Efficient Estimation of the Stochastic Volatility Model by the Empirical Characteristic Function Method

John L. Knight[1]

Department of Economics
The University of Western Ontario
London, Ontario
Canada N6A 5C2

Stephen E. Satchell
Trinity College &
University of Cambridge
Sidgewick Avenue, Cambridge
United Kingdom CB1 9DE

Jun Yu
Department of Economics
University of Auckland
Auckland, New Zealand

November 10, 1999

Abstract

This paper estimates the stochastic volatility model using the empirical characteristic function method. This procedure has the same asymptotic efficiency as maximum likelihood, and is thus a desirable method to use when the likelihood function is unknown. The stochastic volatility model has no closed form for its likelihood but it does have a known characteristic function. A Monte Carlo study shows that the empirical characteristic function method is a viable procedure for the stochastic volatility model. An application is considered for S&P 500 daily returns. Our results suggest much lower persistence than is normally found.

JEL#: C13, C15, C22, G10
Key words and phrases: Empirical Characteristic Function, Stochastic Volatility

[1]This research was started while Yu and Knight were visiting Birkbeck College, University of London. We acknowledge the UWO Department of Statistical and Actuarial Sciences for providing the computer facilities. We would like to thank participants of the Applied Econometrics Workshop at the University of Cambridge for useful comments and discussion, especially A. Harvey and D. Robinson. We would like to thank G. Tauchen for kindly supplying the dataset used in Section 4. Also the first author acknowledges financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC).

1 Introduction

Modeling the volatility of financial and macroeconomic time series has attracted a lot of attention since the introduction of autoregressive conditional heteroskedasticity (ARCH) (Engle (1982)). A feature of the ARCH type model is that the conditional variance is driven by the past observable variables. As an alternative setup to the ARCH-type model, the Stochastic Volatility (SV) model is supposed to describe the financial time series better than the ARCH-type model, since it essentially involves two noise processes. This added dimension makes the model more flexible, for example, the SV model can explain not only volatility clustering but also leverage effects. For further discussion, see Ghysels, Harvey, and 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. It is a consequence of this that direct maximum-likelihood estimation is impossible. Therefore, alternative estimation methods to the maximum likelihood have been proposed to estimate the SV models, which we discuss next.

Melino and Turnbull (1990) use generalized method of moments (GMM) for the discrete SV model. A more efficient GMM is proposed by Andersen and Sorensen (1993) . For the continuous time SV model, a GMM approach is developed by Hansen and 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 and Shephard (1994). The main idea is to treat non Gaussian disturbances as if they are normal and then maximize the quasi likelihood function. Often estimation methods involve the whole family of simulation based methods, including simulated MM/GMM proposed by Duffie and Singleton (1993), indirect inference proposed by Gouriéroux, Monfort and Renault (1993), simulated maximum likelihood (SML) proposed by Danielsson (1994b), and Markov Chain Monte Carlo (MCMC) proposed by Jacquier, Polson and Rossi (1994). The SV model has become a central model to describe financial time series and to

compare the relative merits of estimation procedures.

Although most of these methods are consistent under appropriate regularity conditions, in general they are not efficient. For example, by using only a finite number of moment conditions, MM/GMM may ignore 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. The simulation based methods decrease the efficiency by introducing an extra random error. This raises the question as to whether we can find a methodology with efficiency equivalent to maximum likelihood.

The present paper uses such an alternative approach to estimate the stochastic volatility model – via the empirical characteristic function. The rationale for using the characteristic function is that there is a one to one correspondence between the characteristic function and the distribution function. Consequently, the empirical characteristic function (ECF) should contain the same amount of information as the empirical distribution function (EDF). Theoretically, therefore, inference based on the characteristic function should perform as well as inference based on the likelihood. Moreover, by using the characteristic function, we can overcome the difficulties arising from ignorance of the true density function or the true likelihood function. The paper is organized as follows. The next section introduces a canonical 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 characteristic function of the SV model is obtained as well. Section 4 discusses the implementation of the ECF method as well as a Monte Carlo study and an empirical application. All proofs are collected in the Appendix.

2 The Model

The formulation of the discrete time stochastic volatility model is similar to that of the ARCH-type models. That is, the conditional variance is directly modeled. However,

in contrast to the ARCH-type models, the stochastic volatility model allows a random component in the transition equation. By doing so, the model can explain why large changes can follow stable periods. The model is of the form,

$$x_t = \sigma_t e_t, \ t = 1, 2, \cdots, T, \tag{2.1}$$

where σ_t^2 is the conditional variance based on the information at the end of time t, and e_t is a series of i.i.d. random disturbances which are assumed to have a standard normal distribution. We define

$$\sigma_t = \exp(0.5 h_t) \tag{2.2}$$

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

$$h_t = \lambda + \alpha h_{t-1} + v_t, \ v_t \sim iidN(0, \sigma^2), \tag{2.3}$$

where $\boldsymbol{\theta} = (\alpha, \lambda, \sigma^2)$ 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 will appear 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. Without including the random component, the transition equation is deterministic and the model exhibits time-varying but deterministic volatility. Finally, we assume e_t and v_t are uncorrelated, we shall return to this assumption later.

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 h_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 $E(x_t^2) = \exp(\frac{\sigma^2}{2(1-\alpha^2)})$, and $E(x_t^4) = 3\exp(\frac{2\sigma^2}{1-\alpha^2})$. The kurtosis of x_t is therefore $3\exp(\frac{\sigma^2}{1-\alpha^2})$, so x_t exhibits more kurtosis than a constant variance normal model. Furthermore, Harvey (1993) derives the moments of powers of the absolute

value of x_t,

$$E|x_t|^c = 2^{c/2}\frac{\Gamma(c/2+1/2)}{\Gamma(1/2)}\exp(\frac{c^2\sigma^2}{8(1-\alpha^2)}),\ c>-1, c\neq 0, \tag{2.4}$$

and

$$Var|x_t|^c = 2^c\left\{\frac{\Gamma(c+\frac{1}{2})}{\Gamma(1/2)} - \left[\frac{\Gamma(\frac{c}{2}+\frac{1}{2})}{\Gamma(1/2)}\right]^2\right\}\exp(\frac{c^2\sigma^2}{2(1-\alpha^2)}),\ c>-\frac{1}{2}, c\neq 0. \tag{2.5}$$

Since x_t is a non-linear function of an AR(1) process, however, the process is difficult to work with. For example, there is no closed form expression for the characteristic function of x_t. 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

$$y_t = \log\sigma_t^2 + \log e_t^2 = h_t + \epsilon_t,\ t = 1, 2, \cdots, T, \tag{2.6}$$

where $\epsilon_t = \log e_t^2$ is the logarithm of the chi square 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.

Unfortunately, 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 (2.6) we know that y_t is the convolution of an AR(1) process and an iid logarithmic $\chi^2_{(1)}$ sequence, and hence there is a closed form expression for the characteristic function of y_t which we will 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, inference based on the ECF can achieve efficiency.

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

$$c(r,\boldsymbol{\theta}) = E[\exp(irx)] = \int \exp(irx)\, dF(x;\boldsymbol{\theta}), \tag{3.1}$$

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

$$c_n(r) = \frac{1}{n}\sum_{j=1}^{n} \exp(irx_j) = \int \exp(irx)\, dF_n(x). \tag{3.2}$$

where $F_n(x)$ is the empirical distribution function. Therefore, the CF and ECF are the Fourier transformations of the DF and EDF. Because of the uniqueness of the Fourier-Stieltjes transformation, the CF has the same information as the DF 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 discrete $r_1, \cdots, r_q$, we can minimize the following distance,

$$\sum_{j=1}^{q} |c_n(r_j) - c(r_j;\boldsymbol{\theta})|^2 g(r_j), \tag{3.3}$$

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

$$\int |c_n(r) - c(r;\boldsymbol{\theta})|^2 g(r)\, dr, \tag{3.4}$$

i.e., the distance over an interval. In both cases $g(\cdot)$ is a weight function.

If the observations are an iid sequence, the marginal EDF contains all the information in the sample and so does the marginal ECF in (3.2). In econometrics, previous authors have followed this approach. In order to estimate the mixture of normals, Quandt and Ramsey (1978) give an ordinary least square (OLS) procedure and Schmidt (1982) gives a generalized least square (GLS) procedure where they use

the moment generating function instead of the CF. However, the ECF can be used in the same way, see Tran (1994) . Moreover, there are known convergence results for the empirical characteristic function process $\sqrt{n}(c_n(r) - c(r;\boldsymbol{\theta}))$. These have been established by Fuerverger and Mureika (1977), and Csörgö (1981) for any iid sequence. Furthermore, the procedure can be generalized to stochastic processes, see Feuerverger (1990).

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

$$z_j = (y_j, \cdots, y_{j+p})', \ j = 1, \cdots, T-p, \tag{3.5}$$

and the characteristic function of each block is basically a joint one and is defined as

$$c(\boldsymbol{r}, \boldsymbol{\theta}) = E(\exp(i\boldsymbol{r}'z_j)), \tag{3.6}$$

where $\boldsymbol{r} = (r_1, \cdots, r_{p+1})$, $\boldsymbol{\theta} = (\alpha, \sigma^2, \lambda)$. The ECF is defined as

$$c_n(\boldsymbol{r}) = \frac{1}{n}\sum_{j=1}^{n} \exp(i\boldsymbol{r}'z_j), \tag{3.7}$$

where $n = T - p$.

Several estimation procedures are proposed by Knight and Yu (1997). The common feature of the estimation procedures is to match the ECF with the CF. That is

$$\min \int \cdots \int |c(\boldsymbol{r}, \boldsymbol{\theta}) - c_n(\boldsymbol{r})|^2 g(\boldsymbol{r})\, dr_1 \cdots dr_{p+1}, \tag{3.8}$$

or

$$\int \cdots \int (c(\boldsymbol{r}, \boldsymbol{\theta}) - c_n(\boldsymbol{r}))w(\boldsymbol{r})\, dr_1 \cdots dr_{p+1} = 0, \tag{3.9}$$

where both $g(\boldsymbol{r})$ and $w(\boldsymbol{r})$ are weight functions. The weighted distance between the ECF and the CF is minimized in (3.8), while (3.9) is the first order condition of

(3.8). Since these two methods are equivalent, we only consider the procedure based on (3.8). A key point to note here is that our calculations are with respect to the unconditional (steady-state) joint CF of z_j. One could, of course, do the calculation with the conditional CF as an alternative.

Knight and Yu (1997) present four different versions of the ECF methods. We discuss them in detail. Firstly, when the transformation variables, $\boldsymbol{r}$'s, are chosen discretely and the weight is optimal, the procedure is referred as the the discrete ECF method by Knight and Yu (1997). Hence the weight function $g(\boldsymbol{r})$ in this case is a function with a certain number of jumps. Fuerverger (1990) proves that under some regularity conditions, if p is sufficiently large and the jumps are sufficiently fine and extended, the resulting estimators can achieve the Cramér-Rao lower bound. His results are theoretical and involve no empirical calculations. Considering the estimation of time series by using the discrete ECF method, Knight and Satchell (1996) give a multi-step procedure. The main idea is the following. We first choose moving blocks (i.e. p) such that the most important information of the original sequence is retained by the blocks. Next a function with q jumps is used to be the weight function. When the size of each jump is chosen optimally, minimization of (3.8) boils down to the GLS technique where the ECF is regressed on the CF over a finite number of $\boldsymbol{r}$'s. Equivalently, the ECF is matched with the CF on q discrete points. Unfortunately, several practical questions arise in this procedure. For example, we do not know how to choose the number of jumps and how to make the jumps optimal. Since each choice of $\boldsymbol{r}$ corresponds to a moment condition in the CF, the estimator, in essence, is a GMM estimator of CF.

To overcome the difficulties, we can choose the transformation variables continuously. The procedure is called the continuous ECF method. In the this method, the transformation variables are simply integrated out. By choosing the transformation variables continuously, the procedure basically matches the ECF and the CF over an interval, and hence match all the moments continuously. If an equal weight is chosen, for example, $g(\boldsymbol{r}) = 1$, the procedure is the OLS of the continuous ECF method.

When a non equal weight is used, the procedure corresponds to the WLS (weighted least squares) of the continuous ECF method. Furthermore, when the weight is chosen optimally, the procedure is called the GLS of the continuous ECF method because the resulting estimators can achieve the Cramér-Rao lower bound. For example, the optimal weight function in (3.9), $w(\boldsymbol{r})$, is given below,

$$w(\boldsymbol{r}) = \int \cdots \int \exp(-i\boldsymbol{r}'\boldsymbol{z_j}) \frac{\partial \log f(y_{j+p}|y_j, \cdots, y_{j+p-1})}{\partial \boldsymbol{\theta}} dy_j \cdots dy_{j+p}, \tag{3.10}$$

where $f(\cdot)$ is the conditional probability density function (PDF) of the data. However, this quantity is not calculable if the PDF is unknown, as is our case for the SV model.

Under standard regularity conditions (see Section 2.3 in Knight and Yu (1997)), Knight and Yu (1997) established the strong consistency and asymptotic normality for the resulting estimators when the above procedures are used to estimate a strictly stationary process. Furthermore, by studying the finite sample properties of the ECF estimators, Knight and Yu (1997) note that the continuous ECF method performs better than the discrete ECF method.

Regarding the choice of p, we note that the blocks always contain no less information as p increases and thus the resulting estimators are supposed to be more efficient. However, calculations associated with larger p are numerically more difficult. A small p involves less computation, and the blocks may not retain all important information. For a Markov process such as a Gaussian AR(1) process, fortunately, Knight and Yu (1997) found that blocks with $p = 1$ is enough to capture virtually all the information in the original series.

The model we are going to estimate via the ECF is the one defined by (2.6) since we can derive the closed form of the characteristic function. In order to use the ECF method, of course, we need to find the expression of the joint characteristic function. First, the characteristic function for the logarithm of the $\chi^2_{(1)}$ distribution is given in Lemma 1. And then the joint characteristic function for $y_t, \cdots, y_{t+k-1}$ is obtained in Lemma 2.

Lemma 3.1 *Suppose ϵ_t is the logarithm of the $\chi^2_{(1)}$ distribution. The characteristic function of ϵ_t, $c(r)$, is*

$$c(r) = \frac{\Gamma(0.5+ir)}{\Gamma(0.5)}2^{ir}, \tag{3.11}$$

where i is the imaginary number defined by $\sqrt{-1}$.

Proof: See Appendix A.

Lemma 3.2 *Suppose $\{y_t\}_{t=1}^T$ is defined by (2.6). The joint characteristic function of $y_t, y_{t+1}, \cdots, y_{t+k-1}$ is*

$$\begin{aligned} c(r_1, \cdots, r_k, \boldsymbol{\theta}) &= \exp[\frac{i\lambda}{1-\alpha}\sum_{j=1}^{k} r_j - \frac{\sigma^2}{2(1-\alpha^2)}(\sum_{j=1}^{k} r_j^2 + 2\alpha\sum_{l=1}^{k}\sum_{j=l+1}^{k}\alpha^{j-l-1}r_l r_j)] \\ &\quad \frac{\prod_{j=1}^{k}\Gamma(\frac{1}{2}+ir_j)}{\Gamma^k(\frac{1}{2})}2^{i\sum_{j=1}^{k}r_j}. \end{aligned} \tag{3.12}$$

Proof: See Appendix B.

Using the joint characteristic function we can easily obtain the joint cumulant generating function and consequently the autocorraltion function. The autocorrelation function of $\{y_t\}_{t=1}^T$ is given in the following lemma.

Lemma 3.3 *Suppose $\{y_t\}_{t=1}^T$ is defined by (2.6). The autocorrelation function of $\{y_t\}_{t=1}^T$ is,*

$$\rho_k = \frac{\frac{\alpha^k\sigma^2}{1-\alpha^2}}{\frac{\sigma^2}{1-\alpha^2} + \frac{\Gamma''(0.5)}{\Gamma(0.5)} - (\frac{\Gamma'(0.5)}{\Gamma(0.5)})^2}, \ k = 1, 2, \cdots. \tag{3.13}$$

Proof: See Appendix C.

The y_t process defined by (2.6) is the sum of an AR(1) and white noise, it is well-known that the result is ARMA(1,1). This is confirmed by the formula in (3.13).

In order to use the ECF method to estimate the SV model (2.6), we have to choose a value for p. Although our process is not Markovian, being an ARMA(1,1), we shall choose $p = 1$ at first. Our reason relates to the results of a Monte Carlo study by Knight and Yu (1997) where the ECF method is used to estimate an ARMA(1,1)

process, they found that $p = 1$ works quite well (see Table 8 in their paper). For the SV model (2.6) we note that with $p = 1$,

$$c(r_1, r_2, \boldsymbol{\theta}) = \exp[i\lambda\frac{r_1 + r_2}{1-\alpha} - \frac{\sigma^2}{2(1-\alpha^2)}(r_1^2 + 2\alpha r_1 r_2 + r_2^2)] \frac{\Gamma(\frac{1}{2} + ir_1)\Gamma(\frac{1}{2} + ir_2)}{\Gamma^2(\frac{1}{2})} 2^{ir_1 + ir_2}, \tag{3.14}$$

and

$$c_n(r_1, r_2) = \frac{1}{n}\sum_{j=1}^{n} \exp(ir_1 y_j + ir_2 y_{j+1}). \tag{3.15}$$

Defining $Re\, c(r_1, r_2, \boldsymbol{\theta})$, $Re\, c_n(r_1, r_2)$, $Im\, c(r_1, r_2, \boldsymbol{\theta})$ and $Im\, c_n(r_1, r_2)$ to be the real and imaginary parts of $c(r_1, r_2)$ and $c_n(r_1, r_2)$ respectively, we have,

$$Re\, c_n(r_1, r_2) = \frac{1}{n}\sum_{j=1}^{n} cos(r_1 y_j + r_2 y_{j+1}), \tag{3.16}$$

and

$$Im\, c_n(r_1, r_2) = \frac{1}{n}\sum_{j=1}^{n} sin(r_1 y_j + r_2 y_{j+1}). \tag{3.17}$$

As we mentioned, a clear advantage of choosing the transformation variable continuously is that we do not need to choose q. Furthermore, since the Monte Carlo study conducted by Knight and Yu (1997) shows that the continuous version works better than the discrete version, we use the continuous version to estimate the SV model. However, the optimal weight function in the continuous version is not readily obtained because the conditional score function has no closed form expression for the SV model. Instead the exponential function is considered. The exponential function is chosen because it puts more weight on the points around the origin, consistent with the recognition that the CF contains the most information around the origin. Therefore, the procedure is to choose $(\hat{\alpha}, \hat{\sigma}^2, \hat{\lambda})$ to minimize,

$$\int\int \Big[(Re\, c(r_1, r_2, \boldsymbol{\theta}) - \tfrac{1}{n}\textstyle\sum_{j=1}^{n} cos(ir_1 y_j + ir_2 y_{j+1}))^2 + (Im\, c(r_1, r_2, \boldsymbol{\theta}) - \tfrac{1}{n}\textstyle\sum_{j=1}^{n} sin(ir_1 y_j + ir_2 y_{j+1}))^2\Big] \exp(-ar_1^2 - ar_2^2)\, dr_1 dr_2, \tag{3.18}$$

where $c(r_1, c_2)$ is given by (3.14) and a is an arbitrary positive constant which is chosen to be 32.5 in the Monte Carlo studies and the application.

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

$$\frac{1}{n}\left\{\int\int\left[\frac{\partial Re\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial Re\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T}+\frac{\partial Im\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial Im\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T}\right]g(\boldsymbol{r})d\,r_1 d\,r_2\right\}^{-1}\times$$
$$A(\theta)\times\left\{\int\int\left[\frac{\partial Re\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial Re\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T}+\frac{\partial Im\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial Im\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T}\right]g(\boldsymbol{r})d\,r_1 d\,r_2\right\}^{-1}.$$

In Appendix D the expression for $A(\theta)$ is given along with a proof of the above result.

We should note that the joint characteristic function of y_t is of different functional form if v_t and e_t are correlated. However, the ECF method can be still used in the same way as the uncorrelated case. Of course, the the joint characteristic function of y_t in the correlated case is more difficult.

4 Implementation, Simulation and Application

4.1 Implementation

The implementation of the ECF method essentially requires minimizing (3.18), and thus involves double integrals. Unfortunately, no analytical solutions for either the double integrals or the optimization are available. Consequently, we will numerically evaluate the multiple integral (3.18), followed by numerical minimization of (3.18) with respect to $\boldsymbol{\theta}$. The numerical solutions are the desired estimators.

A 40-points Gauss-Kronod algorithm is used to approximate the two dimensional integrations in (3.18). Since there is no analytical expression for the derivative of the objective functions, a quasi-Newton method is used to find the minimum. The starting point in the optimization is chosen to be the quasi-maximum likelihood estimates proposed by Ruiz (1994). All computations were done in double precision.

By using the implementation procedure, we examine the performance of the ECF method in the estimation of a SV model in a Monte Carlo study. We also apply the procedure to a real dataset.

4.2 Monte Carlo Simulation

The Monte Carlo study is designed to check the viability of the ECF method. We choose the same parameter setting as Jacquier, Polson and Rossi (1994) did in one of their Monte Carlo studies, that is, $\alpha = 0.9$, $\sigma = 0.3629$, $\lambda = -0.736$. The number of observations is set at $T = 2,000$ and the number of replications is set at 500.

Table 1 reports the simulation results. The table shows the mean, the minimum, the maximum, the mean square error (MSE) and the root mean square error (RMSE) for all three estimators, and serves to illustrate that the ECF method works well.

In Table 2 we duplicates the results in Table 9 of Jacquier, Polson and Rossi (1994), where the same experimental design is used but the three alternative methods are employed, ie, the GMM, QML and MCMC. We also report the simulation results provided by Danielsson (1994b) based on the SML for the same experiment. Of course our random numbers may not be the same as those generated by Jacquier, Polson and Rossi (1994) and by Danielsson (1994). However, we believe that the experiments should be comparable. The finite sample performance of the ECF method is better than that of the QML and GMM method, while the MCMC and SML method outperform the ECF method. This can be accounted for by the use of the non-optimal weight function.

4.3 Application

4.3.1 Data

The data we used was supplied by George Tauchen and is the same as that used by Danielsson (1994b). It consists of eight years (2,022 observations) of daily geometric returns (defined as $100(\log P_{t+1} - \log P_t)$) for the S&P 500 index covering period 1980-1987. The data are adjusted as detailed in Danielsson (1994b).

4.3.2 Empirical Results

The empirical results are reported in Table 3, along with the MCMC estimates obtained by Jacquier, Polson and Rossi (1994) using the same data set. To obtain the ECF

estimates the initial values are chosen to be the QML estimates, as well as the MCMC estimates and other starting values. This serves to show that the global minimum is achieved. From Table 3, we note that the ECF estimates are significantly different from the MCMC estimates. For example, $\hat{\alpha}_{ECF}$ is close to 0 while the $\hat{\alpha}_{MCMC}$ is close to 1; $\hat{\sigma}^2_{ECF}$ is 20 times larger than $\hat{\sigma}^2_{MCMC}$. Since the empirical results are so different, the comparison of the goodness of fit is of particular interest.

To compare the goodness of fit, we carry out a number of calculations. Firstly, we simulate two sequences by using the ECF estimates and the MCMC estimates. In Figure 1, we plot the empirical density of the real data and densities of two simulated data sets; these correspond to the steady-state density of the process given by (2.1) to (2.3) (assuming $|\alpha| < 1$). Figure 1 clearly demonstrates that the ECF estimates have better goodness of fit than the MCMC estimates. The Kolmogorov-Smirnov test is performed to test for the goodness of fit. The results are reported in Table 4. For the ECF the Kolmogorov-Smirnov test statistic (0.0257) is much smaller and the p-value (0.498) is very large while the p-value for the MCMC is 0. The MCMC has been rejected at any significant level and the ECF can not be rejected. Therefore, the ECF method is significantly better than the MCMC in the sense of fitting the steady state distribution. This result is very intuitive because the ECF method basically matches all the moments and hence the density.

We next discuss the implication of our results. Firstly, a much smaller α implies smaller volatility clustering, that is less persistence. Consequently, there is not much dependence for the variances between two consecutive trading days. This contrasts with the implication of large α. Secondly, a much larger σ^2 suggests that a large change can possibly follow a stable period and a stable period can follow an unstable period. This happens because with the large variance the random innovation v_t may dominate the deterministic term and hence bring in a significant change. Thus whilst the estimated models have similar means their persistence characteristics seem dramatically different. Evaluating our objective function, given by (3.18), for the two sets of converged estimates in Table 3 we find that the ECF estimates results in a value of

1.6354×10^{-7} while the MCMC eatimates gives 9.73823×10^{-5}. The latter is nearly 600 times larger than the former!

In response to suggestions, we chose larger values of p for the ECF method. Theoretically, we know that with a larger p the moving blocks preserve more information and hence the ECF method can be more efficient. On the other hand, however, a larger p is computationally more time-consuming since higher dimensional numerical integrations are involved. In Table 5, we report the ECF estimates for $p = 2, 3, 4, 5$ where we fit the SV model to the same data set. From this table we note that the empirical results remain almost unchanged for different values of p and are very close to those for $p = 1$.

Since our results are so different from everybody else's we carried out further analysis. We applied the test procedures discussed in Harvey and Streibel (1998). These are methods to test if there is evidence of a unit root in equation (2.3). We report our results in Table 6. The three procedures discussed in their paper provide us with conflicting estimates, namely the von Neumann Ratio (VNR) and chi-squared tests applied to y_t and x_t are not significant for both processes although the $Q(p)$ test, with $p = 45$ is found to be significant for both series. Since the chi-squared test is the one demonstrated by Harvey and Streibel to be the most powerful and since the parametric values found by us are a long way from being a local alternative, we find the above evidence in favour of the traditional high alpha values but not compellingly so.

Finally, we consider a reverse calculation where we evaluate our parameter values (Table 3) with those of Danielsson [1994]. We call these θ_1 and θ_2 respectively. If the likelihood at θ_1 is larger than the likelihood at θ_2, then this, coupled with our finding noted earlier that our objective function at θ_1 is 600 times smaller than θ_2, would be strong supportive evidence that our results are not aberrant but an alternative local (and possibly, global) maximum.

Since the likelihood is not known analytically, we need to resort to numerical procedures. We present details of our likelihood calculations in Appendix E. The procedure we follow is to treat h_t as observable and condition on the sequence $(h_1, ..., h_t)$. The

conditional likelihood can be calcuated and by averaging over different histories of h_t we can get an estimate of the likelihood of y_t.

The values we calculate are based on 500 replications of the history and we set h_0 equal to $\frac{1}{T}\sum_{t=1}^{T} \ln x_t^2 - E(\ln(x_{(1)}^2))$. The values of the two likelihoods are -2270 and -2290 respectively for θ_1 and θ_2. These show supporting evidence for our conjecture that our estimators are the argument of an alternative maximum.

Appendix A

Proof of Lemma 1

According to the definition of the characteristic function, and with $\epsilon_t = \log(\chi^2_{(1)})$, we have

$$\begin{aligned}
c(r) &= E[\exp(ir\epsilon_t)] \\
&= E[(\exp(\epsilon_t))^{ir}] \\
&= E[(\exp(\log(\chi^2_{(1)})))^{ir}] \\
&= E[(\chi^2_{(1)})^{ir}] \\
&= \int x^{ir} f(x)\, dx \\
&= \int x^{ir} \frac{1}{\Gamma(1/2)\sqrt{2}} x^{-1/2} e^{-x/2}\, dx \\
&= \int x^{ir-\frac{1}{2}} \frac{1}{\Gamma(1/2)\sqrt{2}} e^{-x/2}\, dx \\
&= \frac{\Gamma(\frac{1+2ir}{2}) 2^{\frac{1+2ir}{2}}}{\Gamma(\frac{1}{2}) 2^{\frac{1}{2}}} \\
&= \frac{\Gamma(\frac{1}{2}+ir) 2^{ir}}{\Gamma(\frac{1}{2})}, \qquad \text{(A.1)}
\end{aligned}$$

where $f(x)$ is the density function of $\chi^2_{(1)}$ and is given by

$$f(x) = \frac{1}{\Gamma(1/2)\sqrt{2}} x^{-1/2} e^{-x/2}. \quad \blacksquare$$

Appendix B

Proof of Lemma 2

Since y_t is a convolution of a Gaussian AR(1) process and an iid sequence with $\chi^2_{(1)}$ distribution, we have

$$
\begin{aligned}
c(r_1,\cdots,r_k,\boldsymbol{\theta}) &= E[\exp(ir_1y_t + ir_2y_{t+1} + \cdots + ir_ky_{t+k-1})] \\
&= E[\exp(ir_1h_t + ir_1\epsilon_t + ir_2h_{t+1} + ir_2\epsilon_{t+1} + \cdots + ir_kh_{t+k-1} + ir_k\epsilon_{t+k-1}] \\
&= E[\exp(ir_1h_t + ir_2h_{t+1} + \cdots + ir_kh_{t+k-1}] \prod_{j=1}^{k} E[\exp(ir_j\epsilon_{t+j-1})] \\
&= E[\exp(ih_t\sum_{j=1}^{k}\alpha^{j-1}r_j + i\lambda\sum_{j=2}^{k}\frac{1-\alpha^{j-1}}{1-\alpha}r_j + \sum_{l=2}^{k}\nu_{t+l-1}\sum_{j=l}^{k}r_j\alpha^{j-l})] \\
&\quad \prod_{j=1}^{k} E[\exp(ir_j\epsilon_{t+j-1})] \\
&= \exp[i\frac{\lambda}{1-\alpha}\sum_{j=1}^{k}\alpha^{j-1}r_j + i\lambda\sum_{j=2}^{k}\frac{1-\alpha^{j-1}}{1-\alpha}r_j - \frac{1}{2}(\sum_{j=1}^{k}\alpha^{j-1}r_j)^2\frac{\sigma^2}{1-\alpha^2} \\
&\quad -\frac{1}{2}\sum_{l=2}^{k}(\sum_{j=l}^{k}\alpha^{j-l}r_j)^2\sigma^2]\frac{\prod_{j=1}^{k}\Gamma(\frac{1}{2}+ir_j)}{\Gamma^k(\frac{1}{2})}2^{i\sum_{j=1}^{k}r_j} \\
&= \exp[\frac{i\lambda}{1-\alpha}\sum_{j=1}^{k}r_j - \frac{\sigma^2}{2(1-\alpha^2)}(\sum_{j=1}^{k}r_j^2 + 2\alpha\sum_{l=1}^{k}\sum_{j=l+1}^{k}\alpha^{j-l-1}r_lr_j)] \\
&\quad \frac{\prod_{j=1}^{k}\Gamma(\frac{1}{2}+ir_j)}{\Gamma^k(\frac{1}{2})}2^{i\sum_{j=1}^{k}r_j}. \quad \blacksquare
\end{aligned}
$$

Appendix C

Proof of Lemma 3

Defined as the logarithm of the CF, the cumulant function is of the form,

$$\begin{aligned}\phi(r_1,\cdots,r_k) &= \log(c(r_1,\cdots,r_k,\boldsymbol{\theta})) \\ &= \frac{i\lambda}{1-\alpha}\sum_{j=1}^{k} r_j - \frac{\sigma^2}{2(1-\alpha^2)}(\sum_{j=1}^{k} r_j^2 + 2\alpha\sum_{l=1}^{k}\sum_{j=l+1}^{k}\alpha^{j-l-1}r_l r_j) \quad \text{(C.1)} \\ &\quad + \sum_{j=1}^{k}\log(\Gamma(\frac{1}{2}+ir_j)) + i\log 2\sum_{j=1}^{k} r_j.\end{aligned}$$

Therefore, we have

$$\begin{aligned}var(y_t) &= \frac{\partial^2\phi(r_1,\cdots,r_k)}{\partial r_1^2}|_{r_1=0} \\ &= \frac{\sigma^2}{1-\alpha^2} + \frac{\Gamma''(0.5)}{\Gamma(0.5)} - (\frac{\Gamma'(0.5)}{\Gamma(0.5)})^2,\end{aligned}$$

and

$$\begin{aligned}cov(y_t, y_{t+k}) &= \frac{\partial^2\phi(r_1,\cdots,r_k)}{\partial r_1\partial r_k}|_{r_1=0,r_k=0} \\ &= \frac{\alpha^k\sigma^2}{1-\alpha^2},\ k=1,2,\cdots\end{aligned}$$

Hence the autocorrelation functions are,

$$\rho_k = \frac{\frac{\alpha^k\sigma^2}{1-\alpha^2}}{\frac{\sigma^2}{1-\alpha^2} + \frac{\Gamma''(0.5)}{\Gamma(0.5)} - (\frac{\Gamma'(0.5)}{\Gamma(0.5)})^2},\ k=1,2,\cdots \quad \blacksquare$$

Appendix D

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

$$\hat{\theta} = \arg\min s(\theta)$$

and

$$\begin{aligned} s(\theta) &= \int \cdots \int \{(Re\ c_n(r) - Re\ c(r,\theta))^2 \\ &\quad +(Im\ c_n(r) - Im\ c(r,\theta))^2\} dG(r) \end{aligned}$$

Now since

$$Re\ c_n(r) = \frac{1}{n}\sum_{j=1}^{n} \cos r'x_j$$

and

$$Im\ c_n(r) = \frac{1}{n}\sum \sin r'x_j$$

Then

$$\frac{\partial s(\theta)}{\partial \theta} = -\frac{2}{n}\sum_{j=1}^{n} k_j(\theta)$$

where

$$\begin{aligned} k_j(\theta) &= \left[\int \cdots \int \frac{\partial Re\ c(r,\theta)}{\partial \theta}(\cos r'x_j - Re\ c(r,\theta)) \right. \\ &\quad \left. +\frac{\partial Im\ c(r,\theta)}{\partial \theta}(\sin r'x_j - Im\ c(r,\theta))\right] dG(r) \end{aligned}$$

Consequently

$$\sqrt{n}\frac{\partial s(\theta)}{\partial \theta} \xrightarrow{d} N(0, 4A(\theta))$$

where

$$A(\theta) = \lim_{n\to\infty} E\left[\frac{1}{n}\sum_j \sum_k k_j(\theta)k_k(\theta)\right]$$

and is given by:

$$A(\theta) = \lim_{n\to\infty} \frac{1}{n}\left\{ \int \cdots \int \left[\frac{\partial Re\ c(r,\theta)}{\partial \theta}\frac{\partial Re\ c(s,\theta)}{\partial \theta}\sum_j \sum_k cov(\cos r'x_j, \cos s'x_k)\right.\right.$$

$$+\frac{\partial Re\ c(r,\theta)}{\partial\theta}\frac{\partial Im\ c(s,\theta)}{\partial\theta'}\sum_j\sum_k cov(\cos r'x_j, \sin s'x_k)$$
$$+\frac{\partial Im\ c(r,\theta)}{\partial\theta}\frac{\partial Re\ c(s,\theta)}{\partial\theta'}\sum_j\sum_k cov(\sin r'x_j, \cos s'x_k)$$
$$+\frac{\partial Im\ c(r,\theta)}{\partial\theta}\frac{\partial Im\ c(s,\theta)}{\partial\theta'}\sum_j\sum_k cov(\sin r'x_j, \sin s'x_k)\Big]\Big\}dG(r)dG(s)$$

The double summation covariance expressions are readily found and are given in the Lemma in Knight and Satchell [1997, p. 176]. That is, we note that

$$\begin{aligned}
&\sum_j\sum_k cov(\cos r'x_j \cos s'x_k)\\
&= n^2 cov(Re\ c_n(r), Re\ c_n(s))\\
&= n^2\cdot(\Omega_{RR})_{r,s},
\end{aligned}$$

using notation in Knight and Satchell [1997]. Similarly, for the other double sums.

Thus

$$\begin{aligned}
A(\theta) = \lim_{n\to\infty} n\Big\{\int\cdots\int\Big[&\frac{\partial Re\ c(r,\theta)}{\partial\theta}\frac{\partial Re\ c(s,\theta)}{\partial\theta'}\cdot(\Omega_{RR})_{r,s}\\
&+2\frac{\partial Re\ c(r,\theta)}{\partial\theta}\frac{\partial Im\ c(r,\theta)}{\partial\theta'}\cdot(\Omega_{RI})_{r,s}\\
&+\frac{\partial Im\ c(r,\theta)}{\partial\theta}\frac{\partial Im\ c(s,\theta)}{\partial\theta'}\cdot(\Omega_{II})_{r,s}\Big]\Big\}dG(r)dG(s)
\end{aligned}$$

Also we note that

$$\begin{aligned}
E\left[\frac{\partial^2 s(\theta)}{\partial\theta\partial\theta'}\right] &= -\frac{2}{n}\sum_{j=1}^n E\left[\frac{\partial k_j(\theta)}{\partial\theta}\right]\\
&= \frac{2}{n}\sum_{j=1}^n\int\cdots\int\Big[\frac{\partial Re\ c(r,\theta)}{\partial\theta}\frac{\partial Re\ c(r,\theta)}{\partial\theta'}\\
&\quad+\frac{\partial Im\ c(r,\theta)}{\partial\theta}\frac{\partial Im\ c(r,\theta)}{\partial\theta'}\Big]dG(r)\\
&= -2\int\cdots\int\Big[\frac{\partial Re\ c(r,\theta)}{\partial\theta}\frac{\partial Re\ c(r,\theta)}{\partial\theta'}\\
&\quad+\frac{\partial Im\ c(r,\theta)}{\partial\theta}\frac{\partial Im\ c(r,\theta)}{\partial\theta'}\Big]dG(r)\\
&= -2B(\theta)
\end{aligned}$$

Thus standard asymptotic theory results in

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

Appendix E

The numerical computation of the log likelihood ($\ln L$) for the stochastic volatility model is detailed below.

Recall (2.6) which states

$$y_t = h_t + \varepsilon_t$$

with

$$\varepsilon_t = \ln e_t^2.$$

Since $e_t^2 \sim \chi^2_{(1)}$ we can calculate i.e., the distribution of $h_t + \ln \chi^2_{(1)}$. This will be the *conditional* likelihood of y_t and time t. Thus the $\ln L$ function is given by

$$\begin{aligned} \ln L|h_t &= \ln \prod_{t=1}^{T} pdf(y_t|\sigma_t^2) \\ &= \sum_{t=1}^{T} \ln pdf(y_t|\sigma_t^2) \\ &= -\frac{T}{2} \ln 2\pi - \frac{1}{2} \sum_{t=1}^{T} [(y_t/\sigma_t^2) - \ln(y_t/\sigma_t^2)] \end{aligned}$$

To estimate the $\ln L$ associated with our data and with two alternative estimates of the parameters, λ, α, σ we proceed as follows:

i) Letting $\hat{\theta}_j - (\hat{\lambda}_j, \hat{\alpha}_j, \hat{\sigma}_j^2)$, $j = 1, 2$ be the two parameter sets. For each set we can generate a sequence of h_t's and consequently, σ_t^2 :i.e., $h_{jt} = \hat{\lambda}_j + \hat{\alpha}_j h_{jt-1} + \hat{\sigma}_j e_t, j - 1, 2$ with $e_t \sim iid\ N(0, 1)$, and setting h_0 to equal $\frac{1}{T} \sum y_t - E(\ln \chi^2_{(1)})$. Generate a sequence of e_i's $(t = 1, ..., T)$ and use these to generate the *two* sequences for h_i and σ_i^2, i.e.

$$\sigma_{jt}^2 = \exp(h_{jt})$$

ii) With the two generated h_t sequences, calculate the *two* corresponding $\ln L_j|h_{jt}$, $j = 1, 2$.

iii) Go to i) and repeat steps i) and ii) say M times.

iv) Calculate the $\ln L_j$ as the average of $\ln L_j|h_{jt}$ over the M replications, i.e.

$$\ln L_j = \frac{1}{M}\sum_{l=1}^{M} \ln L_{lj}|h_{ljt}$$

The calculated $\ln L_j$ from iv) will be approximations to the true $\ln L$. Each will have been evaluated at the converged parameter values from the two competing estimation methods.

Table 1

Monte Carlo Study of the *ECF* Method

True Values of Parameters $\alpha = 0.9$ $\sigma = 0.3629$, $\lambda = 0.736$

No. of Replications=500 No. of Observations=2,000

	$\alpha = 0.9$	$\sigma = 0.3629$	$\lambda = -0.736$
MEAN	.892	.3812	.-7962
MED	.895	.3763	-.774
MIN	.75	.1985	-1.843
MAX	.95	.6399	-.3401
RMSE	.03	.067	.231

Table 2

Monte Carlo Comparison of GMM, QML, MCMC, and SML Estimates[2]

True Values of Parameters $\alpha = 0.9$ $\sigma = 0.3629$, $\lambda = -0.736$

No. of Replications=500 No. of Observations=2,000

Method	$\alpha = 0.9$	$\sigma = 0.3629$	$\lambda = -0.736$
GMM	.88(.06)	.31(.10)	-.86(.42)
QML	.88(.06)	.383(.11)	-.853(.46)
MCMC	.896(.02)	.359(.034)	-.763(.15)
SML	.902(.02)	.359(.039)	-.721(.15)

[2]The table shows the mean and RMSE(in parentheses).

Table 3

Empirical Comparison of ECF and MCMC Estimates[3]

Method	α	σ	λ
ECF	-.0676	.747	-0.29
MCMC	0.97	.15	-.002

Table 4

Kolmogorov-Smirnov Goodness-of-Fit Test of ECF and MCMC Estimates

	KS statistic	p-value
ECF	0.0257	0.498
MCMC	0.0875	0

Table 5

Empirical Results of ECF Estimates with Different Values of p

Method	α	σ	λ
$p = 2$	-0.0719	0.743	-0.45
$p = 3$	-0.0822	0.740	-0.31
$p = 4$	-0.0927	0.738	-0.38
$p = 5$	-0.0742	0.731	-0.40

Table 6

	x_t	y_t	Control Value (5%)
VNR	3.71588	-0.39550	1.692
χ^2	0.05328	1.2129	0.461
$Q(p)$	82.5	169.6	61.7

[3]The empirical results for the MCMC method are obtained by Jacquier, Polson, and Rossi (1994).

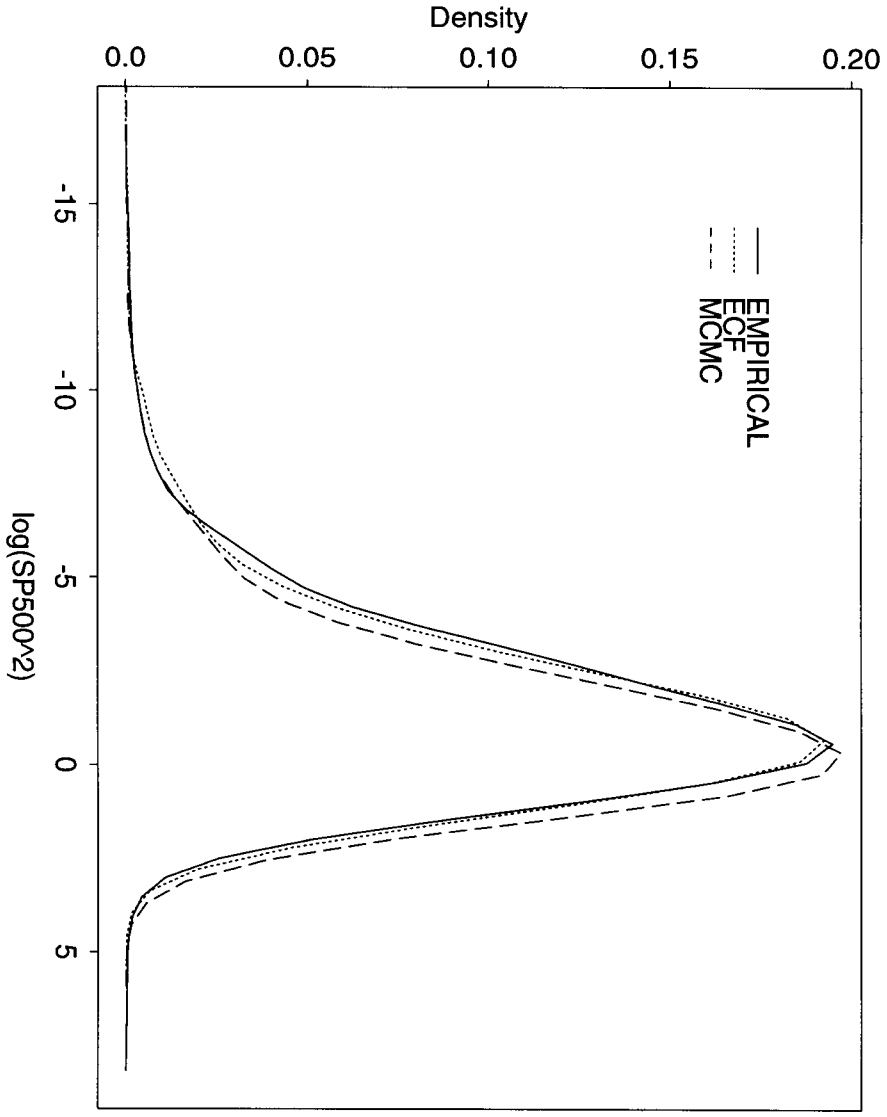
Density
0.0
0.05
0.10
0.15
0.20
log(SP500^2)
-15
-10
-5
0
5
EMPIRICAL
ECF
MCMC

Figure 1

References

[1] Andersen, T. and B. Sorensen. (1994). GMM estiomation of a stochastic volatility models: A Monte Carlo study. *Journal of Business and Economic Statistics.* **14** 329–352.

[2] Csörgő, S. (1981). Limit behaviour of the empirical characteristic function. *The Annals of Probability.* **9** 130–144.

[3] Danielsson, J. (1994). Stochastic volatility in asset prices. Estimation with simulated maximum likelihood. *Journal of Econometrics.* **64** 375–400.

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

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

[6] Feuerverger, A. (1990). An efficiency result for the empirical characteristic function in stationary time-series models. *The Canadian Journal of Statistics.* **18** 155–161.

[7] Feuerverger, A. and R.A. Mureika. (1977). The empirical characteristic function and its applications. *The Annals of Statistics.* **5** 88–97.

[8] Ghysels, E., A.C. Harvey, and E. Renault. (1996). *Stochastic Volatility.* In *Handbook of Stististics-14.*

[9] Gourieroux, C., A. Monfort, and E. Renault. (1993). Indirect inference. *Journal Applied Econometrics.* **85** 85–118.

[10] Hansen, L.P. and J.A. Scheinkman. (1995). Back to future: Generating moment implications for continuous-time markov processes. *Econometrica.* **63** 767–804.

[11] Harvey, A.C. (1993). Long memory in stochastic volatility. London School of Economics.

[12] Harvey, A.C., E. Ruiz, and N. Shephard. (1994). Multivariate stochastic variance models. *Review of Economic Studies.* **61** 247–264.

[13] Jacquier, E, N.G. Polson, and P.E. Rossi. (1994). Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics.* **12** 371–389.

[14] Knight, J.L. and J. Yu. (1997). The empirical characteristic function in time series estimation. The University of Western Ontario.

[15] Knight, J.L. and S.E. Satchell. (1996). Estimation of stationary stochastic processes via the empirical characteristic function. The University of Western Ontario.

[16] Melino, A. and S.M. Turnbull. (1990). Pricing foreign currency options with stochastic volatility. *Journal of Econometrics.* **45** 239–265.

[17] Nelson, D.B. (1988). *Time Series Behavior of Stock Market Volatility and Returns.* PhD thesis, Massachusetts Institution of Technology.

[18] Quandt, R.E. and J.B. Ramsey. (1978). Estimating mixtures of normal distributions and switching regressions. *Journal of the American Statistical Association.* **73** 730–738.

[19] Ruiz, E. (1994). Quasi-maximum likelihood estimation of stochastic volatility models. *Journal of Econometrics.* **63** 289–306.

[20] Schmidt, P. (1982). An improved version of the Quandt-Ramsey MGF estimator for mixtures of normal distributions and switching regressions. *Econometrica.* **50** 501–524.

[21] Tran, K. (1994). *Mixture, Moment and Information-Three Essays in Economietrics.* PhD thesis, The University of Western Ontario.