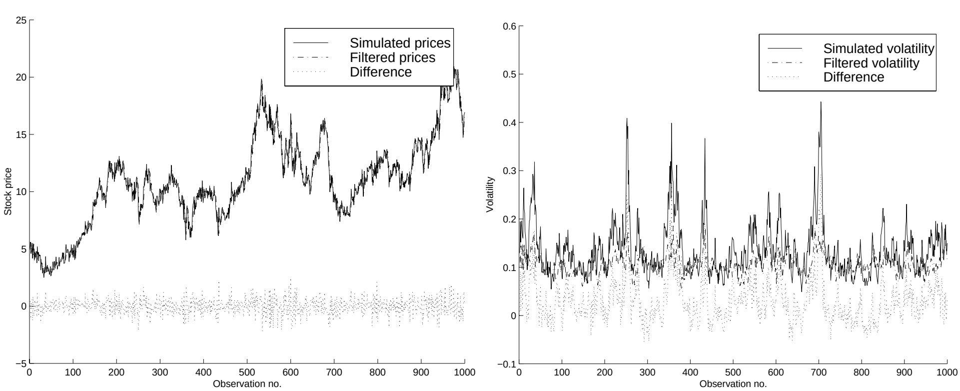
a 40% level giving a strong indication of unbiased estimates. It is seen that the method is able also to provide reasonable estimates of the parameters β and ξ in the unobserved stochastic volatility process (54). This also applies for the correlation coefficient ρ.

In order to verify the assumption of normality of the one-step prediction errors, the tests in Appendix A have been applied to the normalized residuals, see Table 4. The Jarque–Bera statistic rejects the assumption of Gaussianity in 8 out of the 10 simulations. The Box–Ljung test for no autocorrelation in the residuals is only rejected for simulation No. 2 and 7. The Box–Ljung test for no autocorrelation in the squared residuals is rejected for all 10 simulations, except simulation 8. The null hypothesis for no heteroscedasticity is accepted for 6 out of 10 simulations. The exceptions are simulations No. 1, 3, 6 and 10. These test statistics are not consistent with the excellent parameter estimates reported in Table 3, so the rejection of the assumption of Gaussianity must originate from the approximative nature of the applied second order filter. This conclusion seems to confirm the shortcomings of the approximative second order filters reported in [72].

4.2.2. *General case*

Including the κ parameter in the estimation yields the parameter estimates reported in Table 5. For all the parameters the $|t|$-test statistics are larger than in Table 3, but only the estimates of κ and Σ differ significantly (on a 5% level) from their simulated values. In particular, the upwards biased estimate of κ may be explained by the smoothing effect of the filter, which gives rise to the same effect as a high speed-of-adjustment parameter in the mean-reverting drift.

In Fig. 2 the sample paths of the S_t and σ_t processes have been plotted along with their filtered equivalents. It is readily seen that a good estimate of the S_t-process is provided. The estimate of the stochastic volatility process σ_t exhibits behavior similar to the simulated process, but the variations are more limited in magnitude such that the difference is noticeable. This illustrates that the filtering method performs a smoothing of the unobserved stochastic volatility process, which follows from the fact that the optimal predictor is the conditional mean. In spite of this smoothing effect, the parameter estimates suggests that the proposed method captures the dynamics of the simulated time series. The results from Sec. 4.2.1 suggests that the biasedness of the parameter κ is caused by the smoothing effect of the filter, it is not the parameter κ that causes the smoothing effect.



(a) (b)

Fig. 2. Plot of the simulated and filtered stock price process S_{t_i} (a), and the simulated and filtered stochastic volatility σ_{t_i} (b) from the Black–Scholes–Courtadon model (50) + (54).

Table 5. Estimation results for the model (50) + (54), where $\bar{\theta}$ and s_θ denote respectively the mean and standard deviation of the parameter estimates obtained from the 10 independent simulations, and $|t| = (\theta_j - \bar{\theta_j})/(s_{\theta_j}/\sqrt{10})$ is a t-test statistic under the null hypothesis that the estimated parameters are unbiased.

Parameter	α	κ	β	ξ	Σ	ρ
Simulation no.	0.0350000	1.0000000	0.1300000	0.5000000	0.1200000	−0.5000000
1	0.0678985	2.3840620	0.1228596	0.3985797	0.1246835	−0.6301745
2	0.0334874	2.5267987	0.1261216	1.0834183	0.1473596	−0.4257387
3	0.0680778	3.4516496	0.1221268	0.7571279	0.1319948	−0.4423765
4	0.0480507	1.9770827	0.1332794	0.3431033	0.1411092	−0.7409307
5	0.0222548	1.2106302	0.1053715	0.9708950	0.1346723	−0.3175558
6	0.0656793	1.9721291	0.1023120	1.1977318	0.1243178	−0.2089123
7	0.0087756	2.5593195	0.1181904	0.5891934	0.1311660	−0.9665081
8	0.0268085	2.6095114	0.1273615	0.1319957	0.1230215	−0.9819453
9	0.0339221	1.9832019	0.1377474	0.5915995	0.1476529	−0.6779232
10	0.0499857	2.9307014	0.1355141	0.1813567	0.1100203	−0.6962402
mean $\bar{\theta}$	0.0424940	2.3605087	0.1230893	0.6245001	0.1315998	−0.6088305
std. dev. s_θ	0.0207409	0.6153480	0.0118825	0.3724200	0.0117682	0.2589451
$\mid t \mid$-stat	1.1425843	6.9916633	1.8391317	1.0571505	3.1179280	1.3290551

Table 6. Test statistics for the S_{t_i}-process of model (50) + (54). JB is the Jarque–Bera test statistic for normality (A.1) and the critical value is $\chi^2_{95\%}(2) = 5.991$. BL is the Box–Ljung test statistic for autocorrelation in the residuals (A.2) and the critical value is $\chi^2_{95\%}(14) = 23.685$. BL2 is the Box–Ljung test statistic autocorrelation in the squared residuals and the critical value is $\chi^2_{95\%}(20) = 34.410$. NL is a test statistic (A.3) for heteroscedasticity, and the critical set is $C = \{H < 0.81 \wedge H > 1.24\}$ on a 5% level.

Simulation	1	2	3	4	5	6	7	8	9	10
JB	20.98	145.75	9.72	78.48	4.18	10.93	3.15	3.18	152.14	50.76
BL	15.17	34.76	21.05	19.10	21.50	19.85	17.16	18.75	34.93	23.59
BL2	74.33	99.10	70.64	80.72	33.64	70.45	42.44	29.41	100.77	33.61
NL	0.89	0.90	0.94	0.82	0.96	0.97	1.19	0.83	0.90	1.23

The model validation statistics listed in Table 6 lead to rejection of the assumption of Gaussianity (except for simulations No. 5, 7 and 8). The Box–Ljung test is accepted for all simulations except 2 and 9. The test for no heteroscedasticity is accepted for all simulations, whereas the test for no nonlinearities in the model residuals is clearly rejected.[h] Thus the parameter estimates (and their covariances)

[h]The distinction between nonlinearities and heteroscedasticity is not very clear in the literature, so the tests should be interpreted with caution.

and the tests for Gaussian white noise residuals all indicate that a full estimation of the model (50) + (54) is a difficult problem.

4.3. *The Cox–Generalized CIR model — Full information*

Now assume that the stock price is described by a Cox [15] model, where σ is described by a generalized CIR model [11], i.e.

$$\begin{pmatrix} dS_t \\ d\sigma_t \end{pmatrix} = \begin{pmatrix} \alpha S_t \\ \kappa(\beta - \sigma_t) \end{pmatrix} dt + \begin{pmatrix} \xi_1 \sigma_t S_t^{\gamma_1} & 0 \\ 0 & \xi_2 \sigma_t^{\gamma_2} \end{pmatrix} \begin{pmatrix} dW_t^1 \\ dW_t^2 \end{pmatrix} \tag{57}$$

$$\mathbf{Y}_{t_k} = \begin{pmatrix} S_{t_k} \\ \sigma_{t_k} \end{pmatrix} + \begin{pmatrix} e_{t_k} \\ 0 \end{pmatrix}, \; e_{t_k} \in N(0, \Sigma)\,, \tag{58}$$

where W_t^1 and W_t^2 are standard correlated Wiener processes with correlation coefficient ρ.

Let $\mathbf{X}_t = (X_t^1, X_t^2)^T = (S_t, \sigma_t)^T$ denote the state vector. The time propagation equations for the modified Gaussian second order filter are

$$\frac{d\hat{\mathbf{X}}_{t|t_{i-1}}}{dt} = \begin{pmatrix} \alpha \hat{X}^1_{t|t_{i-1}} \\ \kappa(\beta - \hat{X}^2_{t|t_{i-1}}) \end{pmatrix}, \tag{59}$$

$$\frac{d\mathbf{P}_{t|t_{i-1}}}{dt} = \begin{pmatrix} 2\alpha P^{11}_{t|t_{i-1}} & P^{12}_{t|t_{i-1}}(\alpha - \kappa) \\ P^{21}_{t|t_{i-1}}(\alpha - \kappa) & -2\kappa P^{22}_{t|t_{i-1}} \end{pmatrix} + \begin{pmatrix} \tilde{P}^{11}_{t|t_{i-1}} & \tilde{P}^{12}_{t|t_{i-1}} \\ \tilde{P}^{21}_{t|t_{i-1}} & \tilde{P}^{22}_{t|t_{i-1}} \end{pmatrix} \tag{60}$$

where the expressions for $\tilde{P}^{jk}_{t|t_{i-1}}$, with $k, j = 1, 2$ have been left out for brevity.

The update equations are given by (55)–(56). Data has been simulated as described in Sec. 4.3. The simulated parameters and the estimated parameters are listed in Table 7, where it is noticed that the hypothesis of unbiased estimates is rejected only for α.

The usual tests for Gaussian white noise residuals are accepted for the S_t-process. The results for the σ_t-process are reported in Table 8. The test statistics for both processes are in concordance with the parameter estimates reported in Table 7, so the rejection of Gaussianity in the residuals of the σ_t-process most likely originates from the approximative nature of the second order filter.

4.4. *The Cox–Generalized CIR model — Partial information*

Consider the model (57) and replace (58) by

$$Y_{t_k} = S_{t_k} + e_{t_k}\,, \quad e_{t_k} \in N(0, \sigma_e^2)\,. \tag{61}$$

The time propagation equations are given by (59)–(60) and the update equations are again (55)–(56).

Table 7. Estimation results for the correlated Cox–Generalized CIR model (57)–(58).

Parameter	α	ξ_1	γ_1	κ	β	ξ_2	γ_2	σ_e^2	ρ
Simulation no.	0.0300	0.7000	1.5000	1.3000	0.1000	0.8000	1.5000	0.1200	−0.5000
1	0.0494	0.7078	1.4917	1.2507	0.1015	0.3798	1.1633	0.1143	−0.5392
2	0.0383	0.6312	1.5257	1.1916	0.1005	0.6253	1.3984	0.1321	−0.4822
3	0.0494	0.7079	1.4917	1.2507	0.1015	0.3795	1.1630	0.1143	−0.5392
4	0.0402	1.1203	1.3077	1.5115	0.1002	1.0448	1.6082	0.1145	−0.4885
5	0.0360	0.5794	1.5732	1.1575	0.1006	1.2141	1.6874	0.1226	−0.5328
6	0.0530	0.4034	1.7592	1.2837	0.0978	0.8965	1.5406	0.1175	−0.5477
7	0.0252	0.8533	1.3941	1.5926	0.0965	0.5905	1.3848	0.1143	−0.4305
8	0.0505	0.6893	1.4980	1.3414	0.0989	1.3188	1.7183	0.1159	−0.5231
9	0.0276	0.3823	1.8642	1.4752	0.1008	1.2938	1.6820	0.1271	−0.5278
10	0.0184	0.7037	1.4499	1.4789	0.0985	0.6633	1.4137	0.1255	−0.5370
mean ($\bar{\theta}$)	0.0388	0.6778	1.5355	1.3534	0.0997	0.8406	1.4760	0.1198	−0.5148
std. dev. (s_θ)	0.0120	0.2118	0.1647	0.1503	0.0017	0.3627	0.2059	0.0065	0.0368
$\mid t \mid$-stat	2.3132	0.3307	0.6823	1.1227	0.6330	0.3542	0.3694	0.0968	1.2715

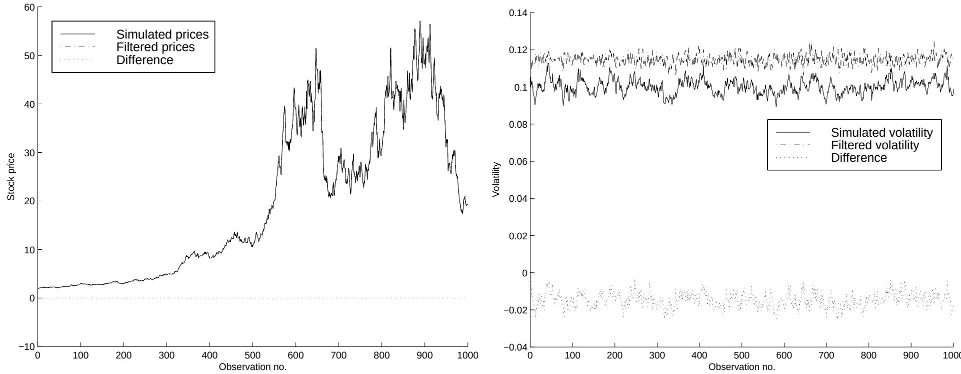
Table 8. Test statistics for the standardized residuals of the σ_t-process of the correlated Cox–Generalized CIR model (57)–(58). JB is the Jarque–Bera test statistic for normality (A.1) and the critical value is $\chi^2_{95\%}(2) = 5.991$. BL is the Box–Ljung test statistic for autocorrelation in the residuals (A.2) and the critical value is $\chi^2_{95\%}(11) = 19.670$. BL2 is the Box–Ljung test statistic autocorrelation in the squared residuals and the critical value is $\chi^2_{95\%}(20) = 34.410$. NL is a test statistic (A.3) for heteroscedasticity, and the critical set is $C = \{H < 0.81 \wedge H > 1.24\}$ on a 5% level.

Simulation	1	2	3	4	5	6	7	8	9	10
JB	18.93	24.90	14.93	26.57	14.64	6.73	3.63	24.74	39.89	36.82
BL	21.19	15.94	12.19	22.45	21.65	12.01	14.06	15.53	22.17	14.24
BL2	0.95	0.92	0.98	0.90	1.08	1.02	0.99	1.01	0.77	0.93
NL	15.57	22.06	15.34	27.66	21.30	26.89	35.04	20.60	14.85	15.97

The estimation results are reported in Table 9. It is noted that the method underestimates ξ_1 and γ_2 and overestimates κ, β and ξ_2. The overestimate of κ originates most likely from the significant smoothing that the filter perform on the volatility process σ_t (see Fig. 3). From this figure it is also noted that the filter provides a biased estimate of the σ_t-process resulting in an overestimate of β. Note that the β parameter cannot, in general, be interpreted as "the average volatility", i.e. the unconditional mean of σ_t depends on the value of γ_2 in a complicated way (the diffusion term $\xi_2\sigma_t^{\gamma_2}$ is only a *local* martingale for $\gamma_2 > 1$), see [50]. For $\rho = 0$, the specification of σ_t in (57) is the CKLS (or the generalized CIR) model [11], and the unconditional mean in this process is computed in [50]. Our results (and others

Table 9. Estimation results for Cox–Generalized CIR model (57)–(61).

Parameter	α	ξ_1	γ_1	κ	β	ξ_2	γ_2	σ_e^2	ρ
Simulation no.	0.0350	0.3000	1.5000	1.3000	0.1000	0.2000	1.5000	0.0012	−0.5000
1	0.0364	0.2628	1.4943	10.7223	0.1148	0.3389	1.1897	0.0017	−0.5358
2	0.0444	0.2190	1.5737	9.2252	0.1110	0.3034	1.2256	0.0016	−0.7013
3	0.0345	0.2435	1.4732	9.5033	0.1390	0.2493	1.2007	0.0009	−0.2880
4	0.0333	0.2340	1.4889	12.0573	0.1308	0.2028	1.3204	0.0009	−0.1880
5	0.0388	0.2258	1.4249	10.0626	0.1619	0.4128	1.1835	0.0011	−0.4569
6	0.0493	0.2224	1.5511	11.1048	0.1173	0.4073	1.3974	0.0012	−0.4575
7	0.0429	0.1665	1.5849	9.7006	0.1457	0.4309	1.3248	0.0017	−0.6146
8	0.0281	0.2297	1.4734	9.8498	0.1334	0.4016	1.0473	0.0014	−0.5893
9	0.0306	0.2676	1.4116	11.2849	0.1433	0.2981	1.8505	0.0009	−0.6132
10	0.0277	0.1666	1.5971	19.9419	0.1464	0.4971	1.1658	0.0013	−0.3676
mean ($\bar{\theta}$)	0.0366	0.2238	1.5073	11.3453	0.1344	0.3542	1.2906	0.0013	−0.4812
std. dev. (s_θ)	0.0072	0.0342	0.0660	3.1505	0.0162	0.0911	0.2200	0.0004	0.1618
$\mid t \mid$-stat	0.7020	7.0391	0.3509	10.0827	6.7057	5.3526	3.0102	0.9014	0.3668



(a) (b)

Fig. 3. Plot of the simulated and filtered stock price process S_t (a) and the simulated and filtered volatility σ_t (b) of the Cox–Generalized CIR model (57)–(61).

not reported here, see [75]) suggest that an alternative parameterization of the drift term for the σ_t process in (57) should be considered. However, a proper identification of bivariate SDEs is outside the scope of this paper. The tests in Appendix A for Gaussian white noise properties are carried out on the standardized residuals of the observed S_t-process. All tests are accepted on a 5% level.

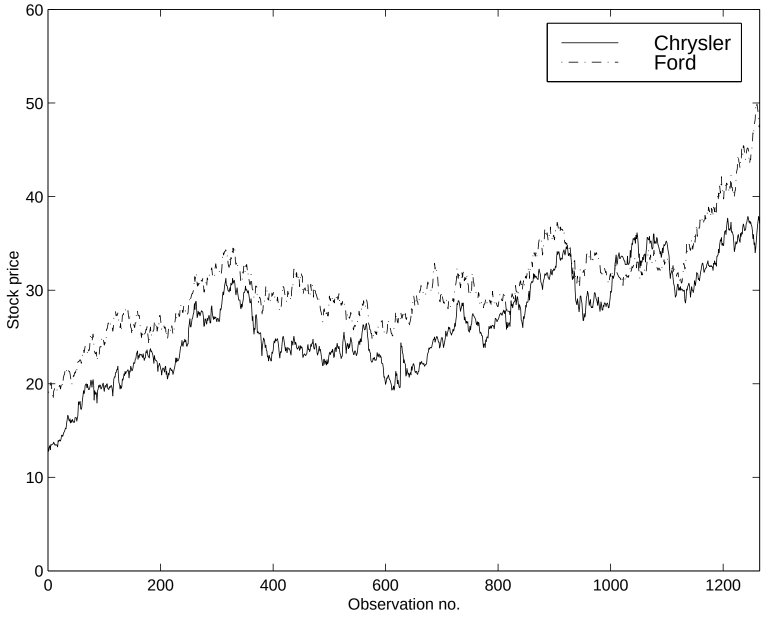


Fig. 4. A time series plot of the price of Chrysler and Ford stocks.

Table 10. Descriptive statistics for log returns of the Chrysler and Ford time series, where $\bar{x}$ is the mean, S^2 is the variance, γ_1 is the skewness, γ_2 is the excess kurtosis, JB is the Jarque–Bera statistic ($\chi^2_{95\%}(2) = 5.991$), BL is the Box–Ljung statistic ($\chi^2_{95\%}(20) = 31.410$) and BL2 is the Box–Ljung statistic for the squared log returns. The sample correlation coefficient between the two time series of stock prices is 0.8866.

Company	$\bar{x}$	s^2	γ_1	γ_2	JB	BL	BL2
Chrysler	0.0008	0.0004	1.19	13.86	10420	22.06	17.36
Ford	0.0007	0.0003	0.18	0.42	16.02	21.08	15.09

5. Empirical Work

In this section the model (50) + (54) will be estimated using real data. The data set consists of 2 time series of 1295 daily observations of Chrysler and Ford stock prices in the time period from the 20th of October 1992 to the 20th of October 1997, see Fig. 4 and Table 10. The sample time is equal to one.[i]

Estimating the parameters in the model (50)–(54) using the Chrysler and Ford time series yields the results reported in Table 11. The estimates of the instantaneous rate-of-return α and the long-term mean of the stochastic volatility β, respectively, are almost identical for the two time series. The estimates of ξ differ markedly

[i]No provisions are made for weekend and holiday effects.

Table 11. The estimated parameters for the Chrysler and Ford time series using the model (50)–(54) (the associated standard deviation in parenthesis).

Company	$\hat{\alpha}$	$\hat{\kappa}$	$\hat{\beta}$	$\hat{\xi}$	$\hat{\Sigma}$	$\hat{\rho}$
Chrysler	0.000897	10.353779	0.017633	0.343640	0.018710	−0.923491
	$(1.38 \cdot 10^{-5})$	$(4.50 \cdot 10^{-4})$	$(1.90 \cdot 10^{-6})$	$(2.72 \cdot 10^{-6})$	$(8.59 \cdot 10^{-7})$	$(4.34 \cdot 10^{-4})$
Ford	0.000896	6.480520	0.015444	0.069672	0.012684	−0.579696
	$(7.64 \cdot 10^{-5})$	$(4.09 \cdot 10^{-5})$	$(4.09 \cdot 10^{-5})$	$(1.91 \cdot 10^{-4})$	$(1.35 \cdot 10^{-4})$	$(1.32 \cdot 10^{-4})$

Table 12. Model validation tests for the Chrysler and Ford time series. JB is the Jarque–Bera test statistic for normality (A.1) and the critical value is $\chi^2_{95\%}(2) = 5.991$. BL is the Box–Ljung test statistic for autocorrelation in the residuals (A.2) and the critical value is $\chi^2_{95\%}(15) = 24.996$. BL2 is the Box–Ljung test statistic autocorrelation in the squared residuals and the critical value is $\chi^2_{95\%}(20) = 34.410$. NL is a test statistic (A.3) for heteroscedasticity, and the critical set is $C = \{H < 0.81 \wedge H > 1.24\}$ on a 5% level.

Jarque–Bera	Box–Ljung	Nonlinear	Heterosce.
16586	32.0803	13.2795	0.7384
21.23	19.6932	14.4921	0.7964

for the 2 time series, where the largest value of ξ is obtained for the Chrysler time series. This result seems reasonable considering that the descriptive statistics in Table 10 exhibit strong departure from normality for the Chrysler time series. In particular, the excess kurtosis is relatively large (13.86), which indicates heavy tails in the unconditional distribution of the returns, and this effect, *ceteris paribus*, is captured by the ξ parameter.[j] The estimates of the speed-of-adjustment parameter κ differ, but these estimates are most likely overestimated (see Table 5). The standard deviation of the Wiener process W_t^2, i.e. ξ, is significantly higher for the Chrysler stock. This also applies for the variance of the observation noise Σ.

The negative correlation between stock returns and changes in volatility have also been documented by [8, 13, 64], and this feature is a major reason for modelling the volatility itself as a diffusion process. The leverage effect is most pronounced for the Chrysler stock. It would be interesting to model the leverage effect originally introduced by [8, 13] dynamically, because [23] suggests that this effect is only a temporary behavior in the stock market. The usual model validation tests are listed in Table 12. The assumption of normality of the standardized model residuals is clearly rejected by the Jarque–Bera test. The Box–Ljung test is only rejected for the Chrysler time series, but the Box–Ljung test for no autocorrelation in the squared standardized model residuals is accepted for both time series. The test for heteroscedasticity is clearly rejected with the same critical set as used before.

[j]This is confirmed by simulation studies (not reported here), which also indicates that increasing the ξ parameter has an adverse effect on the autocorrelations of the squared log returns.

6. Conclusion

In this paper, a maximum likelihood method for direct estimation of parameters in discretely observed continuous-time stochastic volatility models has been presented, which is based on approximative second order filters. The filtering approach may, in principle, also be applied if continuous observations were available. In that case the volatility is observable (using the quadratic variation), but the inherent discreteness of market prices makes the use of SDEs inappropriate. It is shown that the method in most cases provides unbiased estimates, also of the parameters in the unobservable stochastic volatility process. The inherent smoothing of the filter does, however, lead to biased estimates of some state variables. The bias may be reduced by incorporating ODEs for the third moments [76]. Work is in progress along these lines.

The proposed modelling framework allows for a number of interesting generalizations: (i) A dynamic leverage effect may be modelled by making $\mathbf{Q}_t$ time-varying. It is also possible to specify a process for the leverage effect that is, say, consistent with the findings of [23]; and (ii) Log-returns may be used as observations in order to obtain a stationary model. These topics are left for future research.

Appendix A

Consider a time x_i, $i = 1, \ldots, N$, where N denotes the number of observations. The Jarque–Bera test for normality proposed by [41] is given by

$$JB = (N/6)\gamma_1^2 + (N/24)\gamma_2^2 \in \chi^2_{1-\alpha}(2)\,, \tag{A.1}$$

where

$$\gamma_1 = \frac{E[(X - E[X])^3]}{(V[X])^{3/2}}\,; \quad \gamma_2 = \frac{E[(X - E[X])^4]}{(V[X])^2} - 3\,.$$

The Box–Ljung test statistic for autocorrelation in a time series is given by

$$N(N+2)\sum_{\tau=1}^{L} \frac{\rho^2(\tau)}{N - \tau} \in \chi^2_{1-\alpha}(L - p)\,, \tag{A.2}$$

where $\rho^2(\tau)$ is the autocorrelation function computed up to lag L (here $L = 20$) and p is the number of parameters. The Box–Ljung test statistic may be used to test for nonlinearity by using the autocorrelation function of the squared time series [53]. A simple test for heteroscedasticity in a time series is made by comparing estimates of the variance of the first and last third parts of the time series, i.e.

$$H(h) = \sum_{i=N-h+1}^{N} x_i^2 \left(\sum_{t=1}^{1+h} x_i^2\right)^{-1} \in \mathrm{F}_{1-\alpha}(h, h)\,, \tag{A.3}$$

where h is the nearest integer to $N/3$, see e.g. [33].

Acknowledgments

An earlier version of this paper was presented at a workshop held at the University of Copenhagen in August 1998 and at the First Annual Conference on Computational and Quantitative Finance '98 held in New York in September 1998. The authors would like to thank the participants for helpful comments. Further improvements of the paper are due to an anonymous referee.

References

[1] Y. Aït-Sahalia, *Testing continuous-time models of the spot interest rate*, Rev. Financial Studies **9**(2) (1996) 385–426.
[2] Y. Aït-Sahalia, *Do interest rates really follow continuous-time Markov diffusions*, working paper, University of Chicago (1996).
[3] K. I. Amin and V. K. Ng, *Options valuation with systematic stochastic volatility*, J. Finance **48**(3) (1993) 881–910.
[4] P. Artzner, F. Delbaen, J.-M. Eber and D. Heath, *Thinking coherently*, RISK **10** (1997) 68–71.
[5] M. Baadsgaard, J. N. Nielsen, H. Spliid, H. Madsen and M. Preisel, *Estimation in stochastic differential equations with a state dependent diffusion term*, in *SYSID '97 — 11th Symposium on System Identification*, eds. Y. Sawaragi and S. Sagara, IFAC (1997).
[6] W. Bailey and R. M. Stulz, *The pricing of stock index options in a general equilibrium model*, J. Financial and Quantitative Analysis **24**(1) (1989) 1–12.
[7] C. A. Ball and A. Roma, *Stochastic volatility option pricing*, J. Financial and Quantitative Analysis **29**(4) (1994) 589–607.
[8] F. Black, *Studies of stock price volatility changes*, in Proc. 1976 Meeting of the Business and Economics Statistics Section, Amer. Stat. Assoc. (1976) 177–181.
[9] T. Bollerslev, *Generalized autoregressive conditional heteroscedasticity*, J. Econometrics **31** (1986) 307–327.
[10] T. Bollerslev, R. Y. Chou and K. F. Kroner, *ARCH models in finance: A review of the theory and evidence*, J. Econometrics **52** (1992) 5–59.
[11] K. C. Chan, G. A. Karolyi, F. A. Longstaff and A. B. Sanders, *An empirical comparison of alternative models of the short-term interest rate*, J. Finance **47**(3) (1992) 1209–1227.
[12] D. A. Chapman and N. D. Pearson, *Is the short rate drift actually nonlinear?*, J. Finance **55**(1) (2000) 355–388.
[13] A. A. Christie, *The stochastic behaviour of common stock variances: Value, leverage and interest rate effects*, J. Financial Econometrics **10** (1982) 407–432.
[14] G. Courtadon, *The pricing of options on default-free bonds*, J. Financial and Quantitative Analysis **17**(1) (1982) 75–100.
[15] J. C. Cox, *Notes in options pricing 1: Constant elasticity of variance diffusions*, Tech. Report, Stanford Univ. (1975).
[16] J. Cvitanič and I. Karatzas, *On Dynamic Measures of Risk*, Tech. Report, Dept. of Statistics, Columbia Univ., New York, April (1998). To appear in Finance & Stochastics.
[17] F. X. Diebold and M. Nerlove, *The dynamics of exchange rate volatility: A multivariate latent ARCH models*, J. Empirical Finance **1** (1989) 83–106.
[18] F. C. Drost and B. J. M. Werker, *Closing the GARCH gap: Continuous time GARCH modeling*, J. Econometrics **74** (1996) 31–57.

[19] J.-C. Duan, *Augmented GARCH(p,q) process and its diffusion limit*, J. Econometrics **79** (1997) 97–127.

[20] D. Duffie and K. Singleton, *Simulated moments estimation of Markov models of asset prices*, Econometrica **61**(4) (1993) 929–952.

[21] O. Elerian, S. Chib and N. Shephard, *Likelihood inference for discretely observed non-linear diffusions*, Tech. Report, Nuffield College, Oxford Univ. (1998).

[22] R. F. Engle, *ARCH Selected Readings*, Cambridge Univ. Press, Cambridge (1995).

[23] R. F. Engle and G. G. J. Lee, *A permanent and transitory component model of stock return volatility*, in *Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W.J. Granger*, eds. R. Engle and H. White, Oxford Univ. Press (1999) 475–497.

[24] R. F. Engle and G. J. Lee, *Estimating diffusion models of stochastic volatility*, in *Modelling Stock Market Volatility*, ed. P. E. Rossi, Academic Press (1996), Chap. 11, 333–355.

[25] B. Eraker, *MCMC Analysis of Diffusion Models with Applications to Finance*, Manuscript, NHH, Bergen, May (1998).

[26] A. R. Gallant and G. E. Tauchen, *Which moments to match?*, Econometric Theory **12**(4) (1996) 657–681.

[27] A. R. Gallant, D. Hsieh and G. E. Tauchen, *Estimation of stochastic volatility models with diagnostics*, J. Econometrics **81** (1997) 159–192.

[28] A. R. Gallant, C.-T. Hsu and G. E. Tauchen, *Using High/Low Data to Calibrate Volatility Diffusions and Extract the Forward Integrated Variance*, Manuscript, Dept. of Economics, Duke Univ. (1998).

[29] A. R. Gallant and G. Tauchen, *The Relative Efficiency of Method of Moments Estimation*, Manuscript, Dept. of Economics, Duke Univ. (1998).

[30] G. Gourieroux, A. Monfort and E. Renault, *Indirect inference*, J. Appl. Econometrics **8** (1993) 85–118.

[31] A. Grünbichler and F. A. Longstaff, *Valuing futures and options on volatility*, J. Banking and Finance **20** (1996) 985–1001.

[32] L. P. Hansen, J. A. Scheinkman and N. Touzi, *Spectral methods for identifying scalar diffusions*, J. Econometrics **86**(1) (1998) 1–32.

[33] A. C. Harvey, *Forecasting, Structural Models and the Kalman Filter*, Cambridge Univ. Press, New York (1989).

[34] A. C. Harvey, E. Ruiz and E. Sentana, *Unobserved component time series models with ARCH disturbances*, J. Econometrics **52** (1992) 129–157.

[35] A. C. Harvey, E. Ruiz and N. Shephard, *Multivariate stochastic variance models*, R. Economic Studies **61**, 247–264.

[36] S. L. Heston, *A Closed-form solution for options with stochastic volatility with applications to bond and currency options*, Rew. Financial Studies **6**(2) (1993) 327–343.

[37] D. G. Hobson and L. C. G. Rogers, *Complete models with stochastic volatility*, Math. Finance **8** (1998) 27–48.

[38] J. Hull and A. White, *Hedging the risks from writing foreign currency options*, J. International Money and Finance **6** (1987) 131–152.

[39] J. Hull and A. White, *The pricing of options on assets with stochastic volatilities*, J. Finance **42**(2) (1987) 281–300.

[40] J. Hull and A. White, *An analysis of the bias in option pricing caused by a stochastic volatility*, Adv. Futures and Options Research **3** (1988) 29–61.

[41] C. M. Jarque and A. K. Bera, *Efficient tests for normality, heteroscedasticity and serial independence of regression residuals*, Econ. Lett. **6** (1980) 255–259.

[42] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press, New York (1970).

[43] G. J. Jiang and J. L. Knight, *A nonparametric approach to the estimation of diffusion processes — with an application to a short-term interest rate model*, Econometric Theory **13** (1997) 615–645.

[44] G. J. Jiang and J. L. Knight, *Finite Sample Comparison of Alternative Estimators of Itô Diffusion Processes — A Monte Carlo Study*, J. Computational Finance **2**(3) (1999).

[45] H. Johnson and D. Shanno, *Option pricing when the variance is changing*, J. Financial and Quantitative Analysis **22**(2) (1987) 143–152.

[46] C. Jones, *A simple Bayesian approach to the analysis of Markov diffusion processes*, Tech. Report, The Wharton School, Univ. of Pennsylvania (1998).

[47] P. Jorion, *Value at Risk: The New Benchmark for Controlling Market Risk*, Irwin Professional Publ. (1996).

[48] L. A. Klimko and P. I. Nelson, *On conditional least square estimation for stochastic processes*, Ann. Statistics **6**(3) (1978) 629–642.

[49] P. E. Kloeden and E. Platen, *Numerical Solutions of Stochastic Differential Equations*, *2nd Ed.*, Springer-Verlag, Heidelberg (1995).

[50] K. H. Lind, *Properties and estimation of the CKLS-model of the term structure*, Master's Thesis, Dept. of Theoretical Statistics, Univ. of Copenhagen (1997).

[51] L. Ljung, *System Identification: Theory for the User*, Prentice-Hall, New York (1987).

[52] P. S. Maybeck, *Stochastic Models*, *Estimation and Control*, Academic Press, London (1982).

[53] A. I. McLeod and W. K. Li, *Diagnostic checking ARMA time series models using squared-residual autocorrelations*, J. Time Series Analysis **4**(4) (1983) 269–273.

[54] M. Musiela and M. Rutkowski, *Martingale Methods in Financial Modelling*, Springer (1997).

[55] D. B. Nelson, *ARCH models as diffusion approximations*, J. Econometrics **45** (1990) 7–38.

[56] D. B. Nelson, *Filtering and forecasting with misspecified ARCH models I: Getting the right variance with the wrong model*, J. Econometrics **52** (1992) 61–90.

[57] D. B. Nelson, *Asymptotically optimal smoothing with ARCH models*, Econometrica **64**(3) (1996) 561–573.

[58] J. N. Nielsen and H. Madsen, *Applying the EKF to stochastic differential equations with level effects*, Submitted (2000).

[59] J. N. Nielsen, M. Vestergaard and H. Madsen, *Nonlinear filtering of univariate stochastic volatility models*, in *Proc. of the 14th IFAC World Congress*, eds. Han-Fu Chen, Dai-Zhan Cheng and Ji-Feng Zhang **M** (1999) 123–128.

[60] A. R. Pedersen, *A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations*, Scandinavian J. Stat. **22** (1995) 55–71.

[61] M. Pritsker, *Nonparametric density estimation and tests of continuous time interest rate models*, Rev. Financial Studies **11**(3) (1998) 449–487.

[62] P. Protter, *Stochastic Integration and Differential Equations*, Springer-Verlag, New York (1990).

[63] P. E. Rossi, *Modelling Stock Market Volatility — Bridging the Gap to Continuous Time*, Academic Press, San Diego (1996).

[64] R. Schmalensee and R. R. Trippi, *Common stock volatility expectations implied by option premia*, J. Finance **33**(1) (1978) 129–147.

[65] F. Schweppe, *Evaluation of likelihood function for Gaussian signals*, IEEE Trans. on Information Theory **11** (1965) 61–70.

[66] L. O. Scott, *Option pricing when the variance changes randomly: Theory, estimation, and an application*, J. Financial and Quantitative Analysis **22**(4) (1987) 419–438.

[67] L. O. Scott, *Random variance option pricing: Empirical tests of the model and delta-sigma hedging*, Adv. Futures and Options Research **5** (1991) 113–135.

[68] N. Shephard, *Statistical aspects of ARCH and stochastic volatility*, in *Time Series Models: In Econometrics, Finance and Other Fields*, eds. D. R. Cox, D. V. Hinkley and O. E. Barndorff-Nielsen, Chapman & Hall (1996) 1–67.

[69] A. A. Smith, *Estimating nonlinear time series models using simulated vector autoregressions*, J. Appl. Econometrics **8** (1993) S63–S84.

[70] R. Stanton, *A nonparametric model of term structure dynamics and the market price of interest rate risk*, J. Finance **52**(5) (1997) 1973–2002.

[71] E. M. Stein and J. C. Stein, *Stock price distributions with stochastic volatility — An analytic approach*, Rev. Financial Studies **4**(4) (1991) 727–752.

[72] H. Tanizaki, *Nonlinear Filters, 2nd Ed.*, Springer-Verlag, Heidelberg (1996).

[73] S. J. Taylor, *Modeling stochastic volatility: A review and comparative study*, Mathematical Finance **4**(2) (1994) 183–204.

[74] J. Timmer and A. S. Weigend, *Modeling volatility using state space models*, Int. J. Neural Systems **8** (1997) 385–398.

[75] M. Vestergaard, *Nonlinear filtering methods for stochastic volatility models*, Master's Thesis, Dept. of Math. Modelling, Lyngby, Denmark (1998).

[76] D. M. Wiberg and D. G. DeWolf, *A convergent approximation of the continuous-time optimal parameter estimator*, IEEE Trans. on Automatic Control **38**(4) (1993) 529–545.

[77] J. B. Wiggins, *Option values under stochastic volatility: Theory and empirical estimates*, J. Financial Econometrics **19** (1987) 351–372.

[78] Y. Zhu and M. Avellaneda, *A risk-neutral stochastic volatility model*, Int. J. Theoretical and Appl. Finance **1**(2) (1998) 289–310.