Computational Statistics and Data Analysis 58 (2013) 15–26

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

An unscented Kalman smoother for volatility extraction: Evidence from stock prices and options

Junye Li *

ESSEC Business School, Paris-Singapore, 188064, Singapore

ARTICLE INFO

Article history:
Received 10 April 2010
Received in revised form 1 June 2011
Accepted 1 June 2011
Available online 12 June 2011

Keywords:
Nonlinear Gaussian state-space models
Nonlinear Kalman filters
Unscented Kalman smoother
Heston stochastic volatility model
Option pricing

ABSTRACT

A smoothing algorithm based on the unscented transformation is proposed for the nonlinear Gaussian system. The algorithm first implements a forward unscented Kalman filter and then evokes a separate backward smoothing pass by only making Gaussian approximations in the state but not in the observation space. The method is applied to volatility extraction in a diffusion option pricing model. Both simulation study and empirical applications with the Heston stochastic volatility model indicate that in order to accurately capture the volatility dynamics, both stock prices and options are necessary.

1. Introduction

In the state-space model framework, Bayesian optimal smoothing, also known as belief inference, refers to statistical methodology that can be used to infer state estimate using all information, which is available not only in the past and at the current, but also in the future. Optimal smoothing is closely related to optimal filtering, which makes inference based on information only available in the past and at the current. Smoothing and filtering have been successfully applied in many fields, but their applications to financial problems has not attracted enough attention.

At the core of financial econometrics is volatility estimation. Volatility pervades almost everywhere in financial markets. For example, it is used in option pricing, in portfolio allocation to control and manage risks, and in computation of risk-adjusted returns for comparison of relative performance of various financial investments. Time-varying/stochastic volatility is well documented in empirical studies. There are mainly two modeling approaches for volatility. One is the class of ARCH/GARCH models (Engle, 1982; Bollerslev, 1986), where conditional volatility is a deterministic function of past volatility and return innovations, and the other is the stochastic volatility models (Shephard, 2005), which assume that volatility is unobservable and is driven by a different random process. In the past thirty years, the diffusion process has become an common tool used to model dynamics of financial data. The diffusion stochastic volatility models (Hull and White, 1987; Heston, 1993) may be the most popular ones both in academia and in practice because of their flexibility in pricing derivatives and risk management. These models have the following general form under a given filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ for $t > 0$

$$\frac{dS_t}{S_t} = \mu_t dt + \sqrt{V_t} dW_t, \tag{1}$$

$$dV_t = \alpha(t, S_t, V_t)dt + \beta(t, S_t, V_t)dZ_t, \tag{2}$$

* Tel.: +33 1 3443 3097; fax: +33 1 3443 3212.
E-mail address: li@essec.edu.

0167-9473/$ – see front matter
doi:10.1016/j.csda.2011.06.001

where Eqs. (1) and (2) define the stock price process and the volatility process, respectively, W_t and Z_t are two standard Brownian motions, maybe mutually correlated, and $\alpha(\cdot)$ and $\beta(\cdot)$ are some deterministic functions of the stock price and/or volatility.

However, in statistical analysis of stochastic volatility models, we have two main problems. The first problem is that variances of relevant stochastic variables are state-dependent, and the second is that when taking into account derivatives, the pricing formula is nonlinear. These problems make the existent linear techniques (i.e., the standard Kalman filter and smoother) rarely applicable. Even though nonlinear Kalman filters, such as the extended Kalman filter and the unscented Kalman filter, can be used, they do not take into account all available information. Furthermore, when the system becomes highly nonlinear and high-dimensional, the extended Kalman filter may perform poorly (Wan and van der Merwe, 2001).

Recently, some works have been done for solving the above issues in nonlinear Gaussian models. Pedersen et al. (2011) introduce a finite-element method-based smoothing approach. McCausland et al. (2011) present a Gaussian simulation smoothing algorithm. This paper presents a smoothing algorithm for a nonlinear Gaussian system based on Rauch et al. (1965), Anderson and Moore (1979), and Sarkka (2008) using the unscented transformation approach, recently developed in the field of engineering (Julier and Uhlmann, 1997). The unscented transformation is a method for making Gaussian approximation to a random variable that undergoes a nonlinear transformation using the so-called sigma points to cover and propagate information on data. It can reach at least the second-order approximation accuracy. The unscented Kalman filter (Julier and Uhlmann, 1997, 2004; Wan and van der Merwe, 2001) is a straightforward application of the unscented transformation. The unscented transformation can also be applied to Bayesian optimal smoothing. Wan and van der Merwe (2001) present an unscented smoother based on the two-filter smoother (Fraser and Potter, 1969). However, the two-filter smoother needs strong unrealistic assumptions and cannot in general result in the right estimate (Klaas et al., 2006). The algorithm presented here follows the forward–backward smoother approach. It first implements a forward unscented Kalman filter and then evokes a separate backward smoothing pass by only making Gaussian approximations in the state but not in the observation space to obtain the smoothing solution for a nonlinear Gaussian system.

A simulation study is implemented using the Heston stochastic volatility model (Heston, 1993). I find that both the unscented Kalman filter and smoother cannot capture the volatility dynamics if we only use the stock price data. The filtered and smoothed volatility clearly deviate from the true path. However, when we take into account both stock prices and options, the precision of volatility filtering/smoothing gets improved dramatically. The unscented Kalman smoother performs nearly the same as the unscented Kalman filter when we use the stock price data alone, whereas it performs much better than the filter whenever both stock prices and options are used.

I apply the above algorithms to the real data on S&P 500 index and index options. Again, I find that the filtered and smoothed volatility from jointly using stock prices and options is reasonably much better than that obtained from using stock prices alone. The option pricing performance shows that the smoothed volatility can generate smaller pricing errors than the filtered one.

The rest of the paper is organized as follows. Section 2 discusses Bayesian optimal filtering and smoothing and introduces the unscented Kalman filter and smoother for a nonlinear Gaussian system. Section 3 implements a simulation study using the Heston stochastic volatility model. Section 4 presents empirical results using S&P 500 index and index options. Finally, Section 5 concludes the paper.

2. Nonlinear Gaussian systems and state extraction

In this section, I first discuss Bayesian optimal filtering and smoothing for a general state-space model in Section 2.1 and then introduce the unscented Kalman filter and smoother in Sections 2.3 and 2.4 based on the unscented transformation of Section 2.2 for a nonlinear Gaussian system.

2.1. Bayesian optimal filtering and smoothing

Consider a general dynamic state-space model with the following form:

$$y_t = H(x_t, \Theta, w_t), \tag{3}$$

$$x_t = F(x_{t-1}, \Theta, v_t), \tag{4}$$

where the observation y_t is assumed to be conditionally independent given the state x_t with the distribution $p(y_t|x_t)$, the state x_t is modeled as a Markov process with the initial distribution $p(x_0)$ and the transition law $p(x_t|x_{t-1})$, w_t and v_t are mutually independent observation noise and state noise with mean zero and variance R_t^w and R_t^v, respectively, and Θ is a set of static parameters.

Based on past and current observations, filtering is a process to estimate system's current state, that is, to find the filtering distribution $p(x_t|y_{1:t})$, where $y_{1:t}$ represents the information set up to time t, for $t = 1, 2, \ldots, T$. Bayesian optimal filtering can be implemented by the following two steps, the prediction step:

$$p(x_t|y_{1:t-1}) = \int p(x_t|x_{t-1})p(x_{t-1}|y_{1:t-1})\mathrm{d}x_{t-1}, \tag{5}$$

and the update step:

$$p(x_t|y_{1:t}) = \frac{p(y_t|x_t)p(x_t|y_{1:t-1})}{p(y_t|y_{1:t-1})}. \tag{6}$$

Thus, calculation and/or approximation of the prior $p(x_t|y_{1:t-1})$, of the likelihood $p(y_t|x_t)$, and of the evidence $p(y_t|y_{1:t-1})$ is the essence of Bayesian filtering and inference.

In contrast, Bayesian optimal smoothing is to find the smoothing distribution $p(x_t|y_{1:T})$ using all information that is available not only in the past and at the current, but also in the future. There are different smoothing methods, one of which is the two-filter smoother:

$$\begin{aligned} p(x_t|y_{1:T}) &= p(x_t|y_{1:t}, y_{t+1:T}) \\ &= \frac{p(y_{t+1:T}|x_t, y_{1:t})p(x_t|y_{1:t})}{p(y_{t+1:T}|y_{1:t})} \\ &\propto p(x_t|y_{1:t})p(y_{t+1:T}|x_t), \end{aligned} \tag{7}$$

where in (7), the first term is our familiar Bayesian filtering procedure, and the second term is called the backward information filter. However, as shown in Klaas et al. (2006), the computation of this second term does not in general lead to the right result because it is not a probability density of the state and thus its integral might not be finite. In practice, it needs strong unrealistic assumptions.

Alternatively, the smoothing distribution can be found through:

$$\begin{aligned} p(x_t|y_{1:T}) &= \int p(x_t, x_{t+1}|y_{1:T})\mathrm{d}x_{t+1} \\ &= \int p(x_t|x_{t+1}, y_{1:t})p(x_{t+1}|y_{1:T})\mathrm{d}x_{t+1} \\ &= p(x_t|y_{1:t}) \int \frac{p(x_{t+1}|y_{1:T})p(x_{t+1})|x_t}{\int p(x_{t+1}|x_t)p(x_t|y_{1:t})\mathrm{d}x_t}\mathrm{d}x_{t+1}, \end{aligned} \tag{8}$$

where the first term is the filtering density. This is the so-called forward–backward smoother, which can be implemented as follows: first, find the joint distribution of x_t and x_{t+1} conditional on information up to time t, $y_{1:t}$:

$$p(x_t, x_{t+1}|y_{1:t}) = p(x_{t+1}|x_t)p(x_t|y_{1:t}). \tag{9}$$

Second, compute the conditional distribution of x_t given x_{t+1} and $y_{1:t}$:

$$p(x_t|x_{t+1}, y_{1:t}) = \frac{p(x_t, x_{t+1}|y_{1:t})}{p(x_{t+1}|y_{1:t})}. \tag{10}$$

Due to the Markov property of the state-space model, we have $p(x_t|x_{t+1}, y_{1:T}) = p(x_t|x_{t+1}, y_{1:t})$. Finally, the smoothing density at time t can be found by:

$$\begin{aligned} p(x_t|y_{1:T}) &= \int p(x_t, x_{t+1}|y_{1:T})\mathrm{d}x_{t+1} \\ &= \int p(x_t|x_{t+1}, y_{1:T})p(x_{t+1}|y_{1:T})\mathrm{d}x_{t+1}, \end{aligned} \tag{11}$$

where the second term in the integral is the smoothing density at time $t + 1$.

If functions $H(\cdot)$ and $F(\cdot)$ are linear and if Gaussian distributions are assumed for x_t, w_t and v_t, the well-known Kalman filter/smoother can be applied, and the optimal solutions are obtainable. However, when $H(\cdot)$ and $F(\cdot)$ become nonlinear, integrals in filtering and smoothing cannot be solved analytically, and numerical approximations are required. In what follows, I will develop a smoothing algorithm for a nonlinear Gaussian system using the scaled unscented transformation, which is derivative-free and can reach higher-order approximation accuracy.

2.2. The scaled unscented transformation

The scaled unscented transformation (SUT) is a method for calculating statistics of a random variable that undergoes a nonlinear transformation (Julier and Uhlmann, 1997). For a nonlinear function:

$$y = f(x), \tag{12}$$

assume that the mean and covariance of x (with dimension L) are $\bar{x}$ and P_x. The basic idea of SUT is that the mean and covariance of y can be computed by forming a set of $2L+1$ sigma points χ:

$$\chi_0 = \bar{x}, \tag{13}$$

$$\chi_i = \bar{x} + (\sqrt{(L+\lambda)P_x})_i, \quad i = 1, \ldots, L, \tag{14}$$

$$\chi_i = \bar{x} - (\sqrt{(L+\lambda)P_x})_{i-L}, \quad i = L+1, \ldots, 2L, \tag{15}$$

where $\lambda = \alpha^2(L+\kappa) - L$ is a scaling parameter, the constant α determines the spread of sigma points around $\bar{x}$ and is usually set to be a small positive value, and κ is a second scaling parameter with value set to 0 or $3 - L$. These sigma points are propagated through the nonlinear function f:

$$\mathcal{Y}_i = f(\chi_i), \quad i = 0, 1, \ldots, 2L. \tag{16}$$

The mean and covariance of y are then approximated with a weighted sample mean and covariance of posterior sigma points:

$$\bar{y} = \sum_{i=0}^{2L} w_i^{(m)} \mathcal{Y}_i, \qquad P_y = \sum_{i=0}^{2L} w_i^{(c)} (\mathcal{Y}_i - \bar{y})(\mathcal{Y}_i - \bar{y})', \tag{17}$$

where the weights $w^{(m)}$ and $w^{(c)}$ are given by:

$$w_0^{(m)} = \frac{\lambda}{L+\lambda}, \qquad w_0^{(c)} = \frac{\lambda}{L+\lambda} + (1 - \alpha^2 + \beta), \tag{18}$$

$$w_i^{(m)} = w_i^{(c)} = \frac{1}{2(L+\lambda)}, \quad i = 1, 2, \ldots, 2L, \tag{19}$$

and superscripts (m) and (c) indicate that weights are for construction of the posterior mean and covariance, respectively, and β is a covariance correction parameter and is used to incorporate prior knowledge of x.

The scaled unscented transformation can approximate posterior mean and covariance with accuracy up to third order for Gaussian inputs, and for non-Gaussian inputs, the accuracy can be reached at least second order, with accuracy of third and higher order determined by parameters α and β. Typical values for κ, α and β are 0, 10^{-3} and 2, respectively. These values should suffice for most purposes.

2.3. The unscented Kalman filter

The unscented Kalman filter (UKF) is a straightforward application of the scaled unscented transformation. It is proposed by Julier and Uhlmann (1997, 2004). This nonlinear filter does not explicitly approximate or linearize the nonlinear observation and state models. It uses the true nonlinear models and updates state variables through a set of deterministic sigma points generated by the unscented transformation.

To implement the unscented Kalman filter, we first concatenate the state x_{t-1}, the observation noise w_{t-1}, and the state noise v_{t-1} at time $t-1$:

$$x_{t-1}^e = [x_{t-1}\ w_{t-1}\ v_{t-1}]', \tag{20}$$

whose dimension is $L = L_x + L_w + L_v$ with L_x, L_w and L_v being dimensions of the state, the observation noise, and the state noise, respectively, and whose mean and covariance are:

$$\hat{x}_{t-1|t-1}^e = E[x_{t-1}^e], \qquad P_{t-1}^e = \begin{bmatrix} P_{t-1|t-1}^x & 0 & 0 \\ 0 & R_{t-1}^w & 0 \\ 0 & 0 & R_{t-1}^v \end{bmatrix}. \tag{21}$$

We then use the scaled unscented transformation to form a set of $2L+1$ sigma points:

$$\chi_{t-1}^e = \left[\hat{x}_{t-1|t-1}^e\ \hat{x}_{t-1|t-1}^e + \sqrt{(L+\lambda)P_{t-1}^e}\ \hat{x}_{t-1|t-1}^e - \sqrt{(L+\lambda)P_{t-1}^e}\right], \tag{22}$$

and the corresponding weights are given in (18) and (19). With these sigma points, we implement the nonlinear Kalman filter as follows, for the time prediction step:

$$\chi_{t|t-1}^x = F(\chi_{t-1}^x, \chi_{t-1}^v), \qquad \hat{x}_{t|t-1} = \sum_{i=0}^{2L} w_i^{(m)} \chi_{i,t|t-1}^x,$$

$$P_{t|t-1}^x = \sum_{i=0}^{2L} w_i^{(c)} (\chi_{i,t|t-1}^x - \hat{x}_{t|t-1})(\chi_{i,t|t-1}^x - \hat{x}_{t|t-1})',$$

and for the measurement update step:

$$\mathcal{Y}_{t|t-1} = H(\chi^x_{t|t-1}, \chi^w_{t|t-1}), \qquad \hat{y}_{t|t-1} = \sum_{i=0}^{2L} w_i^{(m)} \mathcal{Y}_{i,t|t-1},$$

$$P^y_{t|t-1} = \sum_{i=0}^{2L} w_i^{(c)} (\mathcal{Y}_{i,t|t-1} - \hat{y}_{t|t-1})(\mathcal{Y}_{i,t|t-1} - \hat{y}_{t|t-1})',$$

$$P^{xy}_{t|t-1} = \sum_{i=0}^{2L} w_i^{(c)} (\chi^x_{i,t|t-1} - \hat{x}_{t|t-1})(\mathcal{Y}_{i,t|t-1} - \hat{y}_{t|t-1})',$$

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + P^{xy}_{t|t-1}(P^y_{t|t-1})^{-1}(y_t - \hat{y}_{t|t-1}),$$

$$P^x_{t|t} = P^x_{t|t-1} - [P^{xy}_{t|t-1}(P^y_{t|t-1})^{-1}]P^y_{t|t-1}[P^{xy}_{t|t-1}(P^y_{t|t-1})^{-1}]'.$$

We thus obtain the posterior mean $\hat{x}_{t|t}$ and posterior covariance $P^x_{t|t}$ of the state x_t at time t, for $t = 1, 2, \ldots, T$.

2.4. The unscented Kalman smoother

The Bayesian optimal forward–backward smoother presented in Section 2.1 can also be approximated using the scaled unscented transformation. At each time t, the filtering density $p(x_t|y_{1:t})$ and the predictive density $p(x_{t+1}|y_{1:t})$ can be found by UKF and SUT, respectively, and they are assumed to be normal. Using the basic property of the Gaussian distribution and the Markov property of the system leads to the density $p(x_t|x_{t+1}, y_{1:t}) = p(x_t|x_{t+1}, y_{1:T})$, which is also normal. Assuming that the smoothing density of time $t+1$ is known and normal $x_{t+1}|y_{1:T} \to N(\hat{x}^s_{t+1}, P^s_{x_{t+1}})$ at time t, we can then find the joint normal distribution $p(x_t, x_{t+1}|y_{1:T})$, from which a backward recursive smoothing solution can be obtained.

A single-step smoothing recursion can be performed as follows: first, concatenate the state x_t and the state noise v_t at time t and form sigma points for the augmented random variable $\tilde{x}_t = (x_t, v_t)'$:

$$\tilde{\chi}_t = \left[\hat{\tilde{x}}_{t|t} \; \hat{\tilde{x}}_{t|t} + \sqrt{(L+\lambda)\tilde{P}_t} \; \hat{\tilde{x}}_{t|t} - \sqrt{(L+\lambda)\tilde{P}_t}\right], \tag{23}$$

where

$$\hat{\tilde{x}}_{t|t} = \begin{bmatrix} \hat{x}_t \\ 0 \end{bmatrix}, \qquad \tilde{P}_t = \begin{bmatrix} P^x_{t|t} & 0 \\ 0 & R^v_t \end{bmatrix}.$$

Second, propagate the sigma points through the nonlinear state function and compute the predicted values:

$$\tilde{\chi}^x_{t+1|t} = F(\tilde{\chi}^x_t, \tilde{\chi}^v_t), \qquad \hat{\tilde{x}}_{t+1|t} = \sum_{i=0}^{2L} w_i^{(m)} \tilde{\chi}^x_{i,t+1|t},$$

$$\tilde{P}^x_{t+1|t} = \sum_{i=0}^{2L} w_i^{(c)} (\tilde{\chi}^x_{i,t+1|t} - \hat{\tilde{x}}_{t+1|t})(\tilde{\chi}^x_{i,t+1|t} - \hat{\tilde{x}}_{t+1|t})',$$

$$\tilde{C}_{t+1} = \sum_{i=0}^{2L} w_i^{(c)} (\tilde{\chi}^x_{i,t+1|t} - \hat{\tilde{x}}_{t+1|t})(\tilde{\chi}^x_{i,t} - \hat{x}_{t|t})',$$

where $\tilde{C}_{t+1}$ is the unscented transformation-based Gaussian approximation to the covariance between x_t and x_{t+1} in (9).

Finally, compute the smoothed mean and covariance of the state:

$$\hat{x}^s_t = \hat{x}_{t|t} + \tilde{C}_{t+1}(\tilde{P}^x_{t+1|t})^{-1}(\hat{x}^s_{t+1} - \hat{\tilde{x}}_{t+1|t}),$$

$$P^s_{x_t} = P^x_{t|t} + \tilde{C}_{t+1}(\tilde{P}^x_{t+1|t})^{-1}(P^s_{x_{t+1}} - \tilde{P}^x_{t+1|t})(\tilde{C}_{t+1}(\tilde{P}^x_{t+1|t})^{-1})'.$$

Because the smoothing density is the same as the filtering density at final time T, the above smoothing recursion starts from the last step with $\hat{x}^s_T = \hat{x}_{T|T}$ and $P^s_{x_T} = P^x_{T|T}$ and proceeds backward to the initial time, for $t = T, T-1, \ldots, 1$. This smoothing algorithm has its origin from the linear technique in the sense of Anderson and Moore (1979, p. 189) and Hamilton (1994, p. 394). When the system becomes linear, the above algorithm is exactly the same as the linear smoother.

3. Simulation study

In this section, I implement a simulation study to exemplify algorithms discussed in Section 2. Section 3.1 briefly introduces the Heston stochastic volatility model. Section 3.2 constructs the state-space representation. Section 3.3 conducts simulations and discusses their implications.

3.1. The Heston stochastic volatility model

The Heston stochastic volatility model (Heston, 1993) is one of the most popular models both in academia and in practice. Under a given probability space $(\Omega, \mathcal{F}, P)$ and the complete filtration $\mathcal{F}_t$, it assumes the following stock price and volatility dynamics:

$$\frac{dS_t}{S_t} = (r + \pi_W V_t)dt + \sqrt{V_t}dW_t, \tag{24}$$

$$dV_t = \kappa(\theta - V_t)dt + \sigma\sqrt{V_t}dZ_t, \tag{25}$$

where r is a constant risk-free interest rate, π_W is the market price of diffusion risk, κ is the volatility mean reversion parameter, θ is the long-run volatility parameter, and σ captures volatility of volatility. W_t and Z_t are two correlated Brownian motions with a correlation parameter $\rho \in [-1, 1]$, which accommodates the so-called leverage effect.

Assume that there exists an Equivalent Martingale measure Q, under which the risk-neutral model is defined as:

$$\frac{dS_t}{S_t} = rdt + \sqrt{V_t}dW_t^Q, \tag{26}$$

$$dV_t = [\kappa(\theta - V_t) - \pi_V V_t]dt + \sigma\sqrt{V_t}dZ_t^Q, \tag{27}$$

where W_t^Q and Z_t^Q are two Brownian motions under the risk-neutral measure and are correlated with the same parameter ρ, and π_V is the market price of volatility risk.

For this risk-neutral model, the conditional characteristic function of log returns $R_t = \ln(S_t/S_{t-\tau})$ can be derived using approaches of Duffie et al. (2000) and Carr and Wu (2004). It has the following form:

$$\begin{aligned}\phi_R(u;\tau, V_{t-\tau}) &\equiv E^Q[e^{iuR_t}|\mathcal{F}_{t-\tau}] \\ &= e^{iurt - A(u,\tau) - B(u,\tau)V_{t-\tau}},\end{aligned} \tag{28}$$

where

$$\begin{aligned}A(u,\tau) &= \frac{\kappa\theta}{\sigma^2}\left[2\log\left(1 - \frac{(\gamma - \kappa^*)(1 - e^{-\gamma\tau})}{2\gamma}\right) + (\gamma - \kappa^*)\tau\right], \\ B(u,\tau) &= \frac{2\varphi(u)(1 - e^{-\gamma\tau})}{2\gamma - (\gamma - \kappa^*)(1 - e^{-\gamma\tau})}, \\ \gamma &= \sqrt{\kappa^{*2} + 2\sigma^2\varphi(u)}, \\ \varphi(u) &= \frac{1}{2}(iu + u^2), \\ \kappa^* &= (\kappa + \pi_V) - iu\rho\sigma.\end{aligned}$$

With this conditional characteristic function, we can compute European-type options with the fast Fourier transform method (Carr and Madan, 1999). In this paper, I use a more efficient algorithm, the fractional fast Fourier transform (Chourdakis, 2005).

3.2. The state-space representation

In order to implement filtering and smoothing algorithms, we need to construct an appropriate state-space representation. I first de-correlate the stock price and volatility for applicability of Bayesian filtering and smoothing methods. Taking into account $[dW_t, dZ_t] = \rho dt$, I rewrite the stock price process (24) and the volatility process (25) as follows:

$$d\ln S_t = \left(r + \pi_W V_t - \frac{1}{2}V_t\right)dt + \rho\sqrt{V_t}dZ_t + \sqrt{1 - \rho^2}\sqrt{V_t}dW_t^*, \tag{29}$$

$$\sqrt{V_t}dZ_t = \frac{1}{\sigma}[dV_t - \kappa(\theta - V_t)dt], \tag{30}$$

where W_t^* and Z_t are independent. Note that Brownian motion W_t^* in (29) is different from that in (24). Putting (30) into (29), we obtain:

$$d\ln S_t = rdt + \left(\pi_W - \frac{1}{2}\right)V_t dt + \frac{\rho}{\sigma}[dV_t - \kappa(\theta - V_t)dt] + \sqrt{1 - \rho^2}\sqrt{V_t}dW_t, \tag{31}$$

which can be regarded as one of observation equations. The volatility process (25) is regarded as a state equation. Now we can see that the noise in (31) is independent of that in (25).

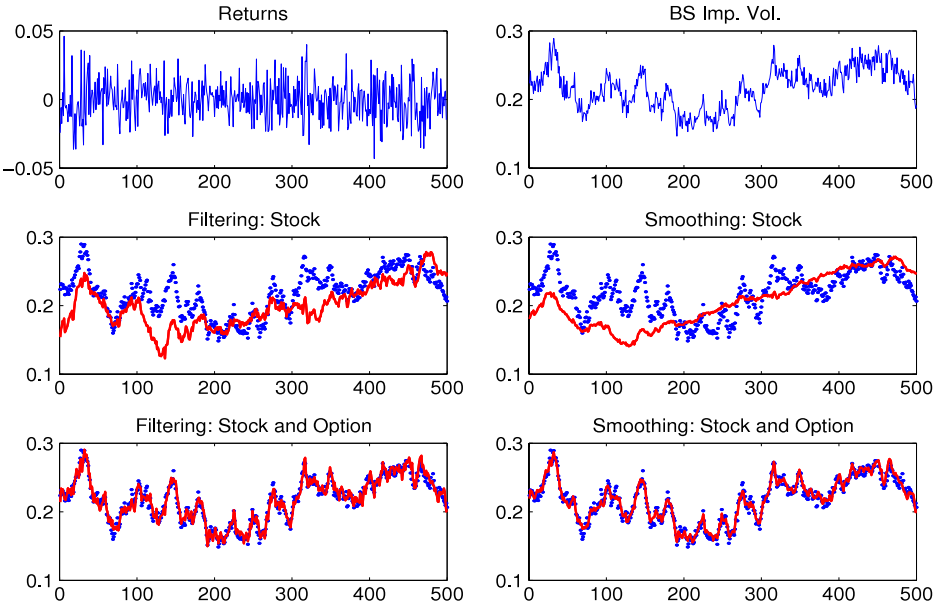

Fig. 1. Simulated data and volatility extraction. *Note:* Upper panels present the simulated stock returns (left) and the BS implied volatility (right) of the simulated at-the-money short maturity call options. The true parameters are $\Theta = (3, 0.04, 0.3, -0.6)$. The middle panels plot the filtered volatility (left) and the smoothed volatility (right) when using stock prices alone. The lower panels show the filtered volatility (left) and the smoothed volatility (right) when jointly using stock prices and options. The dots represents the true values and the lines the estimated ones.

Taking into consideration options and discretizing the model with a time interval τ, we then have the following state-space representation:

Measurement.

$$\ln S_t = \ln S_{t-\tau} + \left(r - \frac{\rho}{\sigma}\kappa\theta\right)\tau + \frac{\rho}{\sigma}V_t + \left[\frac{\rho}{\sigma}(\kappa\tau - 1) + \left(\pi_W - \frac{1}{2}\right)\tau\right]V_{t-\tau} + \sqrt{1-\rho^2}\sqrt{\tau V_{t-\tau}}w_t \tag{32}$$

$$y_t^O = f(S_t, V_t, \Theta) + \epsilon_t^O, \tag{33}$$

State.

$$\begin{pmatrix} V_t \\ V_{t-\tau} \end{pmatrix} = \begin{pmatrix} \kappa\theta\tau \\ 0 \end{pmatrix} + \begin{pmatrix} 1-\kappa\tau & 0 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} V_{t-\tau} \\ V_{t-2\tau} \end{pmatrix} + \begin{pmatrix} \sigma\sqrt{\tau V_{t-\tau}} \\ 0 \end{pmatrix} z_t, \tag{34}$$

where both V_t and $V_{t-\tau}$ are regarded as states, options y_t^O are assumed to be observed with i.i.d normal measurement errors $\epsilon_t^O \to \mathcal{N}(0, \sigma_O^2)$, independent of w_t and z_t, and $f(\cdot)$ is the theoretical option price computed from the model. In this state-space model, the option pricing measurement equation is nonlinear, and return and volatility variances are state-dependent.

3.3. Simulation implementations

For demonstration, I assume zero risk premiums ($\pi_W = 0$, and $\pi_V = 0$). Under this setting, we have parameters $\Theta = \{\kappa, \theta, \sigma, \rho\}$ and the state $x_t = V_t$. In simulations, 500 daily observations of stock prices and options are simulated with the initial values $S_0 = 100$, $V_0 = 0.04$ and the true parameters $\Theta^* = \{3.00,\ 0.04,\ 0.30,\ -0.60\}$, which are the typical values obtained from the empirical studies. The simulated options are those of at-the-money short maturity calls with strike equal to the stock price and with maturity equal to one month. I assume that the risk-free interest rate is known and equal to 5% and that option prices are contaminated by the measurement noise $\epsilon_t^O \to N(0, \sigma_O^2)$, where σ_O is set to 10%. The upper panels of Fig. 1 present a sequence of simulated stock returns and a series of the Black–Scholes implied volatility of the simulated at-the-money short maturity call options.

I first implement the unscented Kalman filter and smoother using the stock price data alone to see whether stock prices contain enough information to effectively capture the volatility dynamics. The middle panels of Fig. 1 plot the filtered volatility (left) and the smoothed volatility (right). We find that both the filtered and smoothed volatility estimates clearly deviate from the true path.

I then combine the data on stock prices and options and see whether the latter are helpful for volatility extraction. The lower panels of Fig. 1 present the filtered volatility (left) and the smoothed volatility (right). The main observations are that the precision of volatility extraction has dramatically improved both in filtering and in smoothing and that the smoothed

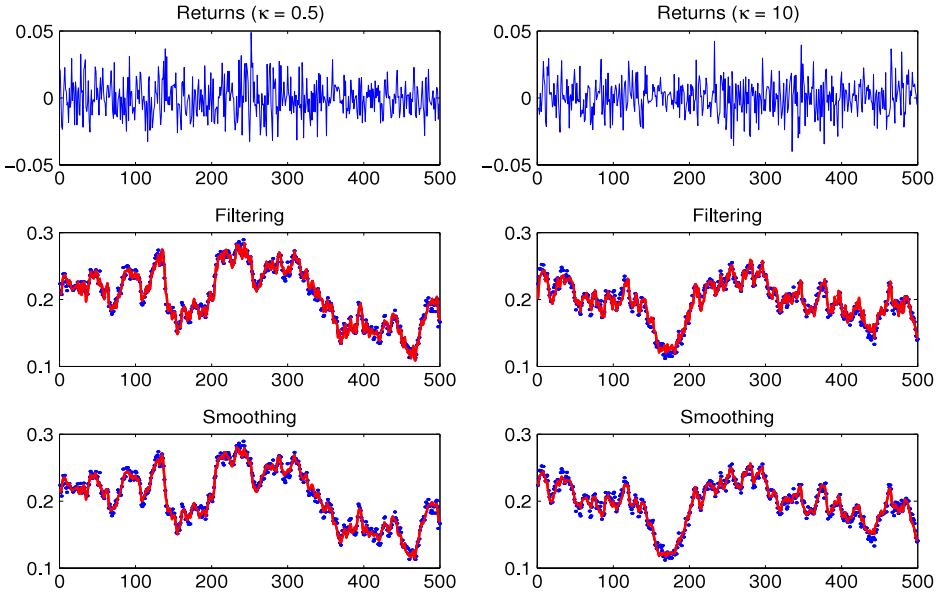

Fig. 2. Joint volatility extraction: persistent vs. non-persistent. *Note:* This figure presents the filtered and smoothed volatility for a persistent volatility process (left panels) and a non-persistent volatility process (right panels). The former is simulated using the true parameters $\Theta = (0.5, 0.04, 0.30, -0.60)$, and the latter is simulated using the true parameters $\Theta = (10, 0.04, 0.30, -0.60)$. Upper panels are the simulated stock returns, middle panels the filtered volatility and lower panels the smoothed volatility. The dots represents the true values and the lines the estimated ones.

volatility has smaller variation than the filtered volatility. Why can we achieve so much improvement? Economically, options contain different information from stock prices. The traded options encode the assessment of market participants of the volatility risk and therefore reflect the expectation of future market movements. They are forward-looking and implicitly embody all available information. However, stock prices mainly contain the historical information. Statistically, the volatility factor enters into the option pricing formula in a nonlinear form, which can further help us pin down volatility in estimation.

The parameter κ controls how fast volatility mean-reverts to its long-run mean θ. The small κ implies that the volatility process is very persistent, whereas the large κ indicates that the volatility shock on returns dissipates very quickly. I investigate these different cases and see whether the unscented Kalman filter/smoother can capture these different persistence of volatility. Fig. 2 presents the true and filtered/smoothed volatility for a persistent process ($\kappa = 0.5$, left panels) and a non-persistent one ($\kappa = 10$, right panels) using both stock prices and options. We find that for these different volatility processes, the unscented Kalman filter/smoother can efficiently capture the true values. We again observe that the smoothed volatility is less noisy than the filtered one.

σ is a volatility of volatility parameter. It governs the variation of volatility. If it is large, volatility can have a big change within a small time interval even under a persistent case. I also study cases with different values of σ. Fig. 3 presents results for a low volatility of volatility ($\sigma = 0.1$, left panels) process and a high volatility of volatility process ($\sigma = 0.8$, right panels). We find that for the low volatility of volatility case, there is no problem to extract the true volatility path using the unscented Kalman filter and/or the unscented Kalman smoother. However, we find that for the high volatility of volatility case, the unscented Kalman filter meets a problem to filter the small volatility values around data point 420. But the unscented Kalman smoother can mostly correct this problem since the smoother uses all available information, whereas the filter only uses the information up to current time.

In order to quantitatively evaluate the performance of the unscented Kalman filter/smoother (UKF/UKS) for volatility extraction, I implement 200 Monte Carlo simulations for different scenarios discussed before. For the purpose of comparison, I also present the performance of the extended Kalman filter/smoother (EKF/EKS; (Cox, 1964)) and the bootstrap particle filter/smoother (PF/PS; Klaas et al., 2006). In EKF/EKS, I use the central-difference method to approximate derivatives, and in PF/PS, I use 5000 particles when using stock prices alone and 1000 particles when using both stock prices and options. Table 1 presents means of root mean square errors (RMSE) of volatility estimates and the relative computational time of each algorithm. The following findings are obtained for the Heston stochastic volatility model. First, the unscented Kalman filter/smoother perform nearly the same as the extended Kalman filter/smoother whether we use stock prices alone or jointly use stock prices and options. Second, the unscented/extended Kalman smoothers outperform the unscented/extended Kalman filters. The improvement of the smoothers upon the filters is much more dramatic when jointly using stock prices and options. Third, with comparison to the particle filter/smoother, UKF/UKS (EKF/EKS) underperform PF/PS when we use stock prices alone. However, when we take into account both stock prices and options, the particle filter performs almost the same as UKF/EKF but underperforms UKS/EKS. Fourth, computationally, EKF/EKS is slightly faster than

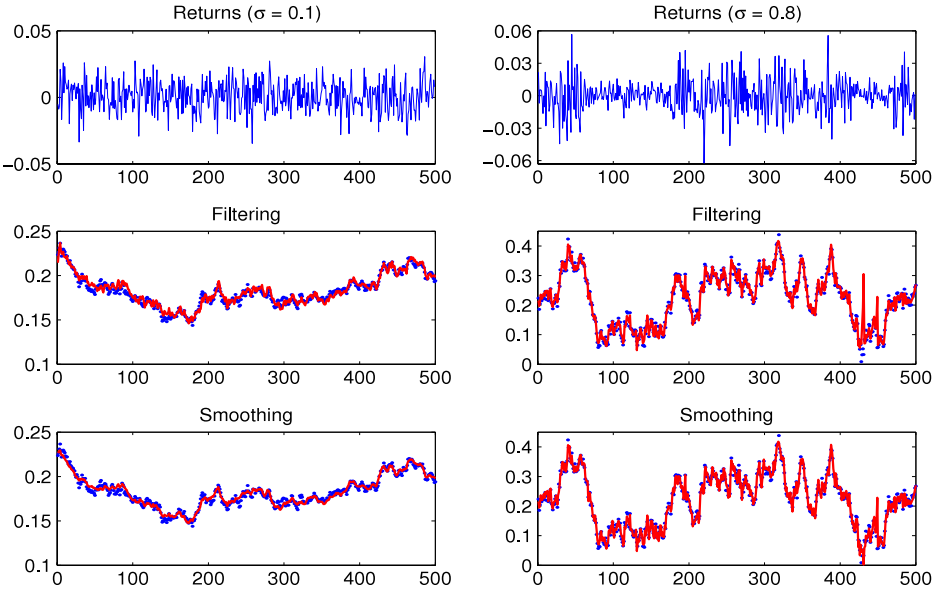

Fig. 3. Joint volatility extraction: low volatility vs. high volatility *Note:* This figure presents the filtered and smoothed volatility for a low volatility of volatility process (left panels) and a high volatility of volatility process (right panels). The former is simulated using the true parameters $\Theta = (3, 0.04, 0.10, -0.60)$, and the latter is simulated using the true parameters $\Theta = (3, 0.04, 0.80, -0.60)$. Upper panels are the simulated stock returns, middle panels the filtered volatility and lower panels the smoothed volatility. The dots represents the true values and the lines the estimated ones.

Table 1
Monte Carlo study.

	Persistence		Volatility of volatility		
	$\kappa = 0.5$	$\kappa = 10$	$\sigma = 0.1$	$\sigma = 0.8$	Avg. time
A. Stock prices alone					
EKF	2.6e−2	1.1e−2	5.9e−3	3.9e−2	0.22
EKS	2.5e−2	1.0e−2	5.8e−3	3.9e−2	0.35
PF	2.2e−2	1.0e−2	5.7e−3	3.0e−2	5.47
PS	2.1e−2	1.0e−2	5.5e−3	2.9e−2	97.2
UKF	2.6e−2	1.1e−2	5.9e−3	3.9e−2	0.52
UKS	2.5e−2	1.0e−2	5.8e−3	3.9e−2	1.00
B. Stock prices and options					
EKF	2.8e−3	3.3e−3	1.9e−3	3.5e−3	0.61
EKS	2.0e−3	2.7e−3	1.3e−3	3.3e−3	0.86
PF	2.6e−3	3.3e−3	1.9e−3	3.5e−3	64.5
PS	–	–	–	–	–
UKF	2.8e−3	3.3e−3	1.9e−3	3.5e−3	0.91
UKS	2.0e−3	2.7e−3	1.3e−3	3.3e−3	1.00

Table presents means of root mean square errors across 200 Monte Carlo simulations for different volatility processes. The persistent volatility process is simulated using the true parameters $\Theta = (0.5, 0.04, 0.30, -0.60)$, and the non-persistent process is simulated using the same parameters as the persistent process except κ, which has a value of 10. The low volatility of volatility process is simulated using the true parameters $\Theta = (3, 0.04, 0.1, -0.60)$, and the high volatility of volatility process is simulated using the same parameters as the low volatility of volatility process except σ, which has a value of 0.8. EKF/EKS stand for the extended Kalman filter/smoother, UKF/UKS for the unscented Kalman filter/smoother, and PF/PS for the bootstrap particle filter/smoother. The last column presents the average relative computational time where the computational time of the UKS is normalized to 1.

UKF/UKS. However, PF/PS are highly computationally demanding. In particular, when we use both stock prices and options, the particle smoother is too computationally demanding to be practically feasible. Finally, the reasons that EKF/EKS perform as well as UKF/UKS are that in the Heston stochastic volatility model, the state system is linear and one-dimensional, and options are priced using exponential affine characteristic function.

4. Empirical applications

This section investigates the unscented Kalman smoother with the real data on S&P 500 index and index options. Section 4.1 describes the data, Section 4.2 presents the filtered and smoothed volatility, and Section 4.3 discusses the option pricing implications.

Table 2
Descriptive statistics of the data and parameter estimates.

	Mean	St. Dev.	Max	Min	Skewness	Kurtosis
A. S&P 500 Index returns						
Weekly	0.050	0.165	0.102	−0.108	−0.230	5.077
B. Constructed calls						
	Mean Mn.	Std Mn.	Mean Mt.	Std Mt.	Mean IV.	Std IV.
OTM	0.952	0.005	29.11	9.911	0.164	0.058
ATM	1.000	0.003	27.55	9.292	0.188	0.063
C. Parameter estimates						
	κ	θ	σ	ρ	π_V	μ
Estimate	2.389	0.042	0.329	−0.819	−1.858	0.048
Std. Dev	(0.428)	(0.007)	(0.064)	(0.105)	(2.947)	(0.013)

Table presents the descriptive statistics of weekly data on the S&P 500 index and index options from January 1996 to September 2008 and the parameter estimates of the Heston stochastic volatility model. In panel A, the mean and standard deviation are annualized. In panel B, Mn stands for moneyness, Mt for maturity (in days), and IV for the Black–Scholes implied volatility. In panel C, parameters are estimated using a EM algorithm based on the unscented Kalman smoother.

4.1. Data

The data used are S&P 500 index and index options traded in the Chicago Board Options Exchange (CBOE) during the period from January, 1996 to September, 2008. They are in weekly frequency and are those traded on Wednesday. If Wednesday is a holiday, we select Thursday options. There are in total 664 weeks. The data are obtained from *OptionMetrics*.

For purpose of volatility extraction, I construct two sets of options. One is the at-the-money short maturity (ATM-SM) calls with maturity greater than 15 days and less than 50 days and with moneyness (S/K) closest to 1, and the other is the out-of-the-money short maturity (OTM-SM) calls with maturity greater than 15 days and less than 50 days and with moneyness (S/K) closest to 0.95. Now we have three sets of observations: one set of stock prices and two sets of option prices. Table 2 presents summary statistics of index returns (panel A) and the constructed options (panel B).

When investigating option pricing implications with the filtered/smoothed volatility, I implement the following filters to the dataset. First, we only consider call options. Second, in order to ensure that options are liquid enough, we select call options with maturity less than 1.5 years and with moneyness greater than 0.90 and less than 1.03. Furthermore, I rule out options with zero trading volume and with open interest less than 100 contracts. Lastly, I exclude call options with maturity less than 10 days and best bid prices less than 3/8 dollar to mitigate market microstructure problems. As a result, there are 28,557 call options in total and 43 options each day on average.

4.2. Volatility extraction

The Heston stochastic volatility model discussed in the previous section is again taken as an example. I first obtain parameter estimates using a likelihood-based method. Panel C of Table 2 presents parameter estimates. The objective mean-reverting parameter is about 2.4, the long-run mean of the volatility process is around 20%, and the volatility of volatility is 0.33. All of them are highly significant. The negative estimate of the volatility risk premium (not significant) indicates that the risk-neutral volatility process is much more persistent than the objective one. With these parameter estimates, I then implement the filtering and smoothing algorithms.

Fig. 4 presents the filtered/smoothed volatility using the S&P 500 index and/or index options. The top panels are index returns and the average Black–Scholes implied volatility, and the middle panels present the filtered (left panel) and smoothed (right panel) volatility using stock prices (dashed line) and jointly using stock prices and options (solid line). With comparison to the evolution of index returns and the Black–Scholes implied volatility, we can clearly see that the filtered and smoothed results from jointly using stock prices and options are reasonably much better than those from using stock prices alone.

The lower panels plot differences between the filtered and smoothed volatility obtained from using stock prices alone (left panel) and from jointly using stock prices and options (right panel). We note that there can be a very big difference between the filtered and smoothed volatility when only using stock prices, whereas more information can make this difference small.

We also notice that even though the smoothed volatility has less variation than the filtered one, they look very similar to each other, especially those obtained from jointly using both datasets. This is because the volatility process is model with a diffusion process, is persistent and cannot have a large change in a short time interval.

4.3. Option pricing implications

We now investigate option pricing implications. I divide the constructed call options into 12 groups whose maturities are 10–60 days, 60–180 days and larger than 180 days, and whose moneyness are 0.90–0.94, 0.94–0.97, 0.97–1.00 and

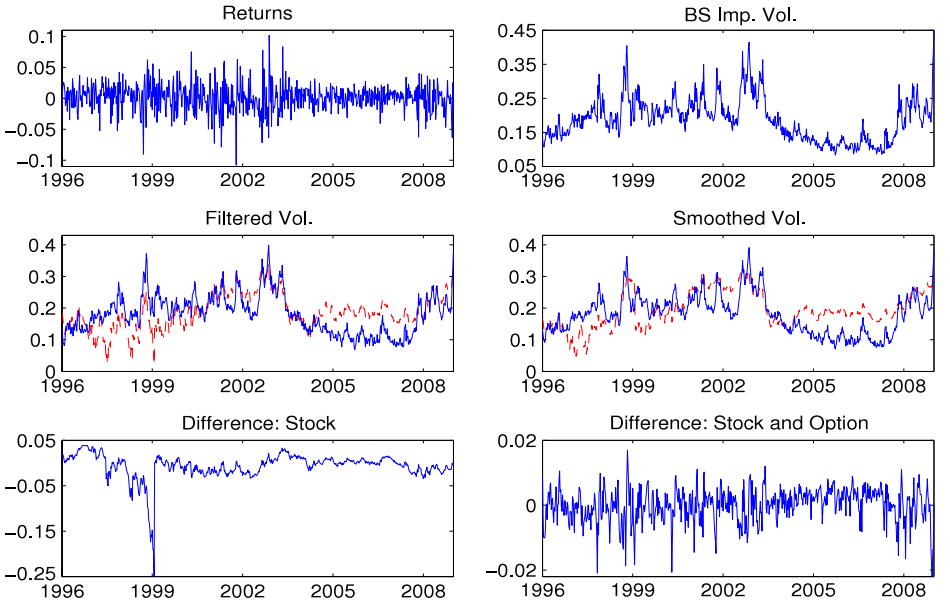

Fig. 4. Volatility extraction using real data. *Note:* This figure presents the filtered and smoothed volatility for the Heston stochastic volatility model using real data on the S&P 500 index and index returns. The upper panels are the index returns (left) and the average BS implied volatility (right). The middle left panel plots the filtered volatility when using stock prices alone (dashed line) and when jointly using stock prices and options (solid line). The middle right panel plots the smoothed volatility when using stock prices alone (dashed line) and when jointly using stock prices and options (solid line). The lower left panel shows difference between the filtered volatility and the smoothed one when using stock prices alone. The lower right shows difference between the filtered volatility and the smoothed one when jointly using stock prices and options.

Table 3
Ratios of absolute option pricing errors.

Maturity	Ratio	Moneyness (S/K)			
		0.90–0.97	0.94–0.97	0.97–1.00	1.00–1.03
10–60	VFJ/VFS	0.591	0.428	0.558	0.407
	VSJ/VSS	0.536	0.478	0.541	0.409
	VSJ/VFJ	1.044	0.905	0.985	1.049
60–180	VFJ/VFS	0.497	0.484	0.592	0.570
	VSJ/VSS	0.552	0.531	0.406	0.419
	VSJ/VFJ	0.873	0.859	0.942	0.975
>180	VFJ/VFS	0.560	0.583	0.531	0.587
	VSJ/VSS	0.547	0.434	0.565	0.455
	VSJ/VFJ	1.078	0.875	0.905	0.964

Table presents ratios of option pricing errors. VFJ/VFS (VSJ/VSS) represents the ratio of the option pricing error using the filtered (smoothed) volatility obtained from jointly using stock prices and options and from using stock prices alone. VSJ/VFJ represents the ratio of the option pricing error using the smoothed volatility and the filtered volatility when jointly using stocks and options.

1.00–1.03. I then compute the absolute option pricing error using the filtered and smoothed volatility obtained from the previous subsection. The absolute option pricing error is defined as follows

$$Aerr = \frac{1}{N}\sum_{t=1}^{T}\sum_{i=1}^{n_t}|P_{ti}^{im} - P_{ti}|, \tag{35}$$

where N is the total number of options we consider, T the number of weeks, n_t the number of options at date t, and P_{ti}^{im} and P_{ti} are the model-implied and the market-observed option prices of the ith option at date t, respectively.

Table 3 presents ratios of the absolute option pricing errors. VFJ/VFS (VSJ/VSS) represents the ratio of the option pricing error using the filtered (smoothed) volatility obtained from jointly using stock prices and options and from using stock prices alone. VSJ/VFJ represents the ratio of the option pricing error using the smoothed volatility and the filtered volatility when jointly using stock prices and options. It is very clear that all of ratios of VFJ/VFS and VSJ/VSS are small (from 0.407 to 0.592 and from 0.406 to 0.565), indicating that the volatility obtained from jointly using stock prices and options is much more accurate than that obtained from using stock prices alone. In most groups, the ratios of VSJ/VFJ are smaller than 1, implying that volatility obtained from the smoother is better than that obtained from the filter. We have three ratios of VSJ/VFJ larger than one. This may be because of the model misspecification problem as we know that the Heston model works poorly for short maturity and out-of-the-money options.

5. Conclusion

This paper proposes a smoothing algorithm based on the unscented transformation and shows how it can be used for volatility extraction in diffusion models. The algorithm firstly implements a forward unscented Kalman filter and then provokes a separate backward smoothing pass to obtain the smoothing solution for nonlinear systems. Simulation study and empirical applications with the Heston stochastic volatility model indicate that in order to accurately capture the volatility dynamics, both stock prices and options are necessary. The paper also finds that volatility obtained from the smoother can in general result in smaller option pricing errors than that obtained from the filter.

The rapid development of sequential Monte Carlo methods (particle filters/smoothers) paves another way to statistically analyze the stochastic volatility models. Particle methods are very general and can be applied to any nonlinear and/or non-Gaussian models. However, in practice, the high computational cost in option pricing makes it impossible to choose large number of particles, and this may make particle methods less efficient. The paper provides a practically efficient method for analyzing the diffusion stochastic volatility models.

References

Anderson, D.O., Moore, J.B., 1979. Optimal Filtering. Prentice-Hall, New Jersey.
Bollerslev, T., 1986. Generalized autoregressive conditional heteroscedasticity. Journal of Econometrics 31, 307–327.
Carr, P., Madan, D.B., 1999. Option valuation using the fast Fourier transform. Journal of Computational Finance 3, 61–73.
Carr, P., Wu, L., 2004. Time-changed Lévy processes and option pricing. Journal of Financial Economics 71, 113–141.
Chourdakis, K., 2005. Option pricing using the fractional FFT. Journal of Computational Finance 8, 1–18.
Cox, H., 1964. On the estimation of state variables and parameters for noisy dynamic systems. IEEE Transactions on Automatic Control 9, 5–12.
Duffie, D., Pan, J., Singleton, K., 2000. Transform analysis and asset pricing for affine jump-diffusions. Econometrica 68, 1343–1376.
Engle, R., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflations. Econometrica 50, 987–1008.
Fraser, D., Potter, J., 1969. The optimum linear smoother as a combination of two optimum linear filters. IEEE Transactions on Automatic Control 14, 387–390.
Hamilton, J.D., 1994. Time Series Analysis. Princeton University Press, New Jersey.
Heston, S.L., 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. Review of Financial Studies 6, 327–343.
Hull, J., White, A., 1987. The pricing of options on assets with stochastic volatilities. Journal of Finance 42, 281–300.
Julier, S.J., Uhlmann, J.K., 1997. A new extension of the Kalman filter to nonlinear systems. In: Proceedings of AeroSense: The 11th International Symposium on Aerospace/Defense Sensing, Simulation and Controls, pp. 182–193.
Julier, S.J., Uhlmann, J.K., 2004. Unscented filtering and nonlinear estimation. Proceedings of the IEEE 92, 401–421.
Klaas, M., Bries, M., de Freitas, N., Doucet, A., Maskell, S., Lang, D., 2006. Fast particle smoothing: if I had a million particles. In: Proceedings of ICML 2006, pp. 481–488.
McCausland, W.J., Miller, S., Pelletier, D., 2011. Simulation smoothing for state-space models: a computational efficiency analysis. Computational Statistics and Data Analysis 55, 199–212.
Pedersen, M.W., Thygesen, U.H., Madsen, H., 2011. Nonlinear tracking in a diffusion process with a Bayesian filter and the finite element method. Computational Statistics and Data Analysis 55, 280–290.
Rauch, H.E., Tung, F., Striebel, C.T., 1965. Maximum likelihood estimates of linear dynamic systems. AIAA Journal 3, 1445–1450.
Sarkka, S., 2008. Unscented Rauch–Tung–Striebel smoother. IEEE Transactions on Automatic Control 53, 845–849.
Shephard, N., 2005. Stochastic Volatility: Selected Readings. Oxford University Press, Oxford.
Wan, E., van der Merwe, R., 2001. The unscented Kalman filter. In: Haykin, S. (Ed.), Kalman Filtering and Neural Networks. John Wiley & Sons, New York.