Aust. N. Z. J. Stat. **44**(3), 2002, 319–335

ESTIMATION OF THE STOCHASTIC VOLATILITY MODEL BY THE EMPIRICAL CHARACTERISTIC FUNCTION METHOD

JOHN L. KNIGHT[1], STEPHEN E. SATCHELL[2] AND JUN YU[2*]

University of Western Ontario,
University of Cambridge and University of Auckland

Summary

The stochastic volatility model has no closed form for its likelihood and hence the maximum likelihood estimation method is difficult to implement. However, it can be shown that the model has a known characteristic function. As a consequence, the model is estimable via the empirical characteristic function. In this paper, the characteristic function of the model is derived and the estimation procedure is discussed. An application is considered for daily returns of Australian/New Zealand dollar exchange rate. Model checking suggests that the stochastic volatility model together with the empirical characteristic function estimates fit the data well.

Key words: empirical characteristic function; financial data; stochastic volatility.

1. Introduction

Modelling the volatility of financial time series has attracted a lot of attention since the introduction of autoregressive conditional heteroscedasticity (ARCH) (Engle, 1982). A feature of the ARCH-type model is that the conditional variance is driven by the past variables. An alternative setup to the ARCH-type model, the stochastic volatility (SV) model, provides a more realistic modelling of financial time series since it essentially involves two noise processes. This added dimension makes the model more flexible (see e.g. Kim, Shephard & Chib (1998) for the comparative advantages of the SV model over the ARCH-type models). For further discussion of the SV model, we refer to Ghysels, Harvey & Renault (1996). Unfortunately, the density function for the SV model has no closed form and hence neither does the likelihood function, even for the simplest version of the SV model; as a consequence, direct maximum-likelihood estimation is very difficult to implement. Therefore, other estimation methods have been proposed for estimating SV models.

Received December 2000; revised July 2001; accepted November 2001.

* Author to whom correspondence should be addressed.
[1] Dept of Economics, University of Western Ontario, London, Ontario, Canada, N6A 5C2.
[2] Trinity College & University of Cambridge, Sidgewick Avenue, Cambridge, United Kingdom CB1 9DE.
[3] Dept of Economics, University of Auckland, Private Bag 92019, Auckland, New Zealand.
e-mail: j.yu@auckland.ac.nz

Acknowledgments. This research was started while Knight and Yu were visiting Birkbeck College, University of London in 1997 and the final version was completed when Yu was visiting National University of Singapore in 2001. The authors thank participants in the Applied Econometrics Workshop at the University of Cambridge, the Canadian Econometrics Study Group meeting in London, the New Zealand Econometrics Study Group meeting in Hamilton, and the Far Eastern meeting of Econometric Society in Singapore for useful comments and discussion, especially R. Garcia, A. Harvey, A. Pagan, D. Robinson, R. Watson (the Technical Editor), the associate editor and an anonymous referee. R. Watson, in particular, showed us how to considerably abbreviate our proof of Propositions 3.1 and 4.1. Also the first author acknowledges financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the third author acknowledges financial support from the Royal Society of New Zealand Marsden Fund under No. 01-UOA-015.

Melino & Turnbull (1990) use the generalized method of moments (GMM) for the SV model. Andersen & Sorensen (1996) propose a more efficient GMM. For the continuous time SV model, a GMM approach is developed by Hansen & Scheinkman (1995). The idea is to match a finite number of sample moments and theoretical moments. Alternatively, the quasi maximum likelihood (QML) approach is suggested by Nelson (1988), Ruiz (1994) and Harvey, Ruiz & Shephard (1994). The main idea is to treat non-normal disturbances as if they are normal and then maximize the quasi-likelihood function. Often estimation methods involve the whole family of simulation-based methods, such as the simulated maximum likelihood proposed by Danielsson (1994), and Markov chain Monte Carlo (MCMC) proposed by Jacquier, Polson & Rossi (1994) and Kim *et al.* (1998), and the simulation method using important sampling and antithetic variables proposed by Durbin & Koopman (2000). The SV model has become a central model for describing financial time series and for comparing the relative merits of estimation procedures.

Although some of above-mentioned methods are consistent under appropriate regularity conditions, in general they are not efficient. For example, when there are only a finite number of moment conditions, MM/GMM can miss important information contained in the realizations. The QML approach simply approximates the true information. Not surprisingly, such an approximation could lose substantial amounts of information. Most of the simulation-based methods are more efficient, but they have higher computational costs.

The present paper uses another approach to estimate the SV model, namely the empirical characteristic function (ECF). The rationale for using the characteristic function (CF) is that there is a one-to-one correspondence between the CF and the cumulative distribution function. Consequently, the ECF should contain the same amount of information as the empirical cumulative distribution function (ecdf). Furthermore, by using the CF, we can overcome difficulties arising from ignorance of the true density function or the true likelihood function. In essence, the ECF method we propose is akin to approximating the likelihood by the product of joint densities of overlapping blocks, of a fixed size, in our data. In this sense it can be viewed as an approximation to the methods which claim to analyse the full likelihood, such as MCMC, etc. It is computationally less intensive than the simulation methods, however.

The paper is organized as follows. Section 2 introduces the simplest version of the SV model and explains why the model is difficult to estimate. Section 3 presents a general discussion of the ECF method, with particular emphasis on ECF estimation for the SV model; the CF of the SV model is obtained as well. In Section 4 we make a few remarks about the ECF estimation method. Section 5 illustrates the method by applying it to daily returns of Australian dollar–New Zealand dollar exchange rate. Section 6 concludes the paper. All proofs are collected in the Appendix.

2. The model

The formulation of the discrete time SV model is similar to that of ARCH-type models. That is, the conditional variance is directly modelled. However, in contrast to ARCH-type models, the SV model allows a random component in the transition equation. Therefore, the model can explain why large changes can follow stable periods. The simplest version of the SV model is of the form,

$$x_t = \sigma_t e_t \qquad (t = 1, 2, \dots, T),$$

where σ_t^2 is the conditional variance based on the information at the end of time t, and the e_t are independent and identically distributed (iid) random disturbances which are assumed to have a standard normal distribution. We define

$$\sigma_t = \exp(0.5h_t)$$

and assume h_t follows a Gaussian AR(1) process, i.e.

$$h_t = \lambda + \alpha h_{t-1} + v_t, \quad v_t \text{ iid } \mathrm{N}(0, \sigma_v^2)\,, \tag{1}$$

where $\boldsymbol{\theta} = (\alpha, \lambda, \sigma_v)$ are the unknown parameters. It is well known that if $|\alpha| < 1$, this process is invertible and stationary. Heuristically, we can say that the conditional variance depends on past conditional variance and a random component. When the effect of the past conditional variance is strong, volatility clustering appears in the series. However, if the random innovation is not dominated, it can bring a large change into a stable period and can smooth large booms and crashes as well. If the random component is omitted, the transition equation is deterministic and the model exhibits time-varying but deterministic volatility. Finally, we assume e_t and v_{t+1} are uncorrelated.

Some statistical properties of x_t are determined by h_t since x_t is a simple function of h_t. For example, h_t is stationary for $|\alpha| < 1$, so x_t is stationary as well. Furthermore, x_t is a martingale difference because e_t is a martingale difference. We also note that x_t has finite moments of all orders and in particular the second and fourth moments are given by $\mathrm{E}(x_t^2) = \exp\big(\lambda/(1-\alpha)+\sigma_v^2/(2(1-\alpha^2))\big)$, and $\mathrm{E}(x_t^4) = 3\exp\big(2\lambda/(1-\alpha)+2\sigma_v^2/(1-\alpha^2)\big)$. The kurtosis of x_t is therefore $3\exp\big(\sigma_v^2/(1-\alpha^2)\big)$, so x_t exhibits more kurtosis than a constant variance normal model. Furthermore, Harvey (1998) derives the moments of powers of the absolute value of x_t,

$$\mathrm{E}(|x_t|^c) = 2^{c/2}\frac{\Gamma(\frac{1}{2}c+\frac{1}{2})}{\Gamma(\frac{1}{2})}\exp\Big(\frac{c\lambda}{2(1-\alpha)} + \frac{c^2\sigma_v^2}{8(1-\alpha^2)}\Big) \qquad (c > -1)\,,$$

$$\mathrm{var}(|x_t|^c) = 2^c\Bigg(\frac{\Gamma(c+\frac{1}{2})}{\Gamma(\frac{1}{2})} - \Big(\frac{\Gamma(\frac{c}{2}+\frac{1}{2})}{\Gamma(\frac{1}{2})}\Big)^2\Bigg)\exp\Big(\frac{c\lambda}{1-\alpha} + \frac{c^2\sigma_v^2}{2(1-\alpha^2)}\Big) \qquad (c > -\tfrac{1}{2})\,.$$

Since x_t is a nonlinear function of an AR(1) process, however, the process is difficult to work with. Observing that the dependence of x_t is completely characterized by the dependence of h_t, we define y_t to be the logarithm of x_t^2. Then we have

$$y_t = \ln\sigma_t^2 + \ln e_t^2 = h_t + \epsilon_t \qquad (t = 1, 2, \ldots, T)\,, \tag{2}$$

where $\epsilon_t = \ln e_t^2$ is the logarithm of the chi-squared random variable with 1 degree of freedom. Hence, the new process y_t still depends on the AR(1) process h_t, but in a linear form. Since the process h_t contains all the parameters of interest, y_t loses no information from the estimation point of view, the only loss of information being the sign of e_t which for a symmetric distribution, uncorrelated in ϵ_t and v_t, contributes nothing to volatility estimation. This is why most of the estimation procedures in the literature are based on y_t, not x_t. Note that equation (2) together with equation (1) forms a non-normal state space model. See Durbin (2000) for more detailed discussion on the state space model.

Neither y_t nor x_t has a closed-form expression for the likelihood function. This property makes the estimation based on the likelihood extremely difficult. However, from (2) we know that y_t is the convolution of an AR(1) process and an iid logarithmic χ_1^2 sequence, and hence there is a closed-form expression for the CF of y_t which we derive in the next section. Since the CF contains the same amount of information as the distribution function, the model is fully and uniquely parameterized by the CF. Therefore, the model is estimable via the ECF.

3. ECF estimation

Before we discuss the estimation of the SV model via the ECF, it is worthwhile to briefly outline the ECF estimation method.

Suppose the cumulative distribution function (cdf) of X is $F(x;\boldsymbol{\theta})$ which depends on a parameter $\boldsymbol{\theta}$. The CF is defined as

$$c(r,\boldsymbol{\theta}) = \mathrm{E}\big(\exp(irx)\big) = \int \exp(irx)\,dF(x,\boldsymbol{\theta})\,,$$

and the ECF is the sample counterpart of the CF, that is,

$$c_n(r) = \frac{1}{n}\sum_{j=1}^{n}\exp(irx_j) = \int \exp(irx)\,dF_n(x)\,,$$

where $F_n(x)$ is the ecdf and r is the transformation variable. Therefore, the CF and ECF are the Fourier transformations of the cdf and ecdf. Because of the uniqueness of the Fourier–Stieltjes transform, the CF has the same information as the cdf and the ECF retains all the information in the sample. We also note that the CF contains only the parameters and the ECF contains only the data. The general idea for the ECF estimation method is to minimize various measures of the distance between the ECF and the CF. For example, by choosing a grid of values for the transformation variable, $r_1,\ldots,r_q$, we can minimize the distance

$$\sum_{j=1}^{q}|c_n(r_j)-c(r_j,\boldsymbol{\theta})|^2 g(r_j)\,,$$

i.e. the distance on q discrete points. Or by choosing the transformation variable r continuously, we can minimize

$$\int |c_n(r)-c(r,\boldsymbol{\theta})|^2 g(r)\,dr\,,$$

i.e. the distance over an interval. In both cases g is a weight function.

If the observations are an iid sequence, the marginal ecdf contains all the information in the sample and so does the marginal ECF. There is an extensive literature on estimation using the ECF for the iid environment, mostly by choosing the transformation variable discretely. To estimate the mixture of normals, for example, Quandt & Ramsey (1978) use the empirical moment generating function instead of the ECF. However, the ECF can be used in the same way; see Tran (1998). Moreover, there are known convergence results for the empirical characteristic function process $\sqrt{n}(c_n(r)-c(r,\boldsymbol{\theta}))$, which have been established by Feuerverger & Mureika (1977) and Csörgő (1981) for any iid sequence.

Estimating a strictly stationary stochastic process using the ECF is not exactly the same as estimating an iid sequence, because the dependence must be taken into account. Since the marginal ecdf does not capture the dependence of a dependent sequence, the marginal ECF would suffer the same problem. Therefore we need to use the joint CF, by a procedure involving moving blocks of data. We first define the overlapping blocks for $y_1, y_2, \ldots, y_T$ as

$$z_j = (y_j,\ldots,y_{j+p}) \qquad (j=1,\ldots,T-p)\,.$$

Hence each block has p periods overlapping with its adjacent blocks. The CF of each block is basically a joint one and is defined as

$$c(\boldsymbol{r},\boldsymbol{\theta}) = \mathrm{E}\big(\exp(i\boldsymbol{r}^{\mathsf{T}} z_j)\big)\,,$$

where the transformation variable $\boldsymbol{r}$ is a $p+1$ dimensional vector, i.e. $\boldsymbol{r} = (r_1,\dots,r_{p+1})$. The joint ECF is defined as

$$c_n(\boldsymbol{r}) = \frac{1}{n}\sum_{j=1}^{n} \exp(i\boldsymbol{r}^{\mathsf{T}} z_j)\,, \quad \text{where } n = T-p\,.$$

As a sequence of overlapping moving blocks, not surprisingly, z_j is dependent. Under standard regularity conditions, $c_n(\boldsymbol{r})$ converges almost surely to $c(\boldsymbol{r},\boldsymbol{\theta})$ for any $\boldsymbol{r}$ (see Feuerverger, 1990). Based on this result, Feuerverger (1990) and Knight & Satchell (1997) propose matching the joint CF with the joint ECF over a grid of discrete points. In practice, however, it is not entirely clear how to choose these discrete points and how many one should use. Rather than choosing the transformation variable discretely, we follow Knight & Yu (2002) and match the joint CF with the joint ECF over an interval. That is

$$\min_{\boldsymbol{\theta}\in\Theta} \int\cdots\int |c(\boldsymbol{r},\boldsymbol{\theta}) - c_n(\boldsymbol{r})|^2 g(\boldsymbol{r})\, dr_1 \cdots dr_{p+1}\,, \tag{3}$$

or solve

$$\int\cdots\int \big(c(\boldsymbol{r},\boldsymbol{\theta}) - c_n(\boldsymbol{r})\big) w(\boldsymbol{r})\, dr_1 \cdots dr_{p+1} = 0\,, \tag{4}$$

where both $g(\boldsymbol{r})$ and $w(\boldsymbol{r})$ are continuous weight functions. Under some regularity conditions on the weight functions, these two methods are equivalent and hence we consider only the procedure based on (3). Our calculations are with respect to the unconditional (steady-state) joint CF of z_j, but they could be done with respect to the conditional CF instead — an approach that has recently been suggested by Singleton (2001) and Chacko & Viceira (2001) for estimating diffusion processes.

In this paper, the procedure based on (3) is referred to as the continuous ECF method. In this method the transformation variable is simply integrated out. Also, the actual implementation requires the specification of a weight function. When the weight is chosen to be optimal, the procedure can lead to an estimator which achieves the Cramér–Rao lower bound. For example, Feuerverger (1990) shows that the ECF estimator is asymptotically efficient when the weight function in (4), $w(\boldsymbol{r})$, is given by

$$w(\boldsymbol{r}) = \int\cdots\int \exp(-i\boldsymbol{r}^{\mathsf{T}} z)\frac{\partial}{\partial\boldsymbol{\theta}} \ln f(y_{j+p} \mid y_j,\dots,y_{j+p-1})\, dy_j \cdots dy_{j+p}\,,$$

where f is the conditional probability density function (pdf) of the data. However, $w(\boldsymbol{r})$ is not calculable if the pdf is unknown, as is the case here for the SV model.

Under standard regularity conditions, Knight & Yu (2002) establish the strong consistency and asymptotic normality for the ECF estimators with a general weight function. In this paper we chose the weight to be an exponential function, $\exp(-\boldsymbol{r}'\boldsymbol{r})$.

The model we are going to estimate via the ECF is the one defined by (2) and (1) since we can derive the closed form expression of the CF. In order to use the ECF method we need to find the expression of the joint CF. Proposition 3.1 gives the CF for the logarithm of the χ_1^2 distribution; then the joint CF for $y_t,\dots,y_{t+k-1}$ is obtained in Proposition 3.2.

Proposition 3.1. *If* $\epsilon_t \overset{\mathrm{d}}{=} \ln \chi_1^2$ *then the* CF *of* ϵ_t *is*

$$c(r) = \frac{\Gamma(\frac{1}{2} + ir)}{\Gamma(\frac{1}{2})} 2^{ir} .$$

Proposition 3.2. *Suppose* $\{y_t\}_{t=1}^T$ *is defined by (2). The joint* CF *of* $y_t, \ldots, y_{t+k-1}$ *is*

$$c(r_1, \ldots, r_k, \boldsymbol{\theta}) = \exp\left(\frac{i\lambda}{1-\alpha}\sum_{j=1}^k r_j - \frac{\sigma_v^2}{2(1-\alpha^2)}\left(\sum_{j=1}^k r_j^2 + 2\alpha \sum_{\ell=1}^k \sum_{j=\ell+1}^k \alpha^{j-\ell-1} r_\ell r_j\right)\right) \times \frac{\prod_{j=1}^k \Gamma(\frac{1}{2} + ir_j)}{\Gamma(\frac{1}{2})^k} 2^{i\Sigma_{j=1}^k r_j} ,$$

where $k - 1 = p$ *is the size of the moving blocks.*

Using the joint CF we can obtain the joint cumulant generating function and consequently the autocorrelation function. The autocorrelation function of $\{y_t\}_{t=1}^T$ is given in Proposition 3.3.

Proposition 3.3. *Suppose* $\{y_t\}_{t=1}^T$ *is defined by (2); its autocorrelation is*

$$\rho_k = \alpha^k \frac{\sigma_v^2/(1-\alpha^2)}{\sigma_v^2/(1-\alpha^2) + c} \quad (k = 1, 2, \ldots), \quad \textit{where } c = \frac{\Gamma''(\frac{1}{2})}{\Gamma(\frac{1}{2})} - \left(\frac{\Gamma'(\frac{1}{2})}{\Gamma(\frac{1}{2})}\right)^2 \approx 0.4948 .$$

The y_t process defined by (2) is the sum of an AR(1) and white noise. It is well known that the result is a non-normal ARMA(1,1) model. This is confirmed by the expression of ρ_k.

To use the ECF method to estimate the SV model (2) and (1), we choose $p = 1$ (i.e. $k = 2$) at first. When $p = 1$ we have

$$\begin{aligned} &c(r_1, r_2, \boldsymbol{\theta}) \\ &\quad = \exp\left(i\lambda \frac{r_1 + r_2}{1-\alpha} - \frac{\sigma_v^2}{2(1-\alpha^2)}(r_1^2 + 2\alpha r_1 r_2 + r_2^2)\right) \frac{\Gamma(\frac{1}{2} + ir_1)\Gamma(\frac{1}{2} + ir_2)}{\Gamma(\frac{1}{2})^2} 2^{ir_1 + ir_2} , \end{aligned} \tag{5}$$

and

$$c_n(r_1, r_2) = \frac{1}{n}\sum_{j=1}^n \exp(ir_1 y_j + ir_2 y_{j+1}) .$$

Using $\operatorname{Re}\omega$ and $\operatorname{Im}\omega$ to denote the real and imaginary parts of ω, we have

$$\operatorname{Re} c_n(r_1, r_2) = \frac{1}{n}\sum_{j=1}^n \cos(r_1 y_j + r_2 y_{j+1}) \quad \text{and} \quad \operatorname{Im} c_n(r_1, r_2) = \frac{1}{n}\sum_{j=1}^n \sin(r_1 y_j + r_2 y_{j+1}) .$$

Therefore, the procedure is to choose $(\hat\alpha, \hat\sigma_v, \hat\lambda)$ to minimize

$$\iint \left(\left(\operatorname{Re} c(r_1, r_2) - \operatorname{Re} c_n(r_1, r_2)\right)^2 + \left(\operatorname{Im} c(r_1, r_2) - \operatorname{Im} c_n(r_1, r_2)\right)^2\right) \exp(-r_1^2 - r_2^2)\, dr_1\, dr_2 , \tag{6}$$

where $c(r_1, r_2) = c(r_1, r_2, \boldsymbol{\theta})$ is given by (5).

It is straightforward to check that the appropriate regularity conditions hold for the application of standard asymptotic theory. The resulting estimators are consistent and asymptotically normal with the asymptotic covariance matrix of the estimators given by

$$\frac{1}{n}\boldsymbol{B}(\boldsymbol{\theta})^{-1}\boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{B}(\boldsymbol{\theta})^{-1},$$

where $$\boldsymbol{B}(\boldsymbol{\theta}) = \int\cdots\int\left(\frac{\partial \mathrm{Re}\, c(\boldsymbol{r},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial \mathrm{Re}\, c(\boldsymbol{r},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^{\mathsf{T}}} + \frac{\partial \mathrm{Im}\, c(\boldsymbol{r},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial \mathrm{Im}\, c(\boldsymbol{r},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^{\mathsf{T}}}\right)g(\boldsymbol{r})\,d\boldsymbol{r}\,.$$

The appendix gives the expression for $\boldsymbol{A}(\boldsymbol{\theta})$ and a proof of the above result.

The actual implementation of the ECF method essentially requires minimizing (6), and thus involves double integrals. We numerically evaluate the multiple integral (6), and then numerically minimize (6) with respect to $\boldsymbol{\theta}$. The numerical solutions are the desired estimators.

We use a 39-points Gauss–Hermitian algorithm to approximate the two-dimensional integrations in (6) and the conjugate direction algorithm proposed by Powell (1964) to find the minimum of the objective function. As starting point in the optimization we chose the quasi-maximum likelihood estimates proposed by Ruiz (1994). All computations have been done in double precision.

4. Further discussion on the ECF method

The ECF estimator is dependent on the pre-specified weight function and the block size. In theory the ECF retains all the information in the sample and hence one should expect that the inference based on the ECF should work as well as that based on the likelihood function, but in practice, efficient estimation via the ECF requires a closed form solution of the conditional score. When the conditional score does not have a closed form expression, the optimal ECF method is not feasible. In Section 3 we proposed an exponential function as the weight, for two reasons. First, it enables us to use a particular quadrature method and hence is numerically convenient. Second, it puts more weight on the points around the origin, consistent with the recognition that the CF contains the most information around the origin. It is important to point out that with the exponential weight the ECF estimator is asymptotically inefficient. The finite sample behaviour of the ECF estimator with various exponential weights has been examined in detail in the SV context and the ARMA context by Yu (1998) and Knight & Yu (2002) respectively.

As to the block size, note that the blocks always contain no less information as p increases and thus the resulting estimators are supposed to be asymptotically more efficient. The relative efficiency is not clear in finite samples since a larger p also reduces the effective sample size (i.e. the total number of blocks). Moreover, calculations associated with larger p are numerically more intensive. Therefore, in practice there are always two trade-offs: the block size versus the sample size and the asymptotic efficiency versus the computational efficiency. In the framework of linear ARMA models, Knight & Yu (2002) show that the choice of p is related to the dimension of the minimal sufficient statistics. For pure AR(ℓ) processes, for instance, the overlapping moving blocks with $p = \ell$ form a set of sufficient statistics and hence it is good enough to set $p = \ell$. For a general ARMA process, however, since any statistics of dimension less than the sample size are not sufficient, the blocks with a larger p always improve asymptotic efficiency. Knight & Yu (2002) further suggest that p should be chosen such that the ARMA process can be well approximated by an AR(p) process.

As we showed in Section 3, the SV model can be represented by a non-normal ARMA(1,1) model. Consequently, it is not clear what the optimal block size should be. Since y_t follows an ARMA(1,1) model, the partial correlogram of y_t can be used to choose a reasonable p. For example, if the partial correlogram cuts off after lag ℓ, p can be set at $\ell - 1$.

The basic SV model discussed in Section 2 is too simplistic for many financial series. Many more flexible SV models have been proposed in the literature. One way to generalize the basic SV model is to assume that disturbances v_t and/or e_t have a fatter tailed distribution. In recent literature, however, evidence has been found to favour the log-normal specification of the volatility process. Using high frequency data, Andersen *et al.* (2001) find that the marginal distribution of the realized volatility can be well approximated by a log-normal distribution.

Another way to make the basic SV model more flexible is to allow a long-range dependence in volatilities (see e.g. Harvey, 1998; Breidt, Crato & De Lima, 1998). Unfortunately, this specification invalidates one of the key assumptions imposed by Knight & Yu (2002) to ensure the strong law of large numbers and the central limit theorem. In particular, the strong mixing condition does not hold for the long memory specification.

A third way to generalize the basic SV model is to allow the so-called leverage effect where shocks and negative shocks change the debt–equity ratio of a firm in different directions and so, at a different scale, change the riskiness of the firm. In the SV context, the model with the leverage effect is defined by

$$x_t = \sigma_t e_t \quad \text{where } \epsilon_t \text{ iid N}(0,1)\,, \; y_t = \ln x_t^2\,, \tag{7}$$

$$\ln\sigma_t^2 = \lambda + \alpha\ln\sigma_{t-1}^2 + v_t \quad \text{where } v_t \text{ iid N}(0,\sigma_v^2) \text{ and } \operatorname{cov}(e_t, v_{t+1}) = \rho\sigma_v^2\,. \tag{8}$$

If $\rho < 0$, the model has a leverage effect.

It can be shown that for the above SV model, the CF has a closed form expression and hence is estimable via the ECF. The joint CF of $y_t, \ldots, y_{t+k-1}$ is given in Proposition 4.1.

Proposition 4.1. *Suppose* $\{y_t\}_{t=1}^T$ *is defined by (7) and (8). The* CF *of* $y_t, \ldots, y_{t+k-1}$ *is*

$$\begin{aligned} c(r_1,\ldots,r_k;\boldsymbol{\theta}) &= \exp\left(\frac{i\lambda}{1-\alpha}\sum_{j=1}^k r_j\right)\exp\left(-\frac{\sigma_v^2\alpha^2}{2(1-\alpha^2)}\left(\sum_{j=1}^k r_j\alpha^{k-j}\right)^2\right) \\ &\quad\times\exp\left(-\frac{\sigma_v^2(1-\rho^2)}{2}\sum_{m=1}^k\left(\sum_{j=1}^k r_j\alpha^j\right)^2\right)\frac{\prod_{j=1}^k\Gamma(\frac12+ir_j)}{\Gamma^k(\frac12)}\,2^{i\sum_{j=1}^k r_j} \\ &\quad\times\prod_{j=1}^k {}_1F_1\left(ir_j+\tfrac12,\tfrac12;-\frac{\sigma_v^2\rho^2}{2}\left(\sum_{j=1}^k r_j\alpha^{k-j}\right)^2\right), \end{aligned}$$

where ${}_1F_1$ *denotes the confluent hypergeometric function.*

All the discrete SV models discussed above can be regarded as the discrete time approximation to pure diffusion processes with stochastic volatility. A related model specification is to mix diffusion with jumps. A special case of the jump-diffusion processes is the affine jump-diffusion introduced by Duffie, Pan & Singleton (2000). For the affine jump-diffusion, they show that the CF has a closed form expression. Consequently, Singleton (2001) and Jiang & Knight (2002) estimate the affine jump-diffusion via the ECF method.

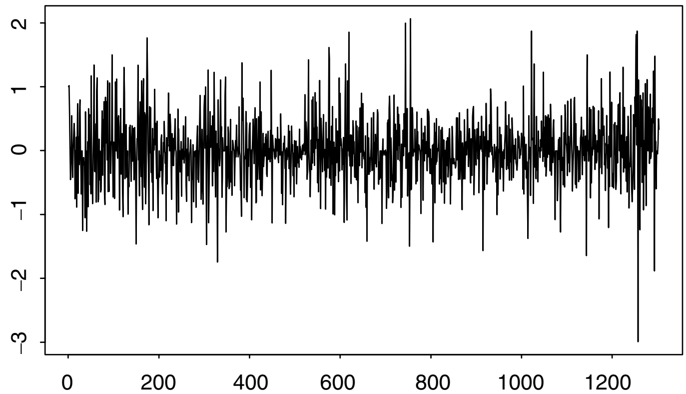

Figure 1. Daily returns of the exchange rate

Since the ECF method is a frequency approach, the parameters of interest exclude the unobserved state variables which are the unobserved volatilities in the SV model. Although the ECF method cannot directly estimate these unobserved volatilities, their interpretation is of interest. To estimate the unobserved volatilities, one can take advantage of the state space representation of the model and make use of any filtering technique for non-Gaussian state space models (such as Frühwirth-Schnatter, 1994).

5. Application

The data we use are the daily returns of Australian dollar–New Zealand dollar exchange rate covering the period from 1 January 1993 to 31 December 1997. Suppose r_t is the exchange rate at time t. The mean-corrected returns are computed as

$$x_t = 100\Big((\ln r_t - \ln r_{t-1}) - \frac{1}{T}(\ln r_T - \ln r_0)\Big).$$

The sample size is $T = 1304$. Figure 1 plots the series. Table 1 presents summary statistics of x_t, $|x_t|$, and x_t^2. There is a high level of excess kurtosis in x_t, which suggests that the marginal distribution of x_t is not normal. The autocorrelation function (ACF) of x_t suggests the series is a martingale difference sequence while both x_t^2 and $|x_t|$ show a high degree of persistence as given by the ACFs. This is evidence of volatility clustering. The argument is further reinforced by the Ljung–Box statistic which indicates that there is not much serial correlation in x_t but significant serial correlation in both x_t^2 and $|x_t|$.

The basic SV model defined in Section 2 is fitted to the dataset and the empirical results with $p = 1$ are reported in Table 2. It can be seen that all three parameters are significant. Compared to daily returns of many other financial time series (see e.g. Jacquier *et al.*, 1994), the stochastic exchange rate volatility is less persistent. That is, the stochastic exchange rate volatility reverts more quickly to its long-term mean as evidenced by a relatively small α estimate.

Larger values of p are also used for the ECF method. In Table 3, we report the ECF estimates for $p = 2, 3, 4, 5$ where we fit the SV model to the same dataset. This table shows that the empirical results remain almost unchanged for different values of p and are very close to those for $p = 1$. This exercise indicates that a small value of p can work well for the ECF method, at least for this problem.

TABLE 1

Summary statistics of the exchange rates

	x_t	$\|x_t\|$	x_t^2
Mean	0	0.3738	0.2684
Standard deviation	0.5183	0.3588	0.5373
Kurtosis	4.997	7.167	65.70
Minimum	–2.991	0.0034	0.00001
Median	–0.004	0.2686	0.072
Maximum	2.066	2.991	8.946
ACF(1)	–0.0071	0.0553	0.0943
ACF(2)	–0.0611	0.0939	0.1420
ACF(3)	–0.0444	0.0587	0.0380
ACF(4)	–0.0011	0.0716	0.1292
ACF(5)	0.0113	0.0496	0.0681
ACF(10)	–0.0421	0.0510	0.0246
ACF(20)	–0.0222	0.0612	0.0715
ACF(30)	–0.0008	0.0088	0.0171
Q(10)	12.12 (0.277)	57.22 (0)	73.96 (0)
Q(20)	24.88 (0.206)	86.72 (0)	98.22 (0)
Q(30)	31.86 (0.374)	113.59 (0)	106.34 (0)

Note: ACF(k) denotes the value of the autocorrelation functions of order k. $Q(k)$ denotes the Ljung–Box test statistic of order k with P-values in parentheses.

TABLE 2

Empirical results for the exchange rates

	α	σ_v	λ
Estimate	0.8247	0.3894	–0.2760
Standard error	0.0756	0.0988	0.100

Note: We choose $p = 1$ for the ECF method.

TABLE 3

Empirical results of ECF estimates with different values of p

	α	σ_v	λ
$p = 2$	0.8089 (0.07471)	0.4047 (0.1026)	–0.3021 (0.12345)
$p = 3$	0.8062 (0.08173)	0.4064 (0.1102)	–0.3063 (0.1307)
$p = 4$	0.8051 (0.07847)	0.4094 (0.1046)	–0.3076 (0.1302)
$p = 5$	0.8118 (0.07449)	0.4022 (0.1001)	–0.2962 (0.1236)

Note: Asymptotic standard errors are presented in parentheses.

Table 4 presents the unconditional model moments of x_t (namely, the variance and the kurtosis of x_t, and the expected value and the variance of $|x_t|$) calculated by inserting the ECF estimates with different values of p into the theoretical expressions and the same moments from data. Also presented are the asymptotic standard errors of point estimates of these moments, using the delta method. Table 5 and Figures 2–3 further compare the autocorrelation of the squared returns and the absolute returns with the implied theoretical autocorrelation of x_t^2 and $|x_t|$ with the same estimates. The analytical expressions of both unconditional model moments and the ACFs of x_t^2 and $|x_t|$ are given in Ghysels *et al.* (1996). Tables 4–5 and Figures 2–3 show that the ECF estimates provide all the four unconditional moments very close to their data counterparts. In terms of the ACFs, the ECF estimates capture the dynamic

TABLE 4

Comparison between unconditional data moments and unconditional model moments

	Data	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$
$\mathrm{var}(x_t)$	0.2686	0.2625	0.2608	0.2606	0.2618	0.2627
		(0.2279)	(0.2113)	(0.2257)	(0.2155)	(0.2129)
$\mathrm{kurt}(x_t)$	4.9966	4.8194	4.8182	4.8098	4.8308	4.8212
		(2.2952)	(2.2407)	(2.2810)	(2.2781)	(2.2165)
$\mathrm{E}(\lvert x_t\rvert)$	0.3738	0.3853	0.3840	0.3840	0.3846	0.3854
		(0.1682)	(0.1557)	(0.1666)	(0.1584)	(0.1564)
$\mathrm{var}(\lvert x_t\rvert)$	0.1287	0.0954	0.0948	0.0947	0.0951	0.0955
		(0.0828)	(0.0768)	(0.0820)	(0.0783)	(0.0774)

Note: Asymptotic standard errors are presented in parentheses.

TABLE 5

Comparison between conditional data moments and conditional model moments

	Data	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$
$\rho_1(\lvert x_t\rvert)$	0.055	0.104	0.101	0.101	0.101	0.102
$\rho_2(\lvert x_t\rvert)$	0.094	0.085	0.081	0.081	0.081	0.082
$\rho_3(\lvert x_t\rvert)$	0.059	0.069	0.065	0.064	0.065	0.066
$\rho_4(\lvert x_t\rvert)$	0.072	0.057	0.053	0.052	0.052	0.053
$\rho_5(\lvert x_t\rvert)$	0.050	0.047	0.042	0.041	0.041	0.043
$\rho_1(\lvert x_t\rvert^2)$	0.094	0.071	0.076	0.069	0.069	0.070
$\rho_2(\lvert x_t\rvert^2)$	0.142	0.056	0.064	0.054	0.053	0.054
$\rho_3(\lvert x_t\rvert^2)$	0.038	0.045	0.054	0.042	0.042	0.043
$\rho_4(\lvert x_t\rvert^2)$	0.129	0.036	0.046	0.033	0.033	0.034
$\rho_5(\lvert x_t\rvert^2)$	0.068	0.029	0.040	0.026	0.026	0.027

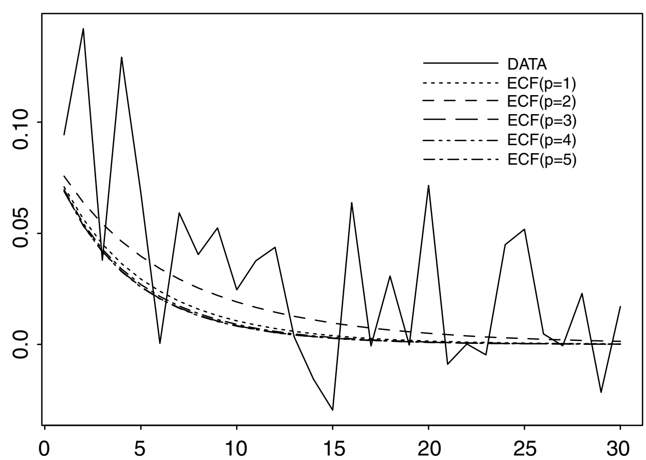

Figure 2. ACF of x^2

properties in the data reasonably well although the sample autocorrelation is more persistent. Moreover, from Tables 4–5 and Figures 2–3 we note that the unconditional moments from the ECF estimates with different values of p remain almost the same and the ACFs converge quickly as p gets larger. This exercise serves to illustrate that the SV model is a reasonable model specification for the data and the ECF technique is a viable estimation method.

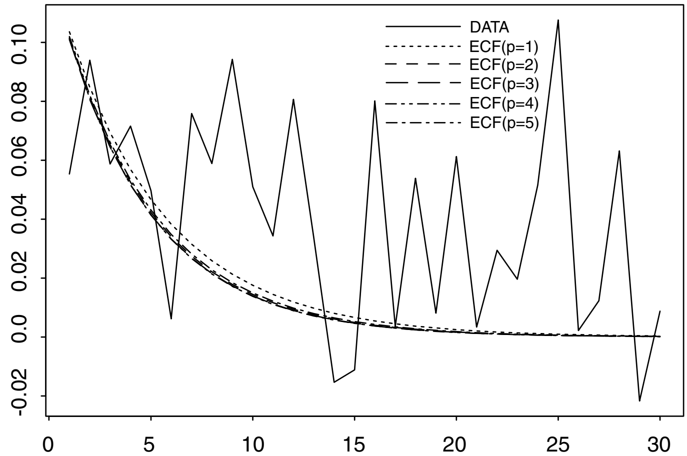

Figure 3. ACF of $|x|$

6. Conclusion

This paper proposes a method for estimating SV model via the ECF. Although the likelihood method does not have a closed form for this model, we show that the CF can be derived analytically and hence the ECF method can be used to estimate the system parameters. The basic idea of the ECF method is to match the theoretical CF derived from the model to the ECF calculated from the sampling observations. The approach yields consistent and asymptotically normal estimates of the parameters. An empirical application to the Australian dollar–New Zealand dollar exchange rate shows the capabilities of the ECF method.

Appendix

Proof of Proposition 3.1. If $X \overset{\mathrm{d}}{=} \gamma(r,\alpha)$, then $\mathrm{E}(X^k) = \Gamma(r+k)/\Gamma(r)\alpha^k$, provided that $\mathrm{Re}(r+k) > 0$. Now we have $\epsilon_t \overset{\mathrm{d}}{=} \ln U$ with $U \overset{\mathrm{d}}{=} \chi_1^2 = \gamma(\frac{1}{2},\frac{1}{2})$, so

$$c(r) = \mathrm{E}\big(\exp(ir\epsilon_t)\big) = \mathrm{E}(U^{ir}) = \frac{\Gamma(\frac{1}{2}+ir)2^{ir}}{\Gamma(\frac{1}{2})}\,.$$

Proof of Proposition 3.2. Since y_t is a convolution of a Gaussian AR(1) process and an iid sequence with χ_1^2 distribution, we have

$$\begin{aligned}
c(r_1,\dots,r_k,\boldsymbol{\theta})&\\
&= \mathrm{E}\big(\exp(ir_1y_t + ir_2y_{t+1} + \cdots + ir_ky_{t+k-1})\big)\\
&= \mathrm{E}\big(\exp(ir_1h_t + ir_1\epsilon_t + ir_2h_{t+1} + ir_2\epsilon_{t+1} + \cdots + ir_kh_{t+k-1} + ir_k\epsilon_{t+k-1})\big)\\
&= \mathrm{E}\big(\exp(ir_1h_t + ir_2h_{t+1} + \cdots + ir_kh_{t+k-1})\big)\prod_{j=1}^{k}\mathrm{E}\big(\exp(ir_j\epsilon_{t+j-1})\big)\\
&= \mathrm{E}\Bigg(\exp\Bigg(ih_t\sum_{j=1}^{k}\alpha^{j-1}r_j + i\lambda\sum_{j=2}^{k}\frac{1-\alpha^{j-1}}{1-\alpha}\,r_j + \sum_{\ell=2}^{k}\nu_{t+\ell-1}\sum_{j=\ell}^{k}r_j\alpha^{j-\ell}\Bigg)\Bigg)\\
&\qquad\times\prod_{j=1}^{k}\mathrm{E}\big(\exp(ir_j\epsilon_{t+j-1})\big)
\end{aligned}$$

$$
\begin{aligned}
&= \exp\Bigg(i\frac{\lambda}{1-\alpha}\sum_{j=1}^{k}\alpha^{j-1}r_j + i\lambda\sum_{j=2}^{k}\frac{1-\alpha^{j-1}}{1-\alpha}r_j - \tfrac{1}{2}\Bigg(\sum_{j=1}^{k}\alpha^{j-1}r_j\Bigg)^2\frac{\sigma_v^2}{1-\alpha^2} \\
&\qquad - \tfrac{1}{2}\sum_{\ell=2}^{k}\Bigg(\sum_{j=\ell}^{k}\alpha^{j-\ell}r_j\Bigg)^2\sigma_v^2\Bigg)\frac{\prod_{j=1}^{k}\Gamma(\frac{1}{2}+ir_j)}{\Gamma(\frac{1}{2})^k}\,2^{i\Sigma_{j=1}^{k}r_j} \\
&= \exp\Bigg(\frac{i\lambda}{1-\alpha}\sum_{j=1}^{k}r_j - \frac{\sigma_v^2}{2(1-\alpha^2)}\Bigg(\sum_{j=1}^{k}r_j^2 + 2\alpha\sum_{\ell=1}^{k}\sum_{j=\ell+1}^{k}\alpha^{j-\ell-1}r_\ell r_j\Bigg)\Bigg) \\
&\qquad \times \frac{\prod_{j=1}^{k}\Gamma(\frac{1}{2}+ir_j)}{\Gamma(\frac{1}{2})^k}\,2^{i\Sigma_{j=1}^{k}r_j}\,.
\end{aligned}
$$

Proof of Proposition 3.3. Defined as the logarithm of the CF, the cumulant function is of the form

$$
\begin{aligned}
\phi(r_1,\dots,r_k) &= \ln\big(c(r_1,\dots,r_k,\boldsymbol{\theta})\big) \\
&= \frac{i\lambda}{1-\alpha}\sum_{j=1}^{k}r_j - \frac{\sigma_v^2}{2(1-\alpha^2)}\Bigg(\sum_{j=1}^{k}r_j^2 + 2\alpha\sum_{\ell=1}^{k}\sum_{j=\ell+1}^{k}\alpha^{j-\ell-1}r_\ell r_j\Bigg) \\
&\qquad + \sum_{j=1}^{k}\ln\big(\Gamma(\tfrac{1}{2}+ir_j)\big) + i\ln 2\sum_{j=1}^{k}r_j\,.
\end{aligned}
$$

Therefore, we have

$$
\operatorname{var}(y_t) = \frac{\partial^2\phi}{\partial r_1^2}\Big|_{r_1=0} = \frac{\sigma_v^2}{1-\alpha^2} + c\,,
$$

where $c = \Gamma''(\frac{1}{2})/\Gamma(\frac{1}{2}) - (\Gamma'(\frac{1}{2})/\Gamma(\frac{1}{2}))^2$; and

$$
\operatorname{cov}(y_t, y_{t+k}) = \frac{\partial^2\phi}{\partial r_1\partial r_k}\Big|_{r_1=0,r_k=0} = \frac{\alpha^k\sigma_v^2}{1-\alpha^2} \qquad (k=1,2,\dots)\,.
$$

Hence the autocorrelation functions are

$$
\rho_k = \alpha^k\frac{\sigma_v^2/(1-\alpha^2)}{\big(\sigma_v^2/(1-\alpha^2)\big)+c} \qquad (k=1,2,\dots)\,.
$$

The covariance matrix of the ECF estimator

We present details of how we calculate the asymptotic covariance matrix of the ECF estimator. Let our ECF estimator be given by $\hat{\boldsymbol{\theta}}$ where

$$
\hat{\boldsymbol{\theta}} = \arg\min\Delta(\boldsymbol{\theta})\,,
$$

and

$$
\Delta(\boldsymbol{\theta}) = \int\!\cdots\!\int\Big(\big(\operatorname{Re}c_n(\boldsymbol{r}) - \operatorname{Re}c(\boldsymbol{r},\boldsymbol{\theta})\big)^2 + \big(\operatorname{Im}c_n(\boldsymbol{r}) - \operatorname{Im}c(\boldsymbol{r},\boldsymbol{\theta})\big)^2\Big)g(\boldsymbol{r})\,d\boldsymbol{r}\,.
$$

Now since

$$\operatorname{Re} c_n(\boldsymbol{r}) = \frac{1}{n}\sum_{j=1}^{n} \cos \boldsymbol{r}^\top \boldsymbol{z}_j \quad \text{and} \quad \operatorname{Im} c_n(\boldsymbol{r}) = \frac{1}{n}\sum \sin \boldsymbol{r}^\top \boldsymbol{z}_j\,,$$

we have

$$\frac{\partial \Delta(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\frac{2}{n}\sum_{j=1}^{n} \delta_j(\boldsymbol{\theta})\,,$$

where
$$\delta_j(\boldsymbol{\theta}) = \int\cdots\int \Big(\partial c_R(\boldsymbol{r},\boldsymbol{\theta})\big(\cos \boldsymbol{r}^\top \boldsymbol{z}_j - \operatorname{Re} c(r,\boldsymbol{\theta})\big) + \partial c_I(\boldsymbol{r},\boldsymbol{\theta})\big(\sin \boldsymbol{r}^\top \boldsymbol{z}_j - \operatorname{Im} c(r,\boldsymbol{\theta})\big)\Big) g(\boldsymbol{r})\, d\boldsymbol{r}\,,$$

where
$$\partial c_R(\boldsymbol{r},\boldsymbol{\theta}) = \frac{\partial \operatorname{Re} c(\boldsymbol{r},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\,, \qquad \partial c_I(\boldsymbol{r},\boldsymbol{\theta}) = \frac{\partial \operatorname{Im} c(\boldsymbol{r},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\,.$$

Consequently

$$\sqrt{n}\,\frac{\partial \Delta(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \xrightarrow{\mathrm{d}} \mathrm{N}\big(0, 4\boldsymbol{A}(\boldsymbol{\theta})\big)\,, \quad \text{where } \boldsymbol{A}(\boldsymbol{\theta}) = \lim_{n\to\infty} \mathrm{E}\bigg(\frac{1}{n}\sum_j\sum_k \delta_j(\boldsymbol{\theta})\delta_k(\boldsymbol{\theta})\bigg)$$

and is given by:

$$\begin{aligned}\boldsymbol{A}(\boldsymbol{\theta}) = \lim_{n\to\infty}\frac{1}{n}\int\cdots\int \bigg(&\partial c_R(\boldsymbol{r},\boldsymbol{\theta})\partial c_R^\top(\boldsymbol{s},\boldsymbol{\theta})\sum_j\sum_k \operatorname{cov}(\cos \boldsymbol{r}^\top \boldsymbol{z}_j, \cos \boldsymbol{s}^\top \boldsymbol{z}_k)\\
&+ \partial c_R(\boldsymbol{r},\boldsymbol{\theta})\partial c_I^\top(\boldsymbol{s},\boldsymbol{\theta})\sum_j\sum_k \operatorname{cov}(\cos \boldsymbol{r}^\top \boldsymbol{z}_j, \sin s^\top \boldsymbol{z}_k)\\
&+ \partial c_I(\boldsymbol{r},\boldsymbol{\theta})\partial c_R^\top(\boldsymbol{s},\boldsymbol{\theta})\sum_j\sum_k \operatorname{cov}(\sin \boldsymbol{r}^\top \boldsymbol{z}_j, \cos s^\top \boldsymbol{z}_k)\\
&+ \partial c_I(\boldsymbol{r},\boldsymbol{\theta})\partial c_I^\top(\boldsymbol{s},\boldsymbol{\theta})\sum_j\sum_k \operatorname{cov}(\sin \boldsymbol{r}^\top \boldsymbol{z}_j, \sin s^\top \boldsymbol{z}_k)\bigg) g(\boldsymbol{r})g(\boldsymbol{s})\, d\boldsymbol{r}\, d\boldsymbol{s}\,.\end{aligned}$$

The double summation covariance expressions are readily found and are given in Knight & Satchell (1997 Lemma p. 176):

$$\sum_j\sum_k \operatorname{cov}(\cos \boldsymbol{r}^\top \boldsymbol{z}_j, \cos \boldsymbol{s}^\top \boldsymbol{z}_k) = n^2 \operatorname{cov}\big(\operatorname{Re} c_n(\boldsymbol{r}), \operatorname{Re} c_n(\boldsymbol{r})\big) = n^2(\Omega_{RR})_{\boldsymbol{r},\boldsymbol{s}}\,;$$

and similarly for the other double sums. Thus

$$\begin{aligned}A(\boldsymbol{\theta}) = \lim_{n\to\infty} n\int\cdots\int \Big(&\partial c_R(\boldsymbol{r},\boldsymbol{\theta})\partial c_R^\top(\boldsymbol{s},\boldsymbol{\theta})(\Omega_{RR})_{\boldsymbol{r},\boldsymbol{s}}\\
&+ 2\partial c_R(\boldsymbol{r},\boldsymbol{\theta})\partial c_I^\top(\boldsymbol{s},\boldsymbol{\theta})(\Omega_{RI})_{\boldsymbol{r},\boldsymbol{s}} + \partial c_I(\boldsymbol{r},\boldsymbol{\theta})\partial c_I^\top(\boldsymbol{s},\boldsymbol{\theta})(\Omega_{II})_{\boldsymbol{r},\boldsymbol{s}}\Big) g(\boldsymbol{r})g(\boldsymbol{s})\, d\boldsymbol{r}\, d\boldsymbol{s}\,.\end{aligned}$$

Furthermore,

$$\begin{aligned}
\mathrm{E}\Big(\frac{\partial^2\Delta(\boldsymbol\theta)}{\partial\boldsymbol\theta\partial\boldsymbol\theta^\top}\Big) &= -\frac{2}{n}\sum_{j=1}^{n}\mathrm{E}\Big(\frac{\partial\delta_j(\boldsymbol\theta)}{\partial\boldsymbol\theta}\Big)\\
&= \frac{2}{n}\sum_{j=1}^{n}\int\cdots\int\big(\partial c_R(\boldsymbol r,\boldsymbol\theta)\partial c_R^\top(\boldsymbol r,\boldsymbol\theta)+\partial c_I(\boldsymbol r,\boldsymbol\theta)\partial c_I^\top(\boldsymbol r,\boldsymbol\theta)\big)g(\boldsymbol r)\,d\boldsymbol r\\
&= -2\int\cdots\int\big(\partial c_R(\boldsymbol r,\boldsymbol\theta)\partial c_R^\top(\boldsymbol r,\boldsymbol\theta)+\partial c_I(\boldsymbol r,\boldsymbol\theta)\partial c_I^\top(\boldsymbol r,\boldsymbol\theta)\big)g(\boldsymbol r)\,d\boldsymbol r\\
&= -2\boldsymbol B(\boldsymbol\theta)\,.
\end{aligned}$$

Thus standard asymptotic theory results in

$$\sqrt{n}\,(\hat{\boldsymbol\theta}-\boldsymbol\theta)\xrightarrow{\mathrm{d}}\mathrm{N}\big(0,\boldsymbol B(\boldsymbol\theta)^{-1}\boldsymbol A(\boldsymbol\theta)\boldsymbol B^{-1}(\boldsymbol\theta)^{-1}\big)\,.$$

Proof of Proposition 4.1. By assumption

$$\begin{bmatrix} e_t\\ v_t\end{bmatrix}\stackrel{\mathrm{d}}{=}\mathrm{N}_2\left(\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}1 & \rho\sigma_v\\ \rho\sigma_v & \sigma_v^2\end{bmatrix}\right)$$

and hence $v_t\mid e_t\sim\mathrm{N}(\rho\sigma_v e_t,\sigma_v^2(1-\rho^2))$. Without loss of generality, we derive only the CF of the SV model with the leverage effect when $k=2$; that is, $c(r_1,r_2)$:

$$\begin{aligned}
c(r_1,r_2) &= \mathrm{E}\big(\exp(ir_1y_t+ir_2y_{t-1})\big)\\
&= \mathrm{E}\big(\exp(ir_1h_t+ir_1\ln e_t^2+ir_2h_{t-1}+ir_2\ln e_{t-1}^2)\big)\,.
\end{aligned}$$

Since $r_1h_t=r_1\lambda+r_1\alpha h_{t-1}+r_1v_t$, we have

$$c(r_1,r_2)=\mathrm{E}\big(\exp(ir_1\lambda)\exp\big(i(r_1\alpha+r_2)h_{t-1}+ir_1v_t\big)\exp(ir_1\ln e_t^2+ir_2\ln e_{t-1}^2)\big)\,.$$

With

$$h_{t-1}=\frac{\lambda}{1-\alpha}+\sum_{j=0}^{\infty}\alpha^j v_{t-1-j}=\frac{\lambda}{1-\alpha}+\sum_{j=1}^{\infty}\alpha^j v_{t-1-j}+v_{t-1}\,,$$

$$\begin{aligned}
c(r_1,r_2) &= \mathrm{E}\Bigg(\exp\bigg(ir_1\lambda+i(r_1\alpha+r_2)\frac{\lambda}{1-\alpha}+i(r_1\alpha+r_2)\sum_{j=1}^{\infty}\alpha^j v_{t-1-j}\\
&\qquad +i(r_1\alpha+r_2)v_{t-1}+ir_1v_t+ir_1\ln e_t^2+ir_2\ln e_{t-1}^2\bigg)\Bigg)\\
&= \exp\Big(i(r_1+r_2)\,\frac{\lambda}{1-\alpha}\Big)\prod_{j=1}^{\infty}\exp\Big(-\frac{\sigma_v^2}{2}\,\alpha^{2j}(r_1\alpha+r_2)^2\Big)\\
&= \mathrm{E}\Big(\exp\big(i(r_1\alpha+r_2)v_{t-1}+ir_2\ln e_{t-1}^2\big)\Big)\mathrm{E}\Big(\exp(ir_1v_t+ir_1\ln e_t^2)\Big)\,. \qquad (\mathrm{A}.1)
\end{aligned}$$

We derive an expression for $\mathrm{E}(\exp(iqv_t+ic\ln e_t^2))$, as (A.1) involves a product of two expressions of this form. So,

$$\begin{aligned}
\mathrm{E}\big(\exp(iqv_t+ic\ln e_t^2)\big) &= \mathrm{E}\Big(\exp(ic\ln e_t^2)\mathrm{E}\big(\exp(iqv_t)\mid e_t\big)\Big)\\
&= \mathrm{E}\Big(\exp(ic\ln e_t^2)\exp(iq\sigma_v\rho e_t)\exp\big(-\tfrac12\sigma_v^2(1-\rho)^2q^2\big)\Big)\\
&= \mathrm{E}\big(e_t^{2ic}\exp(ibe_t)\big)\exp\big(-\tfrac12\sigma_v^2(1-\rho)^2q^2\big)\,,\ \text{where } b=q\sigma_v\rho.
\end{aligned}$$

Let $e_t = S\sqrt{U}$ where $U \overset{\mathrm{d}}{=} \chi_1^2$ and $S = \pm 1$ with probability $\frac{1}{2}$ independent of U. Thus

$$\begin{aligned}
\mathrm{E}\big(e_t^{2ic}\exp(ibe_t)\big) &= \mathrm{E}(U^{ic}e^{ibS\sqrt{U}}) \\
&= \mathrm{E}\big(U^{ic}\tfrac{1}{2}(e^{ib\sqrt{U}} + e^{-ib\sqrt{U}})\big) \\
&= \mathrm{E}\bigg(U^{ic}\sum_{k=0}^{\infty}\frac{(ib\sqrt{U})^{2k}}{(2k)!}\bigg) \\
&= \mathrm{E}\bigg(\sum_{k=0}^{\infty}\frac{(ib)^{2k}}{(2k)!}U^{k+ic}\bigg) \\
&= \sum_{k=0}^{\infty}\frac{(ib)^{2k}}{(2k)!}\frac{\Gamma(k+ic+\frac{1}{2})}{\Gamma(\frac{1}{2})(\frac{1}{2})^{k+ic}} \\
&= \frac{2^{ic}}{\sqrt{\pi}}\sum_{k=0}^{\infty}\frac{(-2b^2)^k}{2^{2k}k!(\frac{1}{2})_k}\Gamma(k+ic+\tfrac{1}{2}),
\end{aligned}$$

where $(x)_k = x(x+1)\cdots(x+k-1)$. Noting that $\Gamma(k+ic+\frac{1}{2}) = (ic+\frac{1}{2})_k\Gamma(ic+\frac{1}{2})$, we obtain

$$\mathrm{E}\big(e_t^{2ic}\exp(ibe_t)\big) = \frac{2^{ic}\Gamma(ic+\frac{1}{2})}{\sqrt{\pi}}\,{}_1F_1(ic+\tfrac{1}{2}, \tfrac{1}{2}; -\tfrac{1}{2}b^2),$$

where ${}_1F_1$ denotes the confluence hypergeometric function. It follows that

$$\begin{aligned}
&\mathrm{E}\big(\exp(iqv_t + ic\ln e_t^2)\big) \\
&\qquad = \frac{2^{ic}\Gamma(ic+\frac{1}{2})}{\sqrt{\pi}}\exp\big(-\tfrac{1}{2}\sigma_v^2(1-\rho)^2q^2\big)\,{}_1F_1\big(ic+\tfrac{1}{2}, \tfrac{1}{2}; -\tfrac{1}{2}q^2\sigma_v^2\rho^2\big).
\end{aligned}$$

Using this expression in (A.1) we obtain

$$\begin{aligned}
c(r_1, r_2; \boldsymbol{\theta}) &= \exp\Big(\frac{i\lambda}{1-\alpha}(r_1+r_2)\Big)\exp\Big(-\frac{\sigma_v^2\alpha^2}{2(1-\alpha^2)}(r_1\alpha+r_2)^2\Big) \\
&\quad\times\exp\Big(-\frac{\sigma_v^2(1-\rho^2)}{2}(r_1\alpha+r_2)^2\Big)\frac{\Gamma(\frac{1}{2}+ir_1)\Gamma(\frac{1}{2}+ir_2)}{\Gamma(\frac{1}{2})^2}2^{i(r_1+r_2)} \\
&\quad\times{}_1F_1\big(ir_1+\tfrac{1}{2}, \tfrac{1}{2}; -\tfrac{1}{2}r_1^2\sigma_v^2\rho^2\big)\,{}_1F_1\big(ir_2+\tfrac{1}{2}, \tfrac{1}{2}; -\tfrac{1}{2}(r_1\alpha+r_2)^2\sigma_v^2\rho^2\big).
\end{aligned}$$

References

Andersen, T. & Sorensen, B. (1996). GMM estimation of a stochastic volatility model: A Monte Carlo study. *J. Bus. Econom. Statist.* **14**, 329–352.

Andersen, T., Bollerslev, T., Diebold, F. & Labys, P. (2001). The distribution of exchange rate volatility. *J. Amer. Statist. Assoc.* **96**, 42–55.

Breidt, F.J., Crato, N. & De Lima, P. (1998). The detection and estimation of long memory in stochastic volatility. *J. Econometrics* **83**, 325–348.

Chacko, G. & Viceira, L.M. (2001). Spectral GMM estimation of continuous-time processes. Working Paper, Harvard University, <http://www.people.hbs.edu/lviceira/cf62.pdf>.

Csörgő, S. (1981). Limit behaviour of the empirical characteristic function. *Ann. Probab.* **9**, 130–144.

DANIELSSON, J. (1994). Stochastic volatility in asset prices: Estimation with simulated maximum likelihood. *J. Econometrics* **64**, 375–400.

DUFFIE, D., PAN, J. & SINGLETON, K.J. (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica* **68**, 1343–1376.

DURBIN, J. (2000). The state space approach to time series analysis and its potential for official statistics (with discussion). *Aust. N. Z. J. Stat.* **42**, 1–23.

DURBIN, J. & KOOPMAN, S.J. (2000). Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives (with discussion). *J. Roy. Stat. Soc. Ser. B* **62**, 3–56.

ENGLE, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**, 987–1007.

FEUERVERGER, A. (1990). An efficiency result for the empirical characteristic function in stationary time-series models. *Canad. J. Statist.* **18**, 155–161.

FEUERVERGER, A. & MUREIKA, R.A. (1977). The empirical characteristic function and its applications. *Ann. Statist.* **5**, 88–97.

FRÜHWIRTH-SCHNATTER, S. (1994). Applied state space modelling of non-Gassian time series using integration-based Kalman filtering. *Statist. Comput.* **4**, 259–269.

GHYSELS, E., HARVEY, A.C. & RENAULT, E. (1996). Stochastic volatility. In *Handbook of Statistics, Vol. 14, Statistical Methods in Finance*, eds G.S. Maddala & C.R. Rao, pp. 119–191. Amsterdam: North-Holland.

HANSEN, L.P. & SCHEINKMAN, J.A. (1995). Back to future: Generating moment implications for continuous-time Markov processes. *Econometrica* **63**, 767–804.

HARVEY, A.C. (1998). Long memory in stochastic volatility. In *Forecasting Volatility in the Financial Markets*, eds J.L. Knight & S.E. Satchell, pp. 307–320. Oxford: Butterworth-Heineman.

HARVEY, A.C., RUIZ, E. & SHEPHARD, N. (1994). Multivariate stochastic variance models. *Rev. Economic Stud.* **61**, 247–264.

JACQUIER, E., POLSON, N.G. & ROSSI, P.E. (1994). Bayesian analysis of stochastic volatility models. *J. Bus. Econom. Statist.* **12**, 371–389.

JIANG, G. & KNIGHT, J.L. (2002). Estimation of continuous time processes via the empirical characteristic function. *J. Bus. Econom. Statist.* **12**, 198–212.

KIM, S., SHEPHARD, N. & CHIB, S. (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *Rev. Economic Stud.* **65**, 361–393.

KNIGHT, J.L. & SATCHELL, S.E. (1997). The cumulant generating function estimation method. *Econom. Theory* **13**, 170–184.

KNIGHT, J.L. & YU, J. (2002). The empirical characteristic function in time series estimation. *Econom. Theory* **18**, 691–721.

MELINO, A. & TURNBULL, S.M. (1990). Pricing foreign currency options with stochastic volatility. *J. Econometrics* **45**, 239–265.

NELSON, D.B. (1988). Time Series Behavior of Stock Market Volatility and Returns. PhD Dissertation, Massachusetts Institute of Technology.

POWELL, M.J.D. (1964). An efficient method for finding the minimum of a function of several variables without calculating derivatives. *Comput. J.* **7**, 155–162.

QUANDT, R.E. & RAMSEY, J.B. (1978). Estimating mixtures of normal distributions and switching regressions. *J. Amer. Statist. Assoc.* **73**, 730–738.

RUIZ, E. (1994). Quasi-maximum likelihood estimation of stochastic volatility models. *J. Econometrics* **63**, 289–306.

SINGLETON, K.J. (2001). Estimation of affine asset pricing models using the empirical characteristic function. *J. Econometrics* **102**, 111–141.

TRAN, K.C. (1998). Estimating mixtures of normal distributions via empirical characteristic function. *Econom. Rev.* **17**, 167–83.

YU, J. (1998). The Empirical Characteristic Function in Time-Series Estimation. PhD Dissertation, The University of Western Ontario.