Available online at www.sciencedirect.com
ScienceDirect
Journal of Empirical Finance 14 (2007) 424–442
Journal of EMPIRICAL FINANCE
www.elsevier.com/locate/jempfin

Specification and estimation of discrete time quadratic stochastic volatility models

Hiroyuki Kawakatsu

Business School, Dublin City University, Dublin 9, Ireland

Received 7 November 2005; received in revised form 13 July 2006; accepted 17 July 2006
Available online 30 August 2006

Abstract

I propose a new class of stochastic volatility models that nests the commonly used log normal autoregressive specification. As with the eigenfunction specification of Meddahi (Meddahi, Nour, 2001. An eigenfunction approach for volatility modeling. Unpublished.), the log-quadratic model can generate high kurtosis, a key feature of asset returns, even with Gaussian innovations. I discuss maximum likelihood estimation based on numerical integration of the log-quadratic specification that allows for leverage effects. A small Monte Carlo simulation experiment demonstrates the feasibility of maximum likelihood estimation and the importance of allowing for leverage effects. I fit the log-quadratic specification to the daily S&P 500 index return series and find that it provides a better fit than the commonly used log autoregressive specification with Gaussian and Student-*t* mean equation innovations.

JEL classification: C22; C51
Keywords: Stochastic volatility; Leverage; Quadratic; Numerical integration

1. Introduction

Estimating volatility of financial assets has become increasingly important with the development of various sophisticated financial instruments such as options. As a result, a number of statistical methods for modelling and measuring volatility have been developed. Of particular importance in the option pricing literature is the class of stochastic volatility models. The canonical log -normal autoregressive (LNAR) model, commonly used in the literature, can capture volatility clustering via a highly persistent (near unit root) autoregressive process.

E-mail address: hiroyuki.kawakatsu@dcu.ie.

0927-5398/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jempfin.2006.07.001

However, the LNAR model does not appear to be able to capture extreme events such as market crashes that occur infrequently (Bakshi et al., 1997; Andersen et al., 2002; Meddahi, 2001).

To better capture the fat tails in the volatility distribution, the canonical LNAR model has been extended in several directions. The most natural extension is to consider distributions other than the normal with heavy tails within the log autoregressive model (Fridman and Harris, 1998; Liesenfeld and Jung, 2000; Bai et al., 2003). Andersen et al. (2002) introduce jumps in the mean specification, while Eraker et al. (2003) introduce jumps in the volatility specification as well. Chernov et al. (2003) consider multifactors in the volatility specification to decouple tail thickness from volatility persistence.

The approach in this paper follows the literature on flexible functional form specifications. Yu et al. (2002) consider Box-Cox transformations in the volatility specification, while Meddahi (2001) considers an eigenfunction approach. These specifications model non-Gaussianity through non-linear terms in the volatility specification. In particular, the approach of Meddahi (2001) can be seen as a semi-parametric expansion of the volatility equation.

There are three main contributions of this paper. First, in Section 2, I propose a new specification in which the log volatility is a quadratic function of a latent state variable. This log-quadratic specification (LQSV) is closely related to the Hermite stochastic volatility model of order 2, HSV(2), proposed by Meddahi (2001). As explained in Section 2, the main advantage of the LQSV specification over the HSV(2) specification is that the variance is ensured to be positive. Section 2.2 characterizes the properties of the LQSV process by deriving the (unconditional) moments and autocovariances of the observed series. As with the HSV(2) model, it is shown that the LQSV specification is able to generate high kurtosis even with Gaussian innovations.

Second, in Section 3, I discuss alternative approaches to approximate maximum likelihood estimation of the LQSV specification with leverage. Of the three alternatives discussed, I find that the numerical quadrature based procedure by Fridman and Harris (1998) extended to handle specifications with leverage effects performs most reliably. Section 3.3 shows that the nonlinear filtering recursions used to evaluate the likelihood can be easily adapted to evaluate statistics from the one-step prediction density for diagnostic checking and value-at-risk applications. A small Monte Carlo simulation experiment is conducted in Section 3.4 to assess the finite sample properties of the proposed estimator. The simulations illustrate the potential gain in efficiency of the finite sample estimates when the model is correctly specified to account for leverage.

Third, in Section 4, I fit the LQSV specification to daily returns from the S&P 500 index for two samples, one that contains the market crash in October 1987 and another more recent sample. The fit of the LQSV model is evaluated by comparing it to two commonly used alternative specifications: the log AR(1) volatility model with Gaussian and Student-t mean equation innovations. For both samples, standard statistical tests reject the LNAR model in favor of the LQSV model. I also illustrate the use of several diagnostics to evaluate the fit of the alternative specifications. These diagnostics and model selection based on the information criterion generally favor the LQSV specification over the log AR(1) model with Student-t innovations.

2. Quadratic class stochastic volatility models

2.1. Specifications

In this section, I describe discrete time quadratic class specifications of stochastic volatility. There is a growing literature on continuous time quadratic class models, especially in the term structure literature (Leippold and Wu, 2002; Ahn et al., 2002; Leippold and Wu, 2003; Cheng and Scaillet, 2004). The eigenfunction specification of Meddahi (2001) is given both in discrete and continuous time. This paper focuses on discrete time quadratic stochastic volatility specifications.

The main difficulty with the eigenfunction approach of Meddahi (2001) is that the specification imposes constraints on the parameters to ensure positivity of the volatility. For example, the discrete time Hermite stochastic volatility model of order p, HSV(p), is specified as

$$y_t = \sigma_t u_t \tag{1a}$$

$$\sigma_t^2 = \sum_{j=0}^{p} a_j H_j(x_t), \qquad \sum_{j=0}^{p} a_j^2 < \infty \tag{1b}$$

$$x_{t+1} = \phi x_t + \sqrt{1-\phi^2} v_{t+1}, \quad x_0 \sim N(0,1), \quad |\phi|<1 \tag{1c}$$

$$\begin{bmatrix} u_t \\ v_{t+1} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right) \tag{1d}$$

where y_t is the observed stochastic process, x_t is the state variable, u_t, v_t are the innovations, and $H_j(x)$ are the Hermite polynomials. The parameters to be estimated are $\theta = (a_j, \phi, \rho)$. ϕ measures the persistence of the standardized latent state variable x_t and a_j are the location (a_0) and scale parameters of the volatility process. The correlation parameter $-1 < \rho < 1$ allows for leverage effects (Yu, 2005).

The key departure of this specification from the standard LNAR model is that the volatility σ_t^2, rather than log σ_t^2, is specified as a polynomial of order p in the underlying state variable x_t. The cost of this generalization is that the parameters a_j must be constrained so that the volatility remains positive. For $p=2$, this requires

$$a_1^2 - \sqrt{8} a_0 a_2 + 2a_2^2 \le 0 \quad \text{and} \quad a_2 > 0 \tag{2}$$

These non-linear constraints need to be imposed when estimating the parameters a_0, a_1 and a_2. Meddahi (2001) derives moment conditions for eigenfunction stochastic volatility models and suggests estimation based on GMM.

To circumvent the positivity constraints required by the eigenfunction specification, I specify the log of volatility as a quadratic in the underlying state variable. That is, in the HSV specification (1), I replace (1b) with

$$\log(\sigma_t^2) = a_0 + a_1 x_t + a_2 x_t^2 \tag{3}$$

This alternative specification is mentioned but not fully developed in Meddahi (2001, p. 12) and is shown to be a special case of HSV(∞).[1]

There are two main advantages of considering this alternative specification, which I call the log-quadratic stochastic volatility (LQSV) model. First, the positivity of volatility is ensured and does not impose constraints on the parameters, except to ensure that higher order moments of the

[1] As mentioned in Meddahi (2001), to generalize the quadratic function to a polynomial of order $p>2$, we need to ensure square integrability of the volatility process by restricting the parameters of higher order monomials.

volatility process exist. As shown below in Section 2.2, the fourth moment of y_t exists provided $a_2<1/4$, a bound constraint that can be easily imposed in estimation if so desired. Second, it is easy to see that, if $a_2=0$, the LQSV model reduces to the canonical LNAR model. The HSV(p) specification (1), on the other hand, nests the LNAR model only as $p\to\infty$ (Meddahi, 2001). Because the LQSV specification nests (or "encompasses") the canonical LNAR model, we can test the adequacy of the LNAR model by classical inference procedures for testing the hypothesis that $a_2=0$.

As mentioned in the Introduction, the main motivation for considering the quadratic class is to capture the heavy-tailed property of asset returns. One could achieve this by simply replacing the Gaussian assumption (1d) with a heavy-tailed non-Gaussian distribution. This is the approach taken by, for example, Liesenfeld and Jung (2000) and Bai et al. (2003) for the canonical log autoregressive specification. However, I do not consider non-Gaussian alternative distributions for quadratic specifications. Replacing the distribution of the volatility innovation v_{t+1} with a non-Gaussian distribution requires care to ensure that the second moment of y_t exists. For example, as noted by Jacquier, Polson and Rossi (2004, p. 190), the moments of y_t do not exist if v_{t+1} has a Student-t distribution. For this reason, the t distribution specification has been used only for the mean equation innovation u_t. As discussed below, the quadratic specification can capture fat tails in the observed asset returns even with Gaussian innovations. Moreover, the moments of the LQSV specification derived below rely heavily on the Gaussian assumption. In particular, derivation of the autocovariance function of the squared returns appears to be intractable without the Gaussian assumption.[2]

2.2. Properties of the LQSV process

The log-quadratic stochastic volatility (LQSV) model examined in this paper consists of Eqs. (1a), (3), (1c) and (1d). To derive the moments of y_t, note that $x_t\sim N(0,1)$ and that $\mathrm{cor}(u_t, v_s)=0$ for $s\neq t+1$. As shown in Appendix A.1, the r-th moment of y_t is given by

$$E[y_t^r]=E[\sigma_t^r u_t^r]=E[\sigma_t^r]E[u_t^r]=\begin{cases}0 & \text{if } r \text{ odd}\\ \dfrac{r!}{2^{\frac{r}{2}}\left(\frac{r}{2}\right)!}\dfrac{1}{\sqrt{1-ra_2}}\exp\left(\dfrac{r}{2}a_0+\dfrac{r^2a_1^2}{8(1-ra_2)}\right) & \text{if } r \text{ even}\end{cases}$$

where the moment exists provided $a_2<1/r$. I note that these moments do not depend on the leverage parameter ρ and the volatility persistence parameter ϕ. The kurtosis of y_t is given by

$$K[y_t]=\frac{E[y_t^4]}{(E[y_t^2])^2}=\frac{3(1-2a_2)}{\sqrt{1-4a_2}}e^{\frac{a_1^2}{(1-2a_2)(1-4a_2)}} \tag{4}$$

for $a_2<1/4$. Note that the kurtosis depends only on two parameters a_1 and a_2. The kurtosis for the HSV(2) specification (1) is given by $K[y_t]=3(1+(a_1^2+a_2^2)/a_0^2)$ (Meddahi, 2001).

[2] Another reason for restricting attention to Gaussian innovations is the robustness property of the QML estimator based on the Gaussian nominal likelihood to certain distributional misspecifications (White, 1994). This has been demonstrated for GARCH models by Newey and Steigerwald (1997). However, I am not aware of a study that examines whether similar results hold for stochastic volatility models.

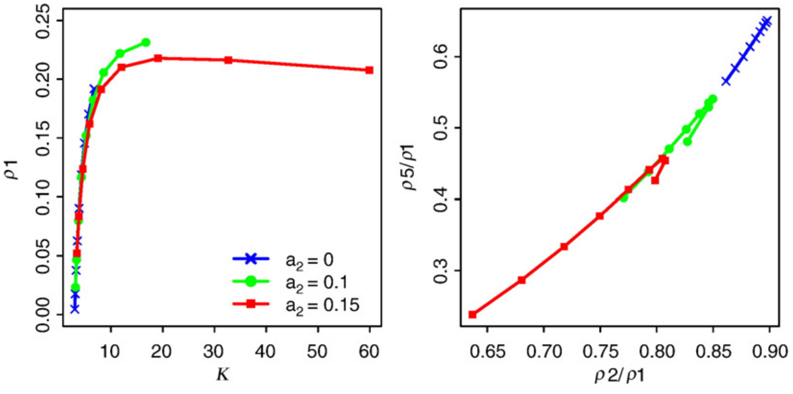

Fig. 1. Moments of the log-quadratic specification for $\phi=0.9$, $\rho=-0.3$ and various values of a_1, a_2. The left panel plots the implied kurtosis $K[y_t]$ and first-order autocorrelation $\rho_1=\mathrm{cor}(y_t^2, y_{t-1}^2)$ for three values of a_2 as we vary a_1 from 0.1 to 0.9. The right panel plots the decay rate of the autocorrelation function $\rho_j=\mathrm{cor}(y_t^2, y_{t-j}^2)$ for the same three values of a_2 as we vary a_1 over the same range as the left panel.

The autocovariances of y_t and y_t^2 for the LQSV are given by

$$E[y_t y_{t-j}] = 0$$

$$E\left[y_t^2 y_{t-j}^2\right] = \frac{1}{(c_1)^{1/2}(c_2)^{5/2}}\left(\frac{c_2 + a_1^2 c_3}{(c_4)^{1/2}} + \frac{4a_1 a_2 c_3 c_5 \phi^j}{(c_4)^{3/2}} + \frac{4a_2^2 c_3 (c_4 + c_5^2)\phi^{2j}}{(c_4)^{5/2}}\right) \times \exp\left(c_6 + \frac{c_5^2}{2c_4}\right)$$

for $j>0$ and the expressions for c_j, $j=1,\ldots,6$ are given in Appendix A.2. For stationary LQSV models with $|\phi|<1$, it is shown in Appendix A.2 that $E[y_t^2 y_{t-j}^2]\to(E[y_t^2])^2$ and $\mathrm{cor}(y_t^2, y_{t-j}^2)\to 0$ as $j\to\infty$.[3]

Given these expressions for the moments of the observable variable y_t, we can examine the observable implications of the additional parameter a_2 in the LQSV specification. As the moments are fairly complex functions of the parameters, I resort to numerical analysis. As pointed out by Liesenfeld and Jung (2000), for the LNAR model without leverage ($\rho=0$), the first-order autocorrelation $\rho_1=\mathrm{cor}(y_t^2, y_{t-1}^2)$ of the squared returns is functionally related to the kurtosis K of y_t via the parameter ϕ. For the LQSV model with leverage, the kurtosis K depends on the parameters a_1 and a_2, while ρ_1 depends on a_1, a_2, ϕ and ρ. Thus, while the parameter a_2 adds a "degree of freedom" to decouple these two moments, the two moments are linked via the parameters a_1 and a_2. To see this dependence, the left panel in Fig. 1 plots the pair (K, ρ_1) as we vary a_1 and a_2 for $\phi=0.9$ and $\rho=-0.3$. As is now widely recognized, the LNAR specification ($a_2=0$) can explain only a limited amount of excess kurtosis. Note that varying the fixed parameters ϕ or ρ will only shift the curves vertically in the panel as K does not depend on those parameters. The panel shows that the additional parameter a_2 enables the LQSV specification to explain a much wider range of excess kurtosis.

Another well-known stylized fact is that squared returns exhibit long-memory type behavior: "the sample autocorrelation usually decreases much faster than exponentially at the beginning

[3] The partial autocorrelation is not addressed in the current paper as it appears to be intractable.

and then decreases much slower and remains significantly positive over long lags" (Ding and Granger, 1996, p. 199). The right panel in Fig. 1 examines the decay rate of the autocorrelation ρ_j of the squared returns over two short lags j. The decay rate is measured by the ratios ρ_2/ρ_1 and ρ_5/ρ_1. Compared to the LNAR specification ($a_2=0$), the autocorrelation function decays more rapidly as we increase a_2. This rapid decay of autocorrelations at short lags for the LQSV specification is in accordance with the stylized fact cited above. However, the model does not exhibit long memory type behavior as the autocorrelations (appear to) decrease exponentially towards zero at long lags.

These expressions for the moments can be used to estimate the parameters of the LQSV model by generalized method of moments (GMM). However, I do not consider GMM estimation as it appears to suffer from poor finite sample performance (Jacquier et al., 1994; Andersen and Sørensen, 1996). Furthermore, moment based estimation methods do not directly provide an estimate of the volatility process and require an additional step such as re-projection (Gallant and Tauchen, 1998). For these reasons, I consider (approximate) maximum likelihood estimation (MLE) of the quadratic class stochastic volatility models instead.

3. Maximum likelihood estimation

This section discusses MLE of the log-quadratic stochastic volatility model with leverage. As is now well-known, the difficulty with MLE for stochastic volatility models is that the likelihood function cannot be obtained in closed form. I consider two methods to approximately evaluate the likelihood function: one based on numerical integration and one based on importance sampling (simulation).

To describe these algorithms, I first cast the model with leverage in state-space form with independent innovations. To this end, define $w_{t+1}=(v_{t+1}-\rho u_t)/\sqrt{1-\rho^2}$ and note that w_t and u_s are independent for any t, s. Then we can write the LQSV model as

$$y_t = \sigma_t u_t \tag{5a}$$

$$\sigma_t^2 = \exp(a_0 + a_1 x_t + a_2 x_t^2) \tag{5b}$$

$$x_{t+1} = \phi x_t + \rho\sqrt{1-\phi^2}\,\frac{y_t}{\sigma_t} + \sqrt{(1-\phi^2)(1-\rho^2)}\,w_{t+1}, \quad x_0 \sim N(0,1) \tag{5c}$$

$$\begin{bmatrix} u_t \\ w_{t+1} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) \tag{5d}$$

(5a) and (5b) are the measurement equations, while (5c) is the state equation. The parameter vector to be estimated is $\theta=(a_0,\, a_1,\, a_2,\, \phi,\, \rho)^\top$. Note that (5a) (5b) (5c) (5d) is not quite in standard state-space form because of the presence of the measurement y_t in the state Eq. (5c). However, as shown below, a simple adjustment to the filtering equations can accommodate this specification. Also note that the QML estimator as suggested in Harvey and Shephard (1996) for the LNAR specification is not readily available for the LQSV specification. The observation equation is non-linear in the state variable even after the square-log transformation because of the quadratic term in (5b).

3.1. Nonlinear filtering via numerical integration

To describe the filtering algorithm, let $p(\cdot)$ denote generic density functions and $y_{s:t}=\{y_s, y_{s+1}, \ldots, y_t\}$. Then the filtering recursions are given by

$$p(x_t|y_{1:t-1},\theta) = \int p(x_t|x_{t-1},y_{t-1},\theta)p(x_{t-1}|y_{1:t-1},\theta)\mathrm{d}x_{t-1} \quad \text{(prediction step)} \tag{6a}$$

$$c_t(\theta) \equiv p(y_t|y_{1:t-1},\theta) = \int p(y_t|x_t,\theta)p(x_t|y_{1:t-1},\theta)\mathrm{d}x_t \quad \text{(updating step)} \tag{6b}$$

$$p(x_t|y_{1:t},\theta) = \frac{1}{c_t(\theta)}p(y_t|x_t,\theta)p(x_t|y_{1:t-1},\theta) \tag{6c}$$

Note how the prediction step has been modified to account for the presence of the measurement y_t in the state equation. The prediction step for the standard state-space form is based on the Markov property $p(x_t|x_{t-1}, y_{1:t-1})=p(x_t|x_{t-1})$. However, for formulation (5a) (5b) (5c) (5d), we have $p(x_t|x_{t-1}, y_{1:t-1})=p(x_t|x_{t-1}, y_{t-1})$ instead. Recursion (6a) (6b) (6c) forms the basis of the filtering approach to likelihood evaluation.

From recursion (6a) (6b) (6c), the log-likelihood value for a given parameter vector θ can be computed as

$$\ell(\theta) = \sum_{t=1}^{T} \log c_t(\theta) \tag{7}$$

MLE of the parameters θ can be obtained by numerically maximizing (7). The difficulty, however, is that the prediction $p(x_t|y_{1:t-1})$ and filtered $p(x_t|y_{1:t})$ densities are not known.

The numerical integration approach approximates these unknown densities by a set of (deterministic) "particles" with associated weights (Kitagawa, 1987). The integrals in (6a) (6b) (6c) are then approximated by an appropriate weighted sum. I follow Fridman and Harris (1998) and use the Gauss-Legendre quadrature nodes and weights to discretize the state-space. To implement the filter, one then needs to choose the number of quadrature nodes m and an interval $[a, b]$ that covers the (univariate) state-space. As the unconditional state distribution is $N(0, 1)$, it is natural to choose the interval $[-a, a]$ where a is the number of standard deviations covered in the unconditional distribution.

The numerical integration approach to likelihood evaluation is straightforward to apply for the LQSV model with leverage (5a) (5b) (5c) (5d). The approximation error decreases with the number of nodes m used to evaluate the likelihood. However, the main problem with this approach is that the computational cost rises rapidly with m.

3.2. Importance sampling

An alternative approach to evaluating the likelihood of nonlinear state-space models is via importance sampling. The idea of this approach is to interpret the (intractable) integral as an expectation and to approximate it by an average using simulation. More specifically,

$$\ell(\theta) = \int \frac{p(y,x,\theta)}{q(x,y,\theta)}q(x,y,\theta)\mathrm{d}x = E\left[\frac{p(y,x,\theta)}{q(x,y,\theta)}\right] \approx \frac{1}{M}\sum_{j=1}^{M}\frac{p(y,x^{(j)},\theta)}{q(x^{(j)},y,\theta)}$$

where $y=(y_1, \ldots, y_T)$ and $x=(x_1, \ldots, x_T)$. The integral is a T-fold integral, the expectation $E[\cdot]$ is with respect to the importance density $q(x, y, \theta)$, and $x^{(j)}$ is a simulated draw from the importance density $q(x, y, \theta)$.

The key implementation issue for the importance sampler approach is the choice of the importance density $q(x, y, \theta)$. A poor choice of $q(\cdot)$ can yield a very "noisy" approximation to the desired integral. Here I consider two approaches to the choice of the importance density $q(\cdot)$ that use information in the sample y. Both of these approaches make a first pass through the data to find the (data dependent) parameters of a Gaussian sampler $q(\cdot)$. The efficient importance sampler of Liesenfeld and Richard (2003) fits a sampler observation by observation using simulated draws from a preliminary sampler. The algorithm in Liesenfeld and Richard (2003) described for the LNAR model without leverage can be easily adapted to evaluate the likelihood of the LQSV model with leverage based on the modified filtering recursions (6a) (6b) (6c).

An alternative sampler that is commonly used is that based on the posterior mode

$$q(\cdot) \sim N(\hat{x}, -H^{-1}), \qquad \hat{x} = \underset{x}{\operatorname{argmax}} \log p(y, x), \qquad H = \left. \frac{\partial^2 \log p(y, x)}{\partial x \partial x^\top} \right|_{x=\hat{x}}$$

For this sampler, one needs to obtain the $T \times 1$ mode $\hat{x}$via numerical optimization. Durham (2006) suggests using the quadratically convergent Newton's method. However, as noted by So (2003), for some specifications Newton's method may fail as $p(y, x)$ may not be globally concave. For such cases, one can use the quadratic hill-climbing algorithm as suggested by So (2003).[4]

3.3. One-step prediction density and diagnostics

A commonly used diagnostic check of the stochastic volatility specification is based on the one-step prediction density (Kim et al., 1998; Liesenfeld and Richard, 2003; Durham, 2006). The diagnostic is to check the probability integral transform

$$z_t = \Pr(Y_t \le y_t | Y_{1:t-1}) \tag{8}$$

Thus, z_t is just the cumulative distribution function corresponding to the one-step density $p(Y_t|Y_{1:t-1}, \theta)$ evaluated at the observations y_t (Diebold et al., 1998).[5] If the model is correctly specified, z_t should be independently and identically distributed uniformly on [0, 1]. Diebold et al. (1998) suggest using the histogram of z_t to check the uniform distribution assumption and to use the autocorrelation functions of powers of z_t to check the independence assumption.

To use this diagnostic, one needs to evaluate the cumulative distribution function of the one-step prediction density at each observation y_t. For the numerical integration approach, this can be done straightforwardly as part of likelihood evaluation. Noting that the updating step (6b) evaluates the one-step density, one merely replaces the conditional density $p(y_t|x_t, \theta)$ in the integrand with the corresponding cumulative distribution function. For the importance sampling approach, however, the one-step density is not used in likelihood evaluation. A common solution is to use an additional

[4] So (2003) considers cases where direct implementation of the quadratic hill-climbing algorithm is too costly and proposes a modified algorithm. However, for the class of stochastic volatility specifications considered in this paper, the Hessian H is tridiagonal and can be evaluated analytically. Thus, the maximum eigenvalue of H needed to calculate the diagonal offset can be efficiently computed.

[5] Kim et al. (1998) use the integral transform of the one-step density of y_t^2 instead which ignores information on the sign of y_t (Durham, 2006).

step of simulation based particle filtering to evaluate the one-step density or cumulative distribution function (Kim et al., 1998; Durham, 2006).[6]

Other statistics of interest from the one-step prediction density can be obtained in a similar fashion. For example, for value-at-risk applications, one requires the α-quantile from the one-step ahead prediction density. For the numerical integration approach, simply replace $p(y_t|x_t, \theta)$ in the integrand of (6b) with the corresponding quantile function. Alternatively, the quantile can be estimated from the simulated prediction density obtained from the particle filter.

3.4. Monte Carlo simulations

To assess the finite sample performance of the maximum likelihood estimators for the LQSV model, a small Monte Carlo simulation experiment was conducted. I simulated data from the LQSV specification (5a) (5b) (5c) (5d) with parameter values

$$a_0 = 0, \quad a_1 = 0.45, \quad a_2 = 0.1, \quad \phi = 0.975, \quad \rho = \{0, -0.3, -0.6\}$$

For $a_2=0$, these parameter values correspond to those typically used in Monte Carlo simulations of LNAR models. For $a_2=0.1$, the fourth moment of y_t exists with kurtosis $K=4.7245$, compared to $K=3.6734$ for the LNAR model with $a_2=0$. For each Monte Carlo replication, I simulated the process with presample value $x_0=0$ for $T=1500$ and discarded the first 500 observations to remove the dependency on x_0. For the retained sample of size $T=1000$, maximum likelihood estimates were obtained by numerically maximizing the log-likelihood value evaluated using the three methods discussed above. In all cases, I use the same starting parameter values

$$a_0 = 0, \quad a_1 = 0.45, \quad a_2 = 0.01, \quad \phi = 0.95, \quad \rho = 0$$

The optimizer achieved convergence in all cases for the numerical integration approach based on Gauss-Legendre nodes. However, for the importance sampler based estimators, a number of cases resulted in convergence failure. An examination of the failed cases suggests that the performance of the importance sampler based estimators is sensitive to the choice of starting values. For the efficient importance sampler of Liesenfeld and Richard (2003), the iterations to update the sampler occasionally do not appear to converge and results in a negative variance for the sampler. For the posterior mode based sampler, the quadratic hill-climbing algorithm occasionally has difficulty finding the mode due to a non-definite Hessian. Though computationally most costly, I have found the numerical integration approach based on the Gauss-Legendre quadrature to be numerically most reliable for estimation of the LQSV model.

The Monte Carlo simulation results for 100 replications are presented in Table 1. Table 1 shows that the MLE performs reasonably well in finite samples. The leverage parameter ρ appears to be slightly downward biased but the persistence parameter ϕ is estimated quite precisely. Note, in particular, that the parameter estimates get more precise with the (absolute) size of the leverage effect ρ. This is as expected since the information contained in the

[6] For an alternative approach, see Liesenfeld and Richard (2003) who use a sequence of importance samplers to evaluate the one-step density.

Table 1
Finite sample performance of MLE based on numerical integration

	a_0	a_1	a_2	ϕ	ρ
DGP	0.0000	0.4500	0.1000	0.9750	0.0000
Mean	−0.0167	0.4312	0.1156	0.9732	0.0213
(RMSE)	(0.1550)	(0.1261)	(0.0907)	(0.0313)	(0.2538)
DGP	0.0000	0.4500	0.1000	0.9750	−0.3000
Mean	−0.0189	0.4362	0.1150	0.9753	−0.3355
(RMSE)	(0.1487)	(0.1136)	(0.0846)	(0.0126)	(0.1991)
DGP	0.0000	0.4500	0.1000	0.9750	−0.6000
Mean	−0.0189	0.4485	0.1147	0.9749	−0.6356
(RMSE)	(0.1299)	(0.0932)	(0.0732)	(0.0121)	(0.1439)

Reported are means and root mean squared errors (RMSE) of MLE for 100 Monte Carlo replications of sample size $T=1000$. DGP is the parameters of the data generating process. MLE is based on Gauss–Legendre quadrature with $m=100$ nodes over the interval $[-3, 3]$.

measurement y regarding the unobserved state x increases with the correlation between the two innovations.

4. Empirical application

In this section, I illustrate the empirical fit of the proposed log quadratic stochastic volatility specification (5a) (5b) (5c) (5d) where y_t is the observed returns of an asset. I fit the model to the daily returns from the S&P 500 index for two samples: January 1980 to December 1987 for 2022 observations and July 1995 to July 2005 for 2538 observations.[7] The daily returns computed as (100 times) the first difference of the log price index are plotted in Fig. 2 for the two samples. As can be seen in Fig. 2, the first sample (January 1980 to December 1987) contains an extreme downward spike at the stock market crash (Black Monday) in October 1987. This sample has become somewhat of a benchmark as several authors (Fridman and Harris, 1998; Liesenfeld and Jung, 2000; Jacquier et al., 2004; Yu, 2005) have used it to fit a number of stochastic volatility models. The more recent sample (July 1995 to July 2005) does not contain as extreme a spike as the first sample and can be considered a more “representative” sample.

In order to assess the fit of the log quadratic stochastic volatility (LQSV) specification, I compare its fit to two alternative commonly used specifications. The first is the baseline LNAR model with $a_2=0$ in (5b). The second is also a linear model with $a_2=0$, except that the mean equation innovation u_t in (5d) is assumed to have a Student-t distribution with $\nu>2$ degrees of freedom and independent of w_s. I denote this model as LTAR. Note that both LTAR and LQSV have five parameters to estimate and model comparison based on the information criterion (such as Akaike) is equivalent to comparing the (maximized) log-likelihood values.

Table 2 reports ML estimates for the three models based on the Gauss-Legendre quadrature with $m=300$ nodes. As the first sample (January 1980 to December 1987) contains an extreme spike at Black Monday, the nodes are chosen to cover the range $[-9, 9]$ which is nine standard deviations of the unconditional state distribution. For the second sample (July 1995 to July 2005), the nodes are chosen to cover the range $[-7, 7]$ as there are no obvious extreme spikes as in the

[7] The price index data were treated as equally spaced daily observations after dropping missing observations due to weekends and holidays.

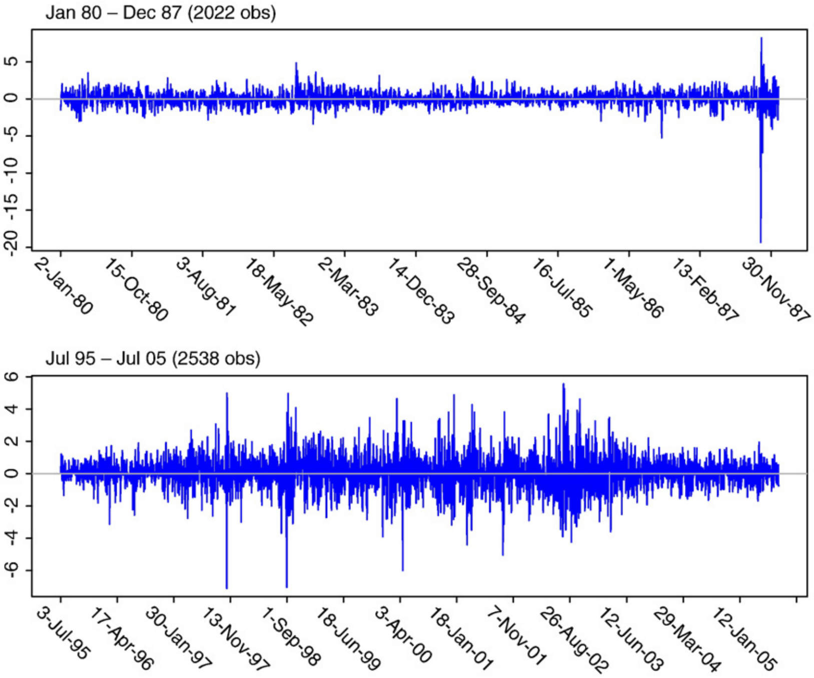

Fig. 2. Daily returns from the S&P 500 index for two samples. The returns were computed as 100 times the first difference in the log price index, removing any missing values due to weekends and holidays.

first sample.[8] The QML standard errors are computed from the QML covariance matrix

$$\frac{1}{T}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2\ell_t}{\partial\theta\partial\theta^\top}\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial\ell_t}{\partial\theta}\frac{\partial\ell_t}{\partial\theta^\top}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2\ell_t}{\partial\theta\partial\theta^\top}\right)^{-1}$$

where the derivatives are evaluated numerically. I make the following observations.

First, as is commonly found for stock index returns, the leverage parameter ρ is statistically significantly negative for all specifications. Moreover, the magnitude of the leverage effect is considerably larger (and much more precisely estimated) in the second sample than in the first sample. These results reinforce the importance of not constraining $\rho=0$ when fitting stochastic volatility models to stock index returns. The dynamics of the volatility process is also highly persistent for all specifications with ϕ near unity.[9] The location parameter a_0 is not significant in all models and samples, while the scale parameter a_1 is statistically significant in all models and of similar magnitude.

[8] For the first sample, a crude grid search was done with $m=100$ nodes covering five up to nine standard deviations of the unconditional state distribution. While the maximized log-likelihood values vary somewhat over these ranges, the parameter estimates do not change much. For the second sample, estimates for nodes that cover $[-7, 7]$ and $[-9, 9]$ were virtually the same.

[9] The parameter estimates for the LNAR model in the first sample are fairly close to those reported by Yu (2005) using Markov Chain Monte Carlo for the same data set. For the parameterization used in this paper, the posterior means reported by Yu (2005) are $a_0=-0.0688$, $a_1=0.6362$, $\phi=0.972$, $\rho=-0.3179$.

Table 2
QML based on Gauss–Legendre quadrature with $m=300$ nodes over the interval $[-9, 9]$ for the sample January 1980–December 1987 and over the interval $[-7, 7]$ for the sample July 1995–July 2005

	January 80–December 87 (2022 obs)			July 1995–July 2005 (2538 obs)		
	LNAR	LTAR	LQSV	LNAR	LTAR	LQSV
a_0	−0.0817	0.0250	−0.2191	0.1035	0.1159	0.0347
	(0.1175)	(0.1544)	(0.1249)	(0.1052)	(0.1032)	(0.1120)
a_1	0.6353*	0.5751*	0.5959*	0.7788*	0.7728*	0.8044*
	(0.1099)	(0.0872)	(0.1067)	(0.0643)	(0.0630)	(0.0713)
a_2			0.1426*			0.0677*
			(0.0613)			(0.0336)
ϕ	0.9680*	0.9858*	0.9720*	0.9794*	0.9809*	0.9799*
	(0.0142)	(0.0071)	(0.0117)	(0.0049)	(0.0044)	(0.0048)
ρ	−0.3346*	−0.4040*	−0.3390*	−0.7701*	−0.7906*	−0.7713*
	(0.0857)	(0.1032)	(0.0892)	(0.0421)	(0.0356)	(0.0420)
v		10.0768*			31.9857	
		(3.5392)			(22.7089)	
ℓ	−2775.69	−2764.58	−2762.62	−3601.86	−3600.46	−3599.51
κ	4.4918	5.5505	10.4040	5.5019	5.8408	8.4746

LNAR is the log -normal AR(1) model, LTAR is the log AR(1) model with Student-t distribution with v degrees of freedom, and LQSV is the log quadratic stochastic volatility model. Numbers in parentheses are QML standard errors and * indicates significance at (two-sided) size 0.05. ℓ is the maximized log-likelihood value and κ is the model implied kurtosis of the returns.

Second, for the LQSV specification, the coefficient on the quadratic term a_2 is statistically significantly positive (at size 0.05) for both samples based on the (robust QML) Wald statistic. The likelihood ratio statistic for the hypothesis $a_2=0$ is LR=26.1462 (p-value 0.0000) for the first sample and LR=4.7115 (p-value 0.0300) for the second sample. Thus, both tests reject the LNAR specification in favor of the LSQV specification, though the evidence in the second sample is weaker (marginally significant) than that in the first sample.

Third, the degrees of freedom parameter v for the Student-t specification LTAR is statistically significant in the first sample but not in the second sample. Moreover, the point estimate $\hat{v}$ from the second sample is much larger than that from the first sample and the implied distribution is much closer to the Gaussian in the second sample. Note also that the parameter estimates in the second sample for LNAR and LTAR are quite similar. Therefore, compared to the LNAR specification, the LTAR specification provides a better fit only in the first sample. Comparing the

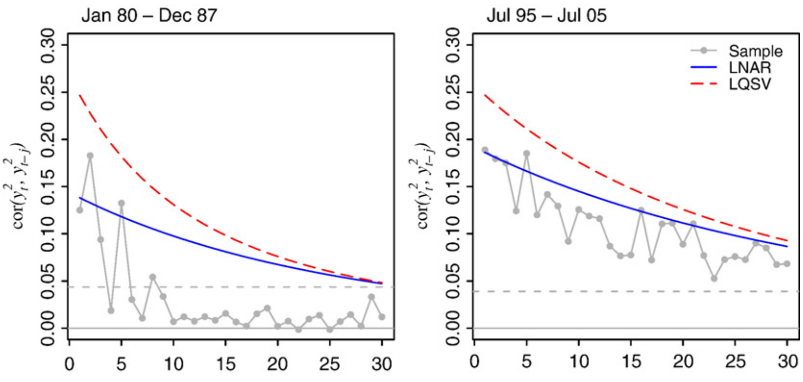

Fig. 3. Sample and model implied autocorrelations of squared returns up to 30 lags.

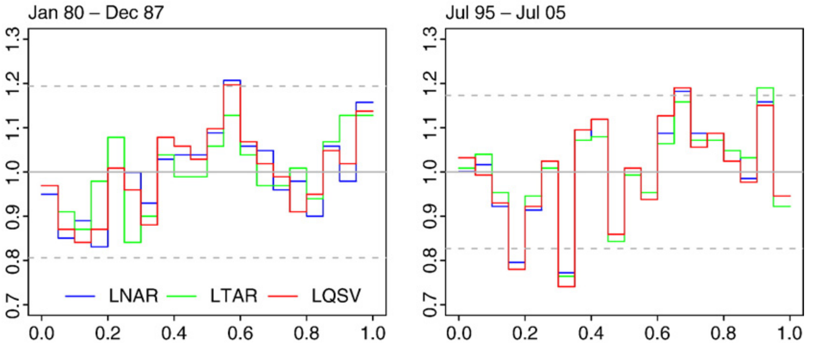

Fig. 4. Histogram of z_t, the probability integral transform of y_t with respect to the one-step ahead density forecasts from three stochastic volatility specifications. The dashed lines indicate the approximate 95% confidence interval under the iid assumption for 20 bins. The left panel is for the January 1980–December 1987 sample (2022 observations) and the right panel is for the July 1995–July 2005 sample (2538 observations).

LTAR and LQSV specifications, we see that the maximized log-likelihood values are larger for the LQSV specification in both samples, though the difference is smaller in the second sample. Thus, model selection based on the information criterion will favor the LQSV specification over the LTAR specification for both samples.

To understand the implications of these estimates, note that the estimated value of the parameter a_2 suggests that the moments exist up to order 7 in the first sample and up to order 14 in the second sample for the LQSV specification. The implied kurtosis (4) of the observed returns for the three specifications are reported in Table 2. The sample kurtosis is 51.6806 for January 1980–December 1987 and 6.0299 for July 1995–July 2005. For the first sample, though the implied kurtosis from the LQSV specification is larger by a factor of about two than that from the other two specifications, it falls far short of the extremely large sample kurtosis due to Black Monday. For the second sample, the implied kurtosis from the LQSV specification is again larger than that from the other two specifications and is larger than the sample kurtosis. The implied skewness of the returns in all specifications is zero. The sample skewness is −2.5550 (p-value 0.000) in the first sample and −0.1036 (p-value 0.033) in the second sample. Fig. 3 plots the sample autocorrelations of the squared returns y_t^2 together with the implied autocorrelations from the LNAR and LQSV models derived in Section 2.2.[10] For both samples, the implied autocorrelations of the squared series from the LQSV model are larger than that from the LNAR model, especially for low order lags.

Figs. 4 and 5 are diagnostics based on the probability integral transform z_t of y_t with respect to the one-step ahead density discussed in Section 3.3. Fig. 4 plots the histogram of z_t which should be uniform on [0, 1] under a correctly specified model. All specifications in both samples appear to pass this diagnostic, though there is some indication of departure from uniformity in the second sample. Fig. 5 plots the autocorrelation function of powers of demeaned z_t to check the independence assumption.[11] The autocorrelations of powers of z_t from the three specifications are nearly identical and do not indicate any obvious departure from the independence assumption. The dependence up to fourth moments appears to be well captured by these specifications.

[10] I am not able to derive the implied autocorrelations of y_t^2 from the LTAR model with leverage ($\rho \neq 0$).

[11] The autocorrelations for the July 1995–July 2005 sample are quite similar to Fig. 5 and are omitted to save space.

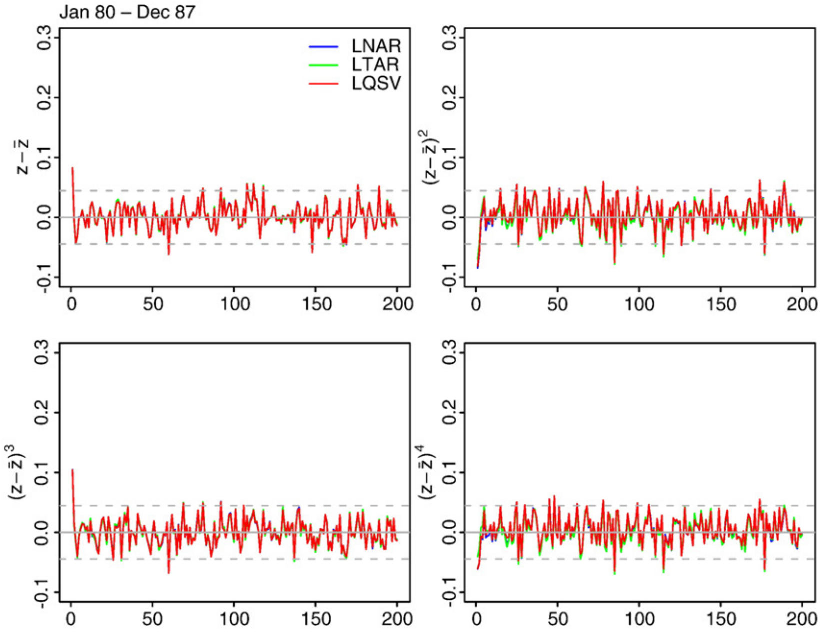

Fig. 5. The autocorrelation functions of powers of z_t, the probability integral transform of y_t with respect to the one-step ahead density forecasts from three stochastic volatility specifications. The dashed lines indicate the approximate 95% confidence interval under the iid assumption.

One of the important uses of volatility models is to measure value-at-risk (VaR). Engle and Manganelli (2004) have proposed tests to check the quality of the estimated VaR based on the idea that the "hit" sequence from a valid VaR should be independent and identically distributed. More specifically, define the hit sequence as

$$h_t(\alpha) = I(y_t < q_t(\alpha)) - \alpha$$

where $I(\cdot)$ is the indicator function and $q_t(\alpha)$ is the predicted α-quantile. A valid VaR $q_t(\alpha)$ should produce a hit sequence $h_t(\alpha)$ with mean zero and orthogonal to the period $t-1$ information set.

Table 3
Dynamic quantile test

	January 1980–December 1987 (2022 obs)			July 1995–July 2005 (2538 obs)		
	LNAR	LTAR	LQSV	LNAR	LTAR	LQSV
$\alpha=0.05$						
$\chi^2(7)$	7.410	21.151*	3.674	9.025	9.376	9.661
[p-value]	[0.388]	[0.004]	[0.816]	[0.251]	[0.227]	[0.209]
$\alpha=0.01$						
$\chi^2(7)$	13.023	25.498*	5.928	13.003	13.130	12.572
[p-value]	[0.072]	[0.001]	[0.548]	[0.072]	[0.069]	[0.083]

The α-quantiles are obtained from the one-step ahead prediction densities for three stochastic volatility specifications, corresponding to a 1-day value-at-risk. The $\chi^2(7)$ statistics are obtained from the auxiliary regression of the "hit" series on a constant, five lags of the hit series and the VaR series.

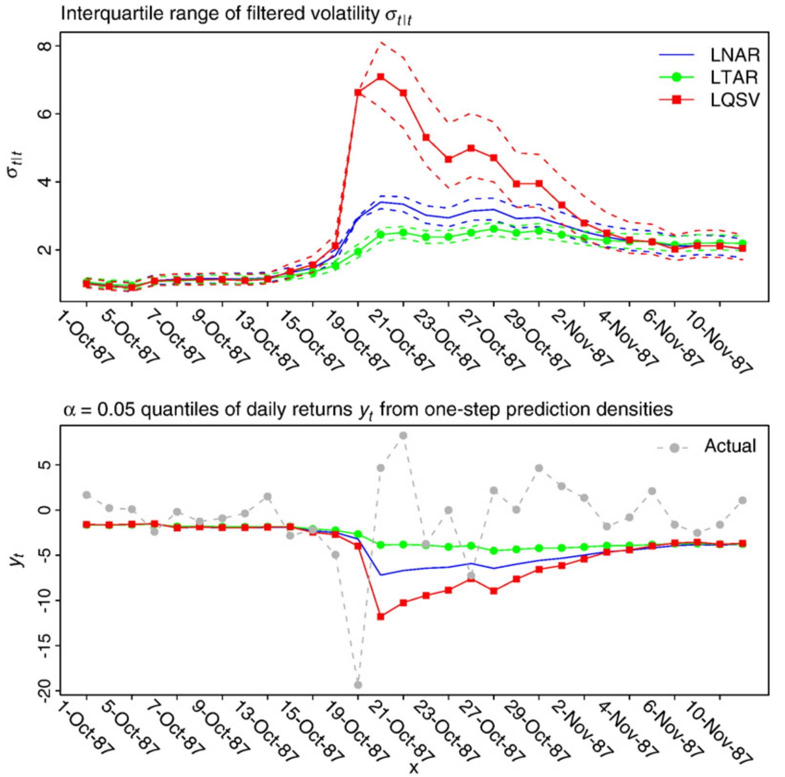

Fig. 6. The top panel shows the median and interquartile range of the filtered volatility from three stochastic volatility specifications around the 19 October 1987 (black Monday) market crash. The bottom panel shows the daily $\alpha=0.05$ value-at-risk based on the one-step prediction densities of the three specifications.

This can be tested with an artificial regression of $h_t(\alpha)$ on a set of regressors X_t in the $t-1$ information set (including a constant term) whose coefficients should all be zero under a valid VaR (Giot and Laurent, 2004).

Table 3 reports the results for testing the validity of 1-day ahead VaRs produced from the three estimated stochastic volatility models. The 1-day VaRs are obtained from the one-step ahead prediction density of y_t as discussed in Section 3.3. Following Giot and Laurent (2004), I use $X_t=(1, h_{t-1}, \ldots, h_{t-5}, q_t)$ as regressors in the artificial regression. The results from the three models are quite similar for the July 1995–July 2005 sample. However, for the January 1980–December 1987 sample, only the VaRs from the LQSV model pass the diagnostic check at both quantiles. I must note, however, that these tests are based on in-sample VaRs that ignore coefficient uncertainty in the estimated VaRs. Fig. 6 displays the filtered volatility estimates and the daily $\alpha=0.05$ VaRs from the three models about the 19 October 1987 market crash.

5. Concluding remarks

I have proposed specification and estimation of a class of discrete time log quadratic stochastic volatility models that nest the commonly used LNAR models. The empirical relevance of this class of specification was illustrated by fitting it to the daily S&P 500 index returns for two sample periods. While the implied moments from the LQSV model do not quite match the sample moments, statistical tests suggest that the LQSV specification might provide a better fit than the commonly used log AR(1) model with Gaussian or Student-t mean equation innovations.

I conclude the paper by suggesting a few areas for further research. Maximum likelihood estimation of stochastic volatility models remains a difficult problem. The numerical integration based method used in this paper, while shown to be feasible and reliable, has the main shortcoming that it can be quite time consuming when fit to a long time series. Given the empirical success of the LQSV specification, it is worthwhile to examine alternative estimation methods. In particular, simulated maximum likelihood based on importance sampling appears to be promising in terms of computational cost. However, as discussed in Section 3.4, the importance samplers that perform well for the LNAR model appear to encounter difficulties when used for estimation of the LQSV model. A further examination of this issue, including development of alternative or modified samplers, would be worthwhile.

The paper has compared the fit of the LQSV specification with two commonly used alternative specifications: the log AR(1) specification with Gaussian and Student-t mean equation innovations. It would be interesting to examine how the quadratic class compares with alternative models such as those with jumps or multifactors. This may require moving from the discrete time formulation used in the paper to a continuous time formulation. Once formulated in continuous time, one can also discuss and compare option pricing based on the alternative stochastic volatility specifications.

Acknowledgement

I thank a referee for pointing out an error in the revised manuscript.

Appendix A

The derivation of the moments uses the following result repeatedly. If $x \sim N(0, v)$, then

$$\begin{aligned}
E\left[x^r e^{a_0+a_1x+a_2x^2}\right] &= \int x^r e^{a_0+a_1x+a_2x^2}\frac{1}{\sqrt{2\pi v}}e^{-\frac{x^2}{2v}}\mathrm{d}x \\
&= \frac{1}{\sqrt{1-2a_2v}}\exp\left(a_0+\frac{a_1^2v}{2-4a_2v}\right)\int x^r\frac{1}{\sqrt{2\pi v_y}}e^{-\frac{(x-\mu_y)^2}{2v_y}}\mathrm{d}x \\
&= \frac{1}{\sqrt{1-2a_2v}}\exp\left(a_0+\frac{a_1^2v}{2-4a_2v}\right)E[y^r] \qquad (9)
\end{aligned}$$

where $y \sim N(\mu_y, v_y)$ with $\mu_y = a_1v/(1-2a_2v)$ and $v_y = v/(1-2a_2v)$. This expectation exists provided $2a_2v < 1$.

A.1. Moments of the LQSV model

As $u_t \sim N(0, 1)$, the moments of u_t are given by

$$E\left[u_t^r\right] = \begin{cases} 0 & \text{for } r \text{ odd} \\ \dfrac{r!}{2^{\frac{r}{2}}\left(\frac{r}{2}\right)!} & \text{for } r \text{ even} \end{cases}$$

To evaluate the expectation of $\sigma_t^r = e^{\frac{r}{2}\left(a_0+a_1x+a_2x_t^2\right)}$, note that $x_t \sim N(0, 1)$ and compute the integral

$$E\left[\sigma_t^r\right] = \int e^{\frac{r}{2}\left(a_0+a_1x+a_2x^2\right)}\phi(x)dx = \frac{1}{\sqrt{1-ra_2}}\exp\left(\frac{ra_0}{2}+\frac{r^2a_1^2}{8(1-ra_2)}\right)$$

for $a_2 < 1/r$.

A.2. Autocovariances

Let $w_{t+1} = \frac{1}{\sqrt{1-\rho^2}}(v_{t+1}-\rho u_t)$ and note that $w_t \sim N(0, 1)$ and $E[w_t u_s]=0$ for $\forall t, s$. Then rewrite the state equation for x_t as

$$\begin{aligned} x_t &= \phi x_{t-1} + \rho\sqrt{1-\phi^2}u_{t-1} + \sqrt{1-\phi^2}\sqrt{1-\rho^2}w_t \\ &= \phi^j x_{t-j} + a_u(u_{t-1} + \phi u_{t-2} + \ldots + \phi^{j-1}u_{t-j}) + a_w(w_t + \phi w_{t-1} \ldots + \phi^{j-1}w_{t-j+1}) \\ &= \phi^j x_{t-j} + a_u\phi^{j-1}u_{t-j} + a_u z_u + a_w z_w \end{aligned}$$

where $a_u \equiv \rho\sqrt{1-\phi^2}, a_w \equiv \sqrt{1-\phi^2}\sqrt{1-\rho^2}$,

$$\begin{aligned} z_u &\equiv \sum_{i=1}^{j-1}\phi^{i-1}u_{t-1}, \quad z_u \sim N(0, s_u^2), \quad s_u^2 = \frac{1-\phi^{2(j-1)}}{1-\phi^2} \\ z_w &\equiv \sum_{i=0}^{j-1}\phi^{i}w_{t-1}, \quad z_u \sim N(0, s_w^2), \quad s_w^2 = \frac{1-\phi^{2j}}{1-\phi^2} \end{aligned}$$

Then

$$\begin{aligned} E[y_t^2 y_{t-j}^2] &= E[u_t^2\sigma_t^2 u_{t-j}^2\sigma_{t-j}^2] = E[u_t^2]E[\sigma_t^2 u_{t-j}^2\sigma_{t-j}^2] = E[\sigma_t^2 u_{t-j}^2\sigma_{t-j}^2] \\ &= E[E_{t-j}[\sigma_t^2]u_{t-j}^2\sigma_{t-j}^2] \end{aligned} \tag{10}$$

where $E_{t-j}[\sigma_t^2]$ is the conditional expectation based on information up to $t-j$. I evaluate this conditional expectation by a further use of the law of iterated expectations

$$E_{t-j}[\sigma_t^2] = E_{t-j}\left[E_{t-j}[\sigma_t^2|z_w]\right]$$

where the outer expectation is with respect to z_w and the inner expectation with respect to z_u. By repeatedly applying (9), we can write

$$E_{t-j}\left[\sigma_t^2\right] = \frac{1}{\sqrt{c_1}}e^{a_0+a_1^2K+a_1(1+4a_2K)y_{t-j}+a_2(1+4a_2K)y_{t-j}^2}$$

where

$$\begin{aligned} c_1 &= 1-2a_2\rho^2(1-\phi^{2(j-1)})-2a_2(1-\rho^2)(1-\phi^{2j}) \\ K &= \frac{c_1\rho^2(1-\phi^{2(j-1)}) + (1-\rho^2)(1-\phi^{2j})}{2c_1(1-2a_2\rho^2(1-\phi^{2(j-1)}))} \\ y_{t-j} &= \phi^j x_{t-j} + a_u\phi^{j-1}u_{t-j} \end{aligned}$$

From (10) and further applications of (9) the required moment can be written as

$$
\begin{aligned}
E\left[y_t^2 y_{t-j}^2\right] &= \frac{1}{\sqrt{c_1}} E\left[u_{t-j}^2 \sigma_{t-j}^2 e^{a_0 + a_1^2 K + a_1(1+4a_2K)y_{t-j} + a_2(1+4a_2K)y_{t-j}^2}\right] \\
&= \frac{1}{\sqrt{c_1}} E\left[E[u_{t-j}^2 \sigma_{t-j}^2 e^{a_0 + a_1^2 K + a_1(1+4a_2K)y_{t-j} + a_2(1+4a_2K)y_{t-j}^2} | x_{t-j}]\right] \\
&= \frac{1}{(c_1)^{1/2}(c_2)^{5/2}} \left(\frac{c_2 + a_1^2 c_3}{(c_4)^{1/2}} + \frac{4a_1 a_2 c_3 c_5 \phi^j}{(c_4)^{3/2}} + \frac{4a_2^2 c_3 (c_4 + c_5^2)\phi^{2j}}{(c_4)^{5/2}}\right) \\
&\quad \times \exp\left(c_6 + \frac{c_5^2}{2c_4}\right)
\end{aligned}
$$

where the outer expectation is with respect to x_{t-j} and the inner expectation is with respect to u_{t-j} and

$$
\begin{aligned}
c_2 &= 1 - \frac{2a_2\rho^2}{c_1}\left(1-\phi^2\right)\phi^{2(j-1)}, \quad c_3 = \frac{\rho^2}{c_1^2}\left(1-\phi^2\right)\phi^{2(j-1)} \\
c_4 &= 1 - 2a_2\left(1 + \frac{\phi^{2j}}{c_1 c_2}\right), \quad c_5 = a_1\left(1 + \frac{\phi^j}{c_1 c_2}\right) \\
c_6 &= 2a_0 + \frac{a_1^2}{2c_1}\left(\frac{c_1\rho^2(1-\phi^{2(j-1)}) + (1-\rho^2)(1-\phi^{2j})}{1-2a_2\rho^2(1-\phi^{2(j-1)})} + \frac{\rho^2(1-\phi^2)\phi^{2(j-1)}}{c_1 c_2}\right)
\end{aligned}
$$

Note that as $j \to \infty$, $c_1 \to 1-2a_2$, $c_2 \to 1$, $c_3 \to 0$, $c_4 \to 1-2a_2$, $c_5 \to a_1$, $c_6 \to 2a_0 + \frac{a_1^2}{2(1-2a_2)}$ and

$$
E\left[y_t^2 y_{t-j}^2\right] \to \frac{1}{1-2a_2} e^{2a_0 + \frac{a_1^2}{1-2a_2}} = \left(E\left[y_t^2\right]\right)^2
$$

and $\operatorname{cor}(y_t^2, y_{t-j}^2) \to 0$.

The expression for $E[y_t^2 y_{t-j}^2]$ simplifies considerably for the canonical log-normal autoregressive (LNAR) model with $a_2=0$. For this case, $c_1 = c_2 = c_4 = 1$, $c_3 = \rho^2\left(1-\phi^2\right)\phi^{2(j-1)}$, $c_5 = a_1 (1+\phi^j)$, $c_6 = 2a_0 + \frac{a_1^2}{2}\left(1-\phi^{2j}\right)$ and

$$
E[y_t^2 y_{t-j}^2] = (1 + a_1^2\rho^2(1-\phi^2)\phi^{2(j-1)})e^{2a_0 + a_1^2(1+\phi^j)}
$$

This expression agrees with the expression given in Jacquier et al. (1994) for $\rho=0$.

References

Dong -Hyun, Ahn, Robert F., Dittmar, Gallant, Ronald A., 2002. Quadratic term structure models: theory and evidence. Review of Financial Studies 15, 243–288.

Andersen, Torben G., Sørensen, Bent E., 1996. GMM estimation of a stochastic volatility model: a Monte Carlo study. Journal of Business and Economic Statistics 14, 328–352.

Andersen, Torben G., Benzoni, Luca, Lund, Jesper, 2002. An empirical investigation of continuous-time equity return models. Journal of Finance 57, 1239–1284.

Xuezheng, Bai, Russell, Jeffrey R., Tiao, George C., 2003. Kurtosis of GARCH and stochastic volatility models with non-normal innovations. Journal of Econometrics 114, 349–360.

Bakshi, Gurdip, Cao, Charles, Chen, Zhiwu, 1997. Empirical performance of alternative option pricing models. Journal of Finance 52, 2003–2049.

Cheng, Peng, Olivier Scaillet. Linear-quadratic jump-diffusion modeling with application to stochastic volatility. 2004. Unpublished.

Chernov, Mikhail, Gallant, A. Ronald, Ghysels, Eric, Tauchen, George, 2003. Alternative models for stock price dynamics. Journal of Econometrics 116, 225–257.

Diebold, Francis X., Gunther, Todd A., Tay, Anthony S., 1998. Evaluating density forecasts with applications to financial risk management. International Economic Review 39, 863–883.

Ding, Zhuanxin, Granger, Clive W.J., 1996. Modeling volatility persistence of speculative returns: a new approach. Journal of Econometrics 73, 185–215.

Durham, Garland B., 2006. Monte Carlo methods for estimating, smoothing, and filtering one- and two-factor stochastic volatility models. Journal of Econometrics 133, 273–305.

Engle, Robert F., Manganelli, Simone, 2004. CAViaR: conditional autoregressive value at risk by regression quantiles. Journal of Business and Economic Statistics 22, 367–381.

Eraker, Bjørn, Johannes, Michael, Polson, Nicholas, 2003. The impact of jumps in volatility and returns. Journal of Finance 58, 1269–1300.

Fridman, Moshe, Harris, Lawrence, 1998. A maximum likelihood approach for non-Gaussian stochastic volatility models. Journal of Business and Economic Statistics 16, 284–291.

Gallant, A. Ronald, Tauchen, George, 1998. Reprojecting partially observed systems with application to interest rate diffusions. Journal of American Statistical Association 93, 10–24.

Giot, Pierre, Laurent, Sébastien, 2004. Modelling daily value-at-risk using realized volatility and ARCH type models. Journal of Empirical Finance 11, 379–398.

Harvey, Andrew C., Shephard, Neil, 1996. Estimation of an asymmetric stochastic volatility model for asset returns. Journal of Business and Economic Statistics 14, 429–434.

Jacquier, Eric, Polson, Nicholas G., Rossi, Peter E., 1994. Bayesian analysis of stochastic volatility models. Journal of Business and Economic Statistics 12, 371–417.

Jacquier, Eric, Polson, Nicholas G., Rossi, Peter E., 2004. Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. Journal of Econometrics 122, 185–212.

Kim, Sangjoon, Shephard, Neil, Chib, Siddhartha, 1998. Stochastic volatility: likelihood inference and comparison with ARCH models. Review of Economic Studies 65, 361–393.

Kitagawa, Genshiro, 1987. Non-Gaussian state-space modeling of nonstationary time series. Journal of American Statistical Association 82, 1032–1041.

Leippold, Markus, Wu, Liuren, 2002. Asset pricing under the quadratic class. Journal of Financial and Quantitative Analysis 37, 271–295.

Leippold, Markus, Wu, Liuren, 2003. Design and estimation of quadratic term structure models. European Finance Review 7, 47–73.

Liesenfeld, Roman, Richard, Jean -François, 2003. Univariate and multivariate stochastic volatility models: estimation and diagnostics. Journal of Empirical Finance 10, 505–531.

Liesenfeld, Roman, Jung, Robert C., 2000. Stochastic volatility models: conditional normality versus heavy-tailed distributions. Journal of Applied Econometrics 15, 137–160.

Meddahi, Nour. An eigenfunction approach for volatility modeling. 2001. unpublished.

Newey, Whitney K., Steigerwald, Douglas G., 1997. Asymptotic bias for quasi-maximum-likelihood estimators in conditional heteroskedasticity models. Econometrica 65, 587–599.

So, Mike K.P., 2003. Posterior mode estimation for nonlinear and non-Gaussian state space models. Statistica Sinica 13, 255–274.

White, Halbert, 1994. Estimation, Inference and Specification Analysis. Cambridge University Press.

Yu, Jun, 2005. On leverage in a stochastic volatility model. Journal of Econometrics 127, 165–178.

Yu, Jun, Zhenlin Yang, Xibin Zhang. A class of nonlinear stochastic volatility models and its implications on pricing currency options. 2002. unpublished.