Estimation of a Self-Exciting Poisson Jump Diffusion Model by the Empirical Characteristic Function Method

JUN YU[1]

Department of Economics
The University of Auckland
Private Bag 92019
Auckland, New Zealand

November 10, 1999

Abstract

This paper proposes a new econometric methodology for the estimation of diffusion models that include a jump component. The jump component's arrival time is endogenously determined, reflecting past volatility in the data and deviations from economic fundamentals. Although the likelihood method does not have closed form for this model, we show that the characteristic function can be derived analytically and hence developed an empirical characteristic function method to estimate the system parameters. This procedure has the same asymptotic efficiency as maximum likelihood, and is thus a desirable method to use when the likelihood function is unknown. A Monte Carlo study shows that the empirical characteristic function method outperforms the GMM procedure for the model. An application is considered for S&P 500 daily returns.

JEL#: C13, C15, C22, G10

Key words and phrases: Empirical Characteristic Function, Diffusion Process, Poisson Jump, Self-exciting, GARCH

[1]I would like to thank Viv Hall, John Knight, Peter C. B. Phillips, Hailong Qian and Julian Wright for their suggestions and comments. Email: j.yu@auckland.ac.nz. Tel: 64-9-373-7599 ext. 5770; fax: 64-9-373-7427.

1 Introduction

Motivated by the apparent non-linearity in financial time series, such as volatility clustering, researchers have resorted to processes which can generate such a property. Among them are ARCH-type models proposed by Engle (1982) and Bollerslev (1986), stochastic volatility (SV) models proposed by Clark (1973), Tauchen and Pitts (1983) and Taylor (1986), and the self-exciting jump diffusion model proposed by Knight and Satchell (1998). A common feature for these models is that they all allow for time dependence. Also, these models are representable as martingales and hence consistent with the efficient market hypothesis in the weak sense. Among these three types of models, the ARCH-type models have thus far attracted the most attention. The reason for this is primarily the relative ease with which the ARCH-type model can be estimated. Unfortunately, there are still some difficulties involved in the ARCH-type models. One of them is the model's inability to explain that large changes are not unusual after stable periods, since the volatility evolves according to a deterministic mechanism in the ARCH-type models. Another difficulty is that the model can not explain the discontinuities existing in the sample path of most financial time series; see Jarrow and Rosenfeld (1984), Ball and Torous (1985) and Jorion (1988).

By introducing another error term and hence treating the volatility as a stochastic process, the SV model can overcome the first difficulty involved in the ARCH-type models. However, a new problem arises from the use of the SV model. Even for the simplest SV model, neither the exact likelihood function nor the conditional likelihood function has a closed form. Therefore, the likelihood based estimation method is extremely difficult to implement. In the recent literature, some alternative estimation methods have been proposed to estimate the SV model; the methods include generalized method of moments (GMM), the quasi-maximum likelihood (QML) method, Markov Chain Monte Carlo (MCMC), the simulated maximum likelihood (SML) method and

efficient method of moments (EMM).

An alternative way to overcome the difficulty involved in the ARCH-type models is to compound a Brownian motion (BM) and a Poisson jump process. The jump component is first introduced by Press (1967) and Merton (1976) and extended by Jorion (1988) to explain the discontinuities in financial time series data. The models proposed by Press and Merton are, however, independent processes. It is now understood that most financial time series are uncorrelated in levels, but not, in general, uncorrelated in squares. Consequently, a model is needed to explain both discontinuity and non-linearity. To allow for non-linearity, the intensity parameter in the Poisson process is assumed to be self-exciting and hence time dependent, making the model different from the model proposed by Jorion. This model was first proposed by Knight and Satchell (1998) and applied to fit the UK stock data by Knight, Satchell and Yoon (1993). Unfortunately, the maximum likelihood (ML) method is not applicable to this model since the likelihood function has no closed form. Instead Knight, Satchell and Yoon employ a non-optimal GMM procedure to estimate the model. When fitting the UK data to the model, however, they find that the GMM estimates sometimes do not make any sense. For example, the estimate of the variance parameter for some stocks is negative. This finding is due to the poor finite sample properties of the GMM estimator.

In this paper, we use an alternative approach to estimate the self-exciting Poisson jump diffusion model – via the empirical characteristic function (ECF). 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 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 function. Moreover, by using the characteristic function, we can overcome the difficulties arising from ignorance of

the true density function or the true likelihood function. Although the self-exciting Poisson jump diffusion model has no closed form density, it has a known form of the characteristic function. Thus the ECF method is a desirable alternative estimation method.

The paper is organized as follows. The next section introduces the self-exciting Poisson jump diffusion 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 this model; the characteristic function of the 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. Section 5 concludes. The appendix collects the proof of the theorem, and the analytic expressions for some moments of the model, and the expression for the covariance matrix of the ECF estimator.

2 The Model

It is common in the financial literature to assume that the price of an asset at time t, $P(t)$, follows a geometric Brownian Motion (BM)

$$dP(t) = \gamma P(t)dt + \sigma P(t)dB(t), \tag{2.1}$$

where $B(t)$ is standard Brownian motion, γ is the instantaneous return and σ^2 is the instantaneous variance. By including the jump component, Knight and Satchell (1998) assume that the price follows the mixed Brownian-Poisson process

$$dP(t) = \gamma P(t)dt + \sigma P(t)dB(t) + P(t)(\exp(Q) - 1)dN(t), \tag{2.2}$$

where Q is a normal variate with mean μ_Q and variance σ_Q^2 in the interval $(t, t + \Delta t]$, and $N(t)$ is a Poisson process with intensity parameter $\lambda(t)$. By using Ito's lemma, we can solve the stochastic differential equation (2.2) for the stock return $X(t)(=$

$\log(P(t)/P(t-1))$,

$$\begin{aligned} X(t) &= (\gamma - \frac{\sigma^2}{2}) + \sigma B(1) + \sum_{n=1}^{\Delta N(t)} Q(n) \\ &= \mu + \sigma B(1) + \sum_{n=1}^{\Delta N(t)} Q(n), \end{aligned} \tag{2.3}$$

where $\mu = \gamma - \frac{\sigma^2}{2}$. Hence, the behavior of $X(t)$ depends not only on the continuous diffusion part $\mu + \sigma B(1)$, but also a discontinuous jump part $\sum_{n=1}^{\Delta N(t)} Q(n)$. The continuous part is responsible for the usual day-to-day price movement, such as temporary imbalance between supply and demand or firm-specific information that only has a marginal effect on the value of the stock. The discontinuous jump part corresponds to the arrival of new important information to the market. When the jump component is relatively more important, a large change can follow stable periods and discontinuities occur. Furthermore, the Poisson process $N(t)$ is assumed to have an intensity function $\lambda(t)$ which is self-exciting as follows:

$$\lambda(t) = \sum_{i=1}^{m} \alpha_i \nu^2(t-i) + \sum_{j=1}^{l} \beta_i Var(X(t-j)|I(t-j-1)), \tag{2.4}$$

where $\nu(t)$ is $N(0,1)$ conditional on $N(t)$, and $I(t)$ is information up to the close of the market on day t.

The motivation for the model is the idea that the flow of information at day t's trading is conditioned by the news known at the close of trading at day $t-1$ or prior to opening at day t. Also, the motivation comes from the idea that the expected number of jumps, which corresponds to the arrival of new important information to the market, depends upon past volatility and deviations from market fundamentals $B^2(1)$ $(= \nu^2(t))$. Due to the dependence of $\lambda(t)$, $X(t)$ could be time dependent. Considering the conditional variance of $X(t)$, we have

$$\begin{aligned} &Var(X(t)|I(t-1)) = \\ &\quad \sigma^2 + (\mu_Q^2 + \sigma_Q^2)\left(\sum_{i=1}^{m} \alpha_i \nu^2(t-i) + \sum_{j=1}^{l} \beta_i Var(X(t-j)|I(t-j-1))\right). \end{aligned} \tag{2.5}$$

Obviously the conditional variances are correlated and hence volatility clustering can be explained by the model. Therefore, as an implication of the model, the non-linearity is due to the dependence of the jump component's arrival time which is endogenously determined. In the model proposed by Jorion (1988), however, the non-linearity is due to an ARCH component which is explicitly specified. Unfortunately, $X(t)$, the self-exciting Poisson jump diffusion process, has no closed form expression for the likelihood function because it basically compounds three processes, the Brownian Motion, a Poisson process and an iid $\chi^2_{(1)}$ sequence. This lack of tractable form for the likelihood function makes the ML method extremely difficult to implement.

For simplicity, Knight, Satchell and Yoon (1993) consider the model with $m = l = 1$ in equation (2.4); that is,

$$\lambda(t) = \alpha\nu^2(t-1) + \beta Var(X(t-1)|I(t-2)). \tag{2.6}$$

In Appendix A, we show that equation (2.6) is equivalent to

$$\lambda(t) = \beta\sigma^2 + \beta(\mu_Q^2 + \sigma_Q^2)\lambda(t-1) + \alpha\nu^2(t-1), \tag{2.7}$$

or

$$Var(X(t)|I(t-1)) = \sigma^2 + \beta(\mu_Q^2+\sigma_Q^2)Var(X(t-1)|I(t-2)) + \alpha(\mu_Q^2+\sigma_Q^2)\nu^2(t-1). \tag{2.8}$$

Knight, Satchell and Yoon claim that $m = l = 1$ is sufficient for most applications. Together with the mean equation

$$E(X(t)|I(t-1)) = \mu + \mu_Q\lambda(t), \tag{2.9}$$

we can tell that if $\mu_Q \neq 0$ the model has an ARCH-M effect which is referred to as "changing lagged conditional variances directly effect the expected return on a portfolio" by Engle, Lilien and Robins (1987). If $\mu_Q = 0$ the model is similar to the GARCH model proposed by Bollerslev (1986). For the model based on equations (2.3) and (2.6),

some conditions on the parameters are needed for the stationarity of the process (see Knight and Satchell (1998) for details). Based on the analytic expressions for some moments, GMM estimators can be obtained. However, since GMM only uses a few moment conditions, it is not surprising that the finite sample performance is not good. In the next section we derive the characteristic function of $X(t)$ and then perform the estimation based on the empirical characteristic function. Since the characteristic function 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 approach should outperform that based on the GMM approach.

3 ECF Estimation

Before we discuss the estimation of the model via 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 (EDF). Therefore, the CF and ECF are the Fourier transforms of the distribution function and EDF respectively. Because of the uniqueness of the Fourier-Stieltjes transform, 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 CF. For example, by choosing discrete $r_1, \cdots, r_q$, we can minimize the

distance on q discrete points

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

Alternatively, by choosing r continuously, we can minimize the distance over an interval

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

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). Therefore, (3.3) with the marginal ECF can be minimized. This is the approach followed by the previous literature. To estimate the mixture of normals, for example, Quandt and Ramsey (1978) adopt an ordinary least square (OLS) procedure while Schmidt (1982) adopts a generalized least squares (GLS) procedure using the moment generating function instead of the CF. 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 $\sqrt{n}(c_n(r) - c(r; \boldsymbol{\theta}))$. These have been established by Feuerverger and Mureika (1977), and Csörgö (1981) for any iid sequence.

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 suffers from the same problem. This is why the joint CF is need. We do this by a procedure involving moving blocks of data. Denote the overlapping blocks for $x_1, x_2, \cdots, x_T$ as

$$\boldsymbol{z}_j = (x_j, \cdots, x_{j+p})', \; j = 1, \cdots, T - p. \tag{3.5}$$

The characteristic function of each block is defined as

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

where $\boldsymbol{r} = (r_1, \cdots, r_{p+1})$, and $\boldsymbol{\theta}$ represents the parameters of interest. The ECF is defined as

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

where $n = T - p$.

Several estimation procedures are proposed by Knight and Yu (1999). 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 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 $\boldsymbol{z}_j$. One could, of course, do the calculation with the conditional CF as an alternative. In fact, a recent paper by Singleton (1999) develops an estimation method based on the conditional CF.

Knight and Yu (1999) present four different versions of the ECF methods. We discuss them in detail. When the transformation variables, $\boldsymbol{r}$'s, are chosen discretely and the weight is optimal, the procedure is referred to as GLS of the discrete ECF method, which is the first version of the ECF method. In this version the weight function $g(\boldsymbol{r})$ is a function with a certain number of jumps. Feuerverger (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. The results are of theoretical interest but involve no empirical calculations. Considering the estimation of time series by using GLS of the discrete ECF method, Knight and

Satchell (1997) give a multi-step procedure.[2] The main idea is the following. We first choose moving blocks (that is 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 in a optimal way, 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 the size of the jumps. Since each choice of $\boldsymbol{r}$ corresponds to a moment condition in the CF, the estimator, in essence, is a GMM estimator.

To overcome the difficulties, we can choose the transformation variables continuously. The procedure is called the continuous ECF method by Knight and Yu (1999). In the continuous ECF method the transformation variables are simply integrated out. By choosing the transformation variables continuously, the procedure basically matches the ECF with the CF over an interval, and hence matches all the moments continuously. If an equal weight is chosen, for example, $g(\boldsymbol{r}) = 1$, the procedure is the OLS-ECF method which is the second version of the ECF method. When an unequal weight is used, the procedure corresponds to the WLS-ECF method which is the third version of the ECF method. Furthermore, when the weight is chosen to be optimal, the procedure is called the GLS-ECF method which is the fourth version of the 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

$$w(\boldsymbol{r}) = \int \cdots \int \exp(-i\boldsymbol{r}'\boldsymbol{z}_j) \frac{\partial \log f(x_{j+p}|x_j, \cdots, x_{j+p-1})}{\partial \boldsymbol{\theta}} dx_j \cdots dx_{j+p}, \tag{3.10}$$

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 jump

[2]In fact, the procedure developed by Knight and Satchell (1997) is based on the cumulant generating function rather than the characteristic function.

diffusion model.

Under standard regularity conditions, Knight and Yu (1999) establish the strong consistency and asymptotic normality for the resulting estimators when the above four procedures are used to estimate strictly stationary processes. Furthermore, by studying the finite sample properties of the ECF estimators for stationary ARMA processes, Knight and Yu 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 should be more efficient. However, calculations associated with larger p are numerically more difficult. For a Markov process such as a Gaussian AR(1) process, Knight and Yu find that blocks with $p = 1$ are enough to capture virtually all the information in the original series.

The model we estimate via ECF is defined by (2.3) and (2.6). In order to use the ECF method, of course, we need to find the expression of the joint characteristic function. The joint characteristic function for $X(t), \cdots, X(t-k)$ is obtained in Theorem 3.1. The joint characteristic function for the model defined by (2.3) and (2.4) can be obtained in the similar way.

Theorem 3.1 *If a random process $\{X(t)\}_{t=1}^{n}$ is a self-exciting Poisson jump diffusion model which is defined by equations (2.3) and equation (2.6), then the joint* CF *of $X(t), \cdots, X(t-k)$ is,*

$$c(r_1, \cdots, r_{k+1}; \boldsymbol{\theta}) = \exp(i\mu \sum_{j=1}^{k+1} r_j - \frac{1}{2}\sigma^2 \sum_{j=1}^{k+1} r_j^2) \exp\left\{\frac{\beta\sigma^2}{1-\phi} \sum_{j=1}^{k+1} G(r_j)\right\} \prod_{l=0}^{\infty}\left\{1 - 2\alpha\phi^l \sum_{j=1}^{k+1} \phi^{k+1-j} G(r_j)\right\}^{-1/2} \prod_{j=2}^{k+1}\left\{1 - 2\alpha \sum_{l=1}^{j} \phi^{j-l} G(r_l)\right\}^{-1/2}, \quad (3.11)$$

where $\phi = \beta(\mu_Q^2 + \sigma_Q^2)$, $G(r) = \exp(ir\mu_Q - \frac{r^2\sigma_Q^2}{2}) - 1$, and $\boldsymbol{\theta} = (\mu, \sigma^2, \alpha, \beta, \mu_Q, \sigma_Q^2)'$ are the parameters of interest.

Using this theorem, one can obtain covariances for the returns and the squared returns; for example, $Cov(X(t), X(t-s)) = 2\mu_Q^2\alpha^2\phi^s/(1-\phi^2)$. Hence, $X(t)$ and $X(t-s)$ are uncorrelated when $\mu_Q = 0$. However, $X(t)$ and $X(t-1)$ are not independent since $c(r_1, r_2; \boldsymbol{\theta}) \neq c(r_1; \boldsymbol{\theta})c(r_2; \boldsymbol{\theta})$. Moreover, the marginal characteristic function and hence moments can be obtained from the joint characteristic function. Appendix B presents the analytic expressions for some moments of the model, including four unconditional moments and three autocovariances.

In order to use the ECF method to estimate the model defined by equations (2.3) and (2.6), we have to choose a value for p. As we argued above, a larger value of p works better than a smaller value of p, however, we shall choose $p = 1$ at first. For the model defined by (2.3) and (2.6) we note that with $p = 1$,

$$\begin{aligned} c(r_1, r_2; \boldsymbol{\theta}) &= \exp\left\{\frac{\beta\sigma^2}{1-\phi}(G(r_1) + G(r_2)) + i\mu(r_1 + r_2) - \frac{1}{2}\sigma^2(r_1^2 + r_2^2)\right\} \\ &\quad \prod_{l=0}^{\infty}\left\{1 - 2\alpha\phi^l(\phi G(r_1) + G(r_2))\right\}^{-1/2}(1 - 2\alpha G(r_1))^{-\frac{1}{2}}, \end{aligned} \tag{3.12}$$

and

$$c_n(r_1, r_2) = \frac{1}{n}\sum_{j=1}^{n}\exp(ir_1x_j + ir_2x_{j+1}). \tag{3.13}$$

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

$$Re\, c_n(r_1, r_2) = \frac{1}{n}\sum_{j=1}^{n} cos(r_1x_j + r_2x_{j+1}) \tag{3.14}$$

and

$$Im\, c_n(r_1, r_2) = \frac{1}{n}\sum_{j=1}^{n} sin(r_1x_j + r_2x_{j+1}). \tag{3.15}$$

As we mentioned above, 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 (1999) shows that the continuous ECF method works better than the discrete ECF method, we use the continuous ECF method to

estimate the model. However, the optimal weight function in the continuous ECF method is not readily obtained because the conditional score function has no closed form expression for the model. Instead, we use the OLS-ECF method with a constant weight function chosen over the interval $[0,1]$. Hence, the procedure is to choose $(\hat{\mu}, \hat{\sigma}^2, \hat{\alpha}, \hat{\beta}, \hat{\mu}_Q, \hat{\sigma}_Q^2)$ to minimize

$$\begin{aligned} &\int_0^1 \int_0^1 \Big[(Re\, c(r_1, r_2; \boldsymbol{\theta}) - \tfrac{1}{n} \textstyle\sum_{j=1}^n cos(ir_1 x_j + ir_2 x_{j+1}))^2 \\ &+ (Im\, c(r_1, r_2; \boldsymbol{\theta}) - \tfrac{1}{n} \textstyle\sum_{j=1}^n sin(ir_1 x_j + ir_2 x_{j+1}))^2 \Big] dr_1 dr_2, \end{aligned} \tag{3.16}$$

where $c(r_1, c_2; \boldsymbol{\theta})$ is given by (3.12).

The resulting estimators are consistent and asymptotically normal with the asymptotic covariance matrix of the estimators given by:

$$\begin{aligned} &\frac{1}{n}\Big\{ \int_0^1 \int_0^1 \Big[\frac{\partial Re\, c(\boldsymbol{r}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial Re\, c(\boldsymbol{r}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} + \frac{\partial Im\, c(\boldsymbol{r}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial Im\, c(\boldsymbol{r}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \Big] dr_1\, dr_2 \Big\}^{-1} \times \\ &A(\boldsymbol{\theta}) \times \Big\{ \int_0^1 \int_0^1 \Big[\frac{\partial Re\, c(\boldsymbol{r}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial Re\, c(\boldsymbol{r}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} + \frac{\partial Im\, c(\boldsymbol{r}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial Im\, c(\boldsymbol{r}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \Big] dr_1\, dr_2 \Big\}^{-1} \quad . \end{aligned}$$

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

4 Implementation, Simulation and Application

4.1 Implementation

The implementation of the ECF method essentially requires minimizing (3.16) which involves double integrals. Unfortunately, no analytic solutions for either the double integrals or the optimization are available. Consequently, we will numerically evaluate the multiple integral (3.16), followed by numerical minimization of (3.16) with respect to $\boldsymbol{\theta}$. The numerical solutions are the desired estimators.

A DCUHRE algorithm proposed by Berntsen, Espelid and Genz (1991) is used to approximate the two dimensional integrations in (3.16). Since there is no analytic expression for the derivatives of the objective functions, the Powell's conjugate direction

algorithm (Powell, 1964)) is used to find the global minimum. All computations are done in double precision.

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

4.2 Monte Carlo Simulation

The Monte Carlo study is designed to check the viability of the ECF method in comparison with a GMM approach. For simplicity, in the model defined by equations (2.3) and (2.6), we let $\mu = \mu_Q = \sigma = 0, \alpha = 1$, and thus $\phi = \beta\sigma_Q^2$. Therefore, the model can be rewritten as

$$X(t) = \sum_{n=1}^{\Delta N(t)} Q(n), \tag{4.1}$$

where $Q(n)|\Delta N(t) \sim i.i.d.\, N(0, \sigma_Q^2), \Delta N(t) \sim P(\lambda(t))$, and

$$\lambda(t) = \phi\lambda(t-1) + \nu^2(t-1). \tag{4.2}$$

Obviously, the model has $Var(X(t)) = \frac{\sigma_Q^2}{1-\phi}$ and $E[X(t) - E(X(t))]^4 = \frac{3\sigma_Q^4}{(1-\phi)^2} + \frac{3\sigma_Q^4}{1-\phi} + \frac{6\sigma_Q^4}{1-\phi^2}$, so the kurtosis in the unconditional distribution is strictly greater than 3 as long as $\sigma_Q^2 \neq 0$. Although μ_Q is set to be 0 and hence implies that the process $\{X(t)\}$ becomes serially uncorrelated, a joint characteristic function is needed since the process is not independent. Furthermore, for the process to be stationary, $|\phi|$ must be smaller than 1. We choose parameters $\sigma_Q^2 = 1$ and $\beta = 0.5$, which imply $\phi = 0.5$. The number of observations is set at $T = 2,000$ and the number of replications is set at 500.

The details of the GMM procedure and the asymptotic properties of the GMM estimator are given by Hansen (1982) in a more general framework. In this specific situation, denote the vector of the sample realizations of the moments at time t by $m_t = (m_{1t}, ..., m_{Qt})$, where Q is the number of the moments selected. The vector of the sample moments are $M_T = (M_{1T}, ..., M_{QT})'$, where $M_{iT} = \sum_{t=j+1}^{T} m_{it}/(T-j)$ for

$j = 1, ..., Q$, and j is the maximum lag between the variables defining the sample moments. Finally, define the vector of corresponding population moments by $f_t(\boldsymbol{\theta})$. The GMM estimator, $\hat{\boldsymbol{\theta}}_{GMM}$, minimizes the distance between $f_t(\boldsymbol{\theta})$ and M_T over the parameter space Θ; that is,

$$\hat{\boldsymbol{\theta}}_{GMM} = \arg\min_{\boldsymbol{\theta}\in\Theta}\{(f_t(\boldsymbol{\theta}) - M_T)'\hat{W}_T(f_t(\boldsymbol{\theta}) - M_T)\}, \tag{4.3}$$

where $\hat{W}_T$ is a positive definite weighting matrix which is chosen to yield the smallest asymptotic covariance matrix of the GMM estimator. In this paper, $\hat{W}_T$ is constructed by the Bartlett kernel with a fixed lag length 10. This is the kernel estimator proposed by Newey and West (1987) and it is widely used in the literature. Under regularity conditions, Hansen (1982) show that the estimator $\hat{\boldsymbol{\theta}}_{GMM}$ is consistent and asymptotically normal, that is,

$$T^{1/2}(\hat{\boldsymbol{\theta}}_{GMM} - \boldsymbol{\theta}) \overset{a}{\sim} N(0, V_T), \tag{4.4}$$

and a consistent estimator of V_T is given by $\hat{V}_T = \frac{1}{T}(\hat{D}'(\boldsymbol{\theta})\hat{S}^{-1}(\boldsymbol{\theta})\hat{D}'(\boldsymbol{\theta}))^{-1}$, where $\hat{D}(\boldsymbol{\theta})$ is the Jacobian matrix of $f_t(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ evaluated at the estimated parameters, and $\hat{S}(\boldsymbol{\theta})$ is a consistent estimator of $S(\boldsymbol{\theta}) = E[f_t(\boldsymbol{\theta})f_t'(\boldsymbol{\theta})]$. Despite the appealing asymptotic properties, GMM does not provide any guide to how many and which moment conditions one needs to use. In fact, there are infinitely many moments that can be used in the GMM estimation, including integer, fractional, and irrational moments. In this paper, we arbitrarily choose seven moment conditions that are listed in Appendix B. The only guide used to select these moments is to avoid high order moments due to the erratic finite-sample behavior caused by the presence of fat-tails in the distribution of the returns, pointed out by Andersen and Sorensen (1996). To obtain the ECF estimates, we choose the starting point to be the GMM estimates in the numerical optimization.

Table 1 presents the simulation results. The table shows the mean, median, minimum, maximum, mean square error (MSE) and root mean square error (RMSE) for

both estimates, and serves to illustrate that the ECF method outperforms the optimal GMM approach. Noticeably, ECF works much better than GMM. For example, the MSE of $\hat{\beta}_{GMM}$ is more than 20 times larger than that of $\hat{\beta}_{ECF}$. The results for the estimates of σ_Q^2 also favor the ECF method. Figures 1 and 2 also show that ECF performs better than GMM.

4.3 Application

4.3.1 Data

The data we used 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 the period 1980-1987 and is plotted in Figure 3.[3] The October 1987 stock-market crash can be clearly identified from the graph. The summary statistics of both static and dynamic properties of the series and the squared series are presented in Tables 2 and 3, from which one can see that the daily returns are skewed and have excess kurtosis (> 3). The dynamic properties reveal that for the return series all the autocorrelations are very small. However, for the squared returns, the lower order autocorrelations are low but not negligible and decay in a very slow fashion.

4.3.2 Empirical Results

In this empirical study, we fit the model defined by (2.2) and (2.6) to the data. The parameters of interest are $\boldsymbol{\theta} = (\mu, \sigma^2, \alpha, \beta, \mu_Q, \sigma_Q^2)'$. Again we let $\phi = \beta(\mu_Q^2 + \sigma_Q^2)$. To simplify the model, we shall assume that $X(t)$ is stationary. Consequently, $\lambda(t)$ has to be stationary. From (2.7) we know that $\lambda(t)$ is stationary only if $|\phi| < 1$. We further require $\alpha, \beta \geq 0$ to guarantee the intensity function to be non-negative. The other two constraints that we impose are $\sigma^2 \geq 0$ and $\sigma_Q^2 \geq 0$. The empirical results are presented in Table 4, where both the GMM and ECF estimates are given.

[3]The series has been filtered to remove calendar effects by Gallant, Rossi and Tauchen (1992).

From Table 4, we note that the ECF estimates of some parameters are not close to the GMM estimates. For example, $\hat{\sigma^2}_{ECF}$ is more than 10 times larger than $\hat{\sigma^2}_{GMM}$ while $\hat{\sigma^2_Q}_{GMM}$ is nearly 3 times as large as $\hat{\sigma^2_Q}_{ECF}$. From equation (2.6) we know that $\phi + \alpha$ captures intertemporal volatility correlation. The ECF estimate of $\phi + \alpha$ is 0.29 which is larger than the GMM estimate of the same parameter (=0.01). Consequently, the ECF estimates imply more persistent autocorrelation of volatilities. Since the empirical results are different, a comparison of the goodness of fit is of particular interest. To compare the goodness of fit, we simulate two sequences using the ECF and GMM estimates respectively. In Figure 4, 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 equations (2.3) and (2.6). Figure 4 clearly demonstrates that the ECF estimates have better goodness of fit than the GMM estimates.

To demonstrate the jump diffusion model is a desirable alternative to use, we first compare it with ARCH-type models. Such a comparison is interesting since most research has found that the ARCH-type models imply much stronger volatility persistence than that implied by the jump diffusion model; see, for example, Lin, Knight and Satchell (1999). In this paper, two ARCH-type models are considered: a GARCH(1,1) model and an ARCH-M model. The GARCH(1,1) is defined by

$$x_t = \mu + \sigma_t \epsilon_t, \tag{4.5}$$

$$\sigma_t^2 = \omega + \alpha(x_{t-1} - \mu)^2 + \beta\sigma_{t-1}^2, \tag{4.6}$$

and the ARCH-M is defined by

$$x_t = \mu + \delta\sigma_t + \sigma_t \epsilon_t, \tag{4.7}$$

$$\sigma_t^2 = \omega + \alpha(x_{t-1} - \mu - \delta\sigma_{t-1})^2 + \beta\sigma_{t-1}^2, \tag{4.8}$$

where $\epsilon_t \sim iidN(0,1)$.[4] Both models involve conditionally normal density, so the ML

[4]In the ARCH-M model we use the conditional standard deviation in place of the conditional variance in the mean equation because the former specification is preferable according to the Schwartz information criterion (SIC) defined by Schwartz (1978).

estimators are obtained. Table 5 presents the ML estimates of the parameters. As expected, both models suggest that a shock to conditional variance is highly persistent since the estimate of $\alpha + \beta$ is very close to 1. To compare the two ARCH models with the jump diffusion model, we use the available analytic expressions for moments derived from all three models to calculate some descriptive statistics based on the parameter estimates. Table 6 presents the exact unconditional moments of the jump diffusion model with the ECF estimates, the GARCH(1,1) model and the ARCH-M model, together with those calculated from the data. All three models match the standard deviation quite well, but the skewness and kurtosis are not matched closely by any model. Nevertheless, the jump diffusion model provides the skewness and kurtosis which are closest to their data counterpart. It is well known that under some conditions the ARCH-type models can be approximated by continuous-path diffusion processes and vice versa (see Nelson, 1990). Therefore, it is not surprising that the conventional ARCH-type models can not account for the discontinuities present in the data, such as the 1987 crash. This may explain why the ARCH-type model is unable to generate much excess kurtosis for the data set used. On the other hand, however, with a jump component included into a diffusion process, discontinuities are modeled explicitly and hence the rare event such as the 1987 crash can be better explained.

We also compare the jump diffusion model with a SV model. The SV model has been found to provide better goodness-of-fit than the ARCH-type models for the same series (see Danielsson, 1994). Moreover, Jacquier, Polson and Rossi (1994) and Hsieh (1991) find that the ARCH-type filter does not remove all non-linear dependencies present in financial time series. The SV model is defined by

$$x_t = \exp(0.5h_t)e_t, \tag{4.9}$$

$$h_t = \lambda + \alpha h_{t-1} + v_t, \tag{4.10}$$

where $e_t \sim iidN(0,1)$, $v_t \sim iidN(0,\sigma^2)$ and $corr(e_t, v_t) = 0$. This model has been

fitted to the same series by Jacquier, Polson and Rossi using MCMC.[5] The exact unconditional moments of the SV model with the MCMC estimates are calculated and also presented in Table 6. Similar to the other three models, the SV model also matches the standard deviation well. However, the SV model fails to provide better skewness and kurtosis than the jump diffusion model.

We can now interpret the ECF estimates of the jump diffusion model. Firstly, $\phi = 0.29$ implies a positive correlation of the conditional variance between two consecutive trading days. The conditional variance at day t has much less dependence on the deviations from market fundamentals $\nu(t-1)$. This dependence is captured by $\alpha(\mu_Q^2 + \sigma_Q^2)$ which equals 0.001. Secondly, the instantaneous mean and instantaneous variance in the continuous part of the jump diffusion model equal 0.055 and 0.79 respectively. As we argued in Section 2, the continuous part is due to the arrival of general economic information to the market. Therefore, the arrival of general economic information generates daily log returns of 0.055% or annual log returns of 13.75%. Thirdly, the mean and variance of the variate in the jump part of the jump diffusion model equal -0.16 and 6.18 respectively. Since the jump part is due to the arrival of new important information to the market, on average important information leads to daily log returns of -0.16% or annual log returns of -40%. The finding that important information incurs a negative return is not very surprising if one believes that important information is more likely to cause downward pressure on the market. Also, the size of the drift which corresponds to small movements is much smaller than the average size of the jumps which correspond to large movements. Finally, the ECF estimate of σ_Q^2 is statistically significant and indicates that discontinuities are present in the sample path of S&P 500 returns over the period from 1980 through 1987, consistent with the findings by Ball and Torous (1985) in the sample path of daily stock returns and by Jorion (1988) in the sample path of weekly exchange rates and weekly stock returns.

[5]The MCMC estimates of λ, α and σ are -0.002, 0.97, 0.15 respectively.

5 Conclusion

This paper proposes a new econometric methodology for the estimation of diffusion models that include a jump component. The jump component's arrival time is endogenously determined, reflecting past volatility in the data and deviations from economic fundamentals. Although the likelihood method does not have closed form for this model, we show that the characteristic function can be derived analytically and hence developed an empirical characteristic function method to estimate the system parameters. The basic idea of the ECF method is to match the theoretical characteristic function derived from the model to the ECF calculated from the sampling observations. The approach yields consistent and asymptotically normal estimates of the parameters. Simulations demonstrate that the ECF method works well in comparison with a GMM approach. An empirical application to S&P 500 index returns shows the capabilities of the new method and reveals some differences with ARCH-type models.

Appendix A

Proof of Theorem 1

We start the proof by showing that equations (2.6), (2.7) and (2.8) are equivalent. From the definition of $I(t-1)$, we get

$$Var(X(t)|I(t-1)) = Var(\mu + \sigma B(1) + \sum_{n=1}^{\Delta N(t)} Q(n)|\lambda(t)) = \sigma^2 + \lambda(t)(\mu_Q^2 + \sigma_Q^2). \quad \text{(A.1)}$$

Substituting out $\lambda(t)$ in equation (2.6), we then get

$$Var(X(t)|I(t-1)) = \sigma^2 + \alpha(\mu_Q^2 + \sigma_Q^2)\nu(t-1)^2 + \beta(\mu_Q^2 + \sigma_Q^2)Var(X(t-1)|I(t-2)). \quad \text{(A.2)}$$

which is equation (2.8). Therefore, the conditional variance of this model follows a GARCH(1, 1). Furthermore, by substituting out $Var(X(t-1)|I(t-2))$ in equation (2.6), we get

$$\lambda(t) = \beta\sigma^2 + \beta(\mu_Q^2 + \sigma_Q^2)\lambda(t-1) + \alpha\nu^2(t-1), \quad \text{(A.3)}$$

which is equation (2.7). Since the intensity represents how fast new information arrives, equation (A.3) means that the speed of the arrival of new information on day t depends on how frequent new information has arrived on day $t-1$, as well as a random component. By applying backward induction to equation (A.3), we get

$$\lambda(t) = \frac{\beta\sigma^2}{1-\phi}(1-\phi^{t-1}) + \phi^{t-1}\lambda(1) + \alpha\sum_{j=0}^{t-2}\phi^j\nu^2(t-j-1). \tag{A.4}$$

If $|\phi| < 1$, as $t \to \infty$,

$$\lambda(t) = \frac{\beta\sigma^2}{1-\phi} + \sum_{j=1}^{\infty}\phi^j\nu^2(t-j-1). \tag{A.5}$$

Consequently, the characteristic function of $\lambda(t)$ is

$$\begin{aligned}
\Psi(r) &= E(\exp(ir\lambda(t))) \\
&= E\left\{\exp\left(is(\frac{\beta\sigma^2}{1-\phi} + \sum_{j=1}^{\infty}\phi^j\nu^2(t-j-1))\right)\right\} \\
&= \exp(ir\frac{\beta\sigma^2}{1-\phi})\prod_{j=0}^{\infty}(1-2i\alpha\phi^j)^{-\frac{1}{2}}.
\end{aligned} \tag{A.6}$$

Considering

$$\sum_{n=1}^{\Delta N(t-l)} Q(n)|\Delta N(t-l) \sim N(\mu_Q\Delta N(t-l), \sigma_Q^2\Delta N(t-l)),$$

we get

$$\begin{aligned}
& E\left\{\prod_{l=1}^{k+1}\exp\left(ir_l\sum_{n=1}^{\Delta N(t-l)}Q(n)\right)\right\} \\
= \; & E\left\{E\left\{E\left[\prod_{l=1}^{k+1}\exp(ir_l\sum_{n=1}^{\Delta N(t-l)}Q(n))|\Delta N(t-l), I(t-k)\right]\right\}\right\} \\
= \; & E\left\{E\left[\prod_{l=1}^{k+1}\exp(i\mu_Q r_l\Delta N(t-l) - \frac{r_l^2}{2}\sigma_Q^2\Delta N(t-l))|I(t-k)\right]\right\} \\
= \; & E\left\{\prod_{l=1}^{k+1}\exp[\lambda(t-l)(\exp(i\mu_Q r_l - \frac{r_l^2}{2}\sigma^2)-1)]\right\} \\
= \; & E\{\prod_{l=1}^{k+1}\exp[(\frac{\beta\sigma^2}{1-\phi}(1-\phi^{k-l}) + \phi^{k-l}\lambda(t-k)
\end{aligned}$$

$$
\begin{aligned}
&+\alpha(\nu^2(t-j-1)+\cdots+\phi^{k-l-1}\nu^2(t-k)))(\exp(i\mu_Q r_l-\frac{r_l^2}{2}\sigma^2)-1)]\} \\
= \;& \exp\left[\sum_{l=1}^{k+1}\frac{\beta\sigma^2}{1-\phi}(1-\phi^{k-l})(\exp(i\mu_Q r_l-\frac{r_l^2}{2}\sigma^2)-1)\right] \\
& E\left\{\exp\left[\lambda(t-k)\sum_{l=1}^{k+1}\left(\phi^{k+1-l}(\exp(i\mu_Q r_l-\frac{r_l^2}{2}\sigma^2)-1)\right)\right]\right\} \\
& E\left\{\prod_{l=1}^{k+1}\exp\left[\alpha(\nu^2(t-j-1)+\cdots+\phi^{k-l-1}\nu^2(t-k))(\exp(i\mu_Q r_l-\frac{r_l^2}{2}\sigma^2)-1)\right]\right\} \\
= \;& \exp\left[\sum_{l=1}^{k+1}\frac{\beta\sigma^2}{1-\phi}(1-\phi^{k-l})(\exp(i\mu_Q r_l-\frac{r_l^2}{2}\sigma^2)-1)\right] \\
& \exp\left\{\frac{\beta\sigma^2}{1-\phi}\sum_{l=1}^{k+1}\left[\phi^{k+1-l}(\exp(i\mu_Q r_l-\frac{r_l^2}{2}\sigma^2)-1)\right]\right\} \\
& \prod_{j=0}^{\infty}\left\{1-2\alpha\phi^j\sum_{l=1}^{k+1}\left[\phi^{k+1-l}(\exp(i\mu_Q r_l-\frac{r_l^2}{2}\sigma^2)-1)\right]\right\} \\
& E\left\{\exp\left[\sum_{j=2}^{k+1}\left(\alpha\sum_{l=1}^{j}\phi^{j-l}(\exp(i\mu_Q r_l-\frac{r_l^2}{2}\sigma^2)-1)\right)\right]\right\} \\
= \;& \exp\left[\sum_{l=1}^{k+1}\frac{\beta\sigma^2}{1-\phi}(1-\phi^{k-l})G(r_l)\right]\exp\left\{\frac{\beta\sigma^2}{1-\phi}\sum_{l=1}^{k+1}\left[\phi^{k+1-l}G(r_l)\right]\right\} \\
& \prod_{l=0}^{\infty}\left\{1-2\alpha_1\phi^l\sum_{j=1}^{k+1}\phi^{k+1-j}G(r_j)\right\}^{-1/2}\prod_{j=2}^{k+1}\left\{1-2\alpha\sum_{l=1}^{j}\phi^{j-l}G(r_l)\right\}^{-1/2} \\
= \;& \exp\left\{\frac{\beta\sigma^2}{1-\phi}\sum_{j=1}^{k+1}G(r_j)\right\}\prod_{l=0}^{\infty}\left\{1-2\alpha\phi^l\sum_{j=1}^{k+1}\phi^{k+1-j}G(r_j)\right\}^{-1/2} \\
& \prod_{j=2}^{k+1}\left\{1-2\alpha\sum_{l=1}^{j}\phi^{j-l}G(r_l)\right\}^{-1/2}. \qquad \text{(A.7)}
\end{aligned}
$$

Thus, the characteristic function of $X(t),\cdots,X(t-k)$ is

$$
\begin{aligned}
& c(r_1,\cdots,r_{k+1};\boldsymbol{\theta}) \\
= \;& E\left\{\exp(ir_1X(t)+\cdots+ir_{k+1}X(t-k))\right\} \\
= \;& E\left\{\exp(ir_1(\mu+\sigma B(1)+\sum_{n=1}^{\Delta N(t-l)}Q(n))+\cdots+ir_{k+1}(\mu+\sigma B(1)+\sum_{n=1}^{\Delta N(t-l)}Q(n)))\right\} \\
= \;& E\left\{\exp(i\mu\sum_{l=1}^{k+1}r_l)\exp(i\sigma(r_1B(1)+\cdots+r_{k+1}B(1)))\exp(\sum_{l=1}^{k+1}ir_l\sum_{n=1}^{\Delta N(t-l)}Q(n))\right\}
\end{aligned}
$$

$$
\begin{aligned}
&= \exp(i\mu\sum_{j=1}^{k+1} r_j - \frac{1}{2}\sigma^2\sum_{j=1}^{k+1} r_j^2)E\left\{\prod_{l=1}^{k+1}\exp\left(ir_l\sum_{n=1}^{\Delta N(t-l)} Q(n)\right)\right\} \\
&= \exp(i\mu\sum_{j=1}^{k+1} r_j - \frac{1}{2}\sigma^2\sum_{j=1}^{k+1} r_j^2)\exp\left\{\frac{\beta\sigma^2}{1-\phi}\sum_{j=1}^{k+1} G(r_j)\right\} \\
&\quad \prod_{l=0}^{\infty}\left\{1-2\alpha\phi^l\sum_{j=1}^{k+1}\phi^{k+1-j}G(r_j)\right\}^{-1/2}\prod_{j=2}^{k+1}\left\{1-2\alpha\sum_{l=1}^{j}\phi^{j-l}G(r_l)\right\}^{-1/2}. \quad \blacksquare
\end{aligned}
$$

Appendix B

Analytic Expressions for Moments of the Self-Exciting Jump Diffusion Model

The model is defined by equations (2.3) and (2.6). The joint cumulant generating function can be obtained by applying log operator to the joint characteristic function (3.11). The coefficients on the Taylor series expansion of the joint cumulant generating function are the cumulants of the model and given by

$$
\begin{aligned}
\kappa_1 &= \frac{\beta\sigma^2+\alpha}{1-\phi}\mu_Q + \mu \\
\kappa_2 &= \frac{\beta\sigma^2+\alpha}{1-\phi}(\mu_Q^2+\sigma_Q^2) + \sigma^2 + \frac{2\alpha^2\mu_Q^2}{1-\phi^2} \\
\kappa_3 &= \frac{\beta\sigma^2+\alpha}{1-\phi}(\mu_Q^3+3\mu_Q\sigma_Q^2) + \frac{8\alpha^3\mu_Q^3}{1-\phi^3} + \frac{\alpha^2(\mu_Q^2+\sigma_Q^2)\mu_Q}{1-\phi^2} \\
\kappa_4 &= \frac{\beta\sigma^2+\alpha}{1-\phi}(\mu_Q^4+6\mu_Q^2\sigma_Q^2+3\sigma_Q^4) + \frac{\alpha^2}{1-\phi^2}(14\mu_Q^4+36\mu_Q^2\sigma_Q^2+6\sigma_Q^4) \\
&\quad + \frac{\alpha^3}{1-\phi^3}(48\mu_Q^4+48\mu_Q^2\sigma_Q^2) + \frac{48\alpha^4\mu_Q^4}{1-\phi^4} \\
\kappa_{11} &= \frac{2\alpha^2\mu_Q^2\phi}{1-\phi^2} \\
\kappa_{12} &= \frac{2\alpha^2\mu_Q\phi(\mu_Q^2+\sigma_Q^2)}{1-\phi^2} + \frac{2\alpha^3\mu_Q^3\phi}{1-\phi^3} \\
\kappa_{22} &= \frac{8\alpha^3\mu_Q^2\phi(\mu_Q^2+\sigma_Q^2)}{1-\phi^3} + \frac{2\alpha^2\phi(\mu_Q^2+\sigma_Q^2)^2}{1-\phi^2} + \frac{16\alpha^3\mu_Q^2\phi^2(\mu_Q^2+\sigma_Q^2)}{1-\phi^3} + \frac{48\alpha^4\phi^2\mu_Q^4}{1-\phi^4}.
\end{aligned}
$$

The analytic expressions of corresponding moments can be then calculated using the relationship given by Kendall and Stuart (1974),

$$\mu_1 = \kappa_1$$

$$
\begin{aligned}
\mu_2 &= \kappa_2 \\
\mu_3 &= \kappa_3 \\
\mu_4 &= \kappa_4 + 3\mu_2^2 \\
\mu_{11} &= \kappa_{11} \\
\mu_{12} &= \kappa_{12} \\
\mu_{22} &= \kappa_{22} + \mu_2^2 + 2\mu_{11}^2. \quad \blacksquare
\end{aligned}
$$

Appendix C

Asymptotic 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

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

and

$$
\begin{aligned}
s(\boldsymbol{\theta}) &= \int_0^1 \int_0^1 \{(Re\ c_n(\boldsymbol{r}) - Re\ c(\boldsymbol{r}, \boldsymbol{\theta}))^2 \\
&\quad + (Im\ c_n(\boldsymbol{r}) - Im\ c(\boldsymbol{r}, \boldsymbol{\theta}))^2\}\, dr_1\, dr_2
\end{aligned}
$$

Now since

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

and

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

Then

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

where

$$
\begin{aligned}
k_j(\boldsymbol{\theta}) &= \Big[\int_0^1 \int_0^1 \frac{\partial Re\ c(\boldsymbol{r}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} (\cos(\boldsymbol{r}'\boldsymbol{z}_j) - Re\ c(\boldsymbol{r}, \boldsymbol{\theta})) \\
&\quad + \frac{\partial Im\ c(\boldsymbol{r}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} (\sin(\boldsymbol{r}'\boldsymbol{z}_j) - Im\ c(\boldsymbol{r}, \boldsymbol{\theta})) \Big]\, dr_1\, dr_2
\end{aligned}
$$

Consequently

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

where

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

and is given by:

$$\begin{aligned}
A(\boldsymbol{\theta}) \;=\; & \lim_{n\to\infty}\frac{1}{n}\Big\{\int_0^1\int_0^1\int_0^1\int_0^1\Big[\frac{\partial Re\ c(\boldsymbol{r},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial Re\ c(\boldsymbol{s},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\sum_j\sum_k cov(\cos(\boldsymbol{r}'\boldsymbol{z}_j),\cos(\boldsymbol{s}'\boldsymbol{z}_k)) \\
& +\frac{\partial Re\ c(\boldsymbol{r},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial Im\ c(\boldsymbol{s},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\sum_j\sum_k cov(\cos(\boldsymbol{r}'\boldsymbol{z}_j),\sin(\boldsymbol{s}'\boldsymbol{z}_k)) \\
& +\frac{\partial Im\ c(\boldsymbol{r},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial Re\ c(\boldsymbol{s},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\sum_j\sum_k cov(\sin(\boldsymbol{r}'\boldsymbol{z}_j),\cos(\boldsymbol{r}'\boldsymbol{z}_k)) \\
& +\frac{\partial Im\ c(\boldsymbol{r},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial Im\ c(\boldsymbol{s},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\sum_j\sum_k cov(\sin(\boldsymbol{r}'\boldsymbol{z}_j),\sin(\boldsymbol{s}'\boldsymbol{z}_k))\Big]\Big\}\,dr_1\,dr_2\,ds_1\,ds_2
\end{aligned}$$

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

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

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

$$\begin{aligned}
A(\boldsymbol{\theta}) \;=\; & \lim_{n\to\infty} n\Big\{\int_0^1\int_0^1\int_0^1\int_0^1\Big[\frac{\partial Re\ c(\boldsymbol{r},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial Re\ c(\boldsymbol{s},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\cdot(\Omega_{RR})_{\boldsymbol{r},\boldsymbol{s}} \\
& +2\frac{\partial Re\ c(\boldsymbol{r},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial Im\ c(\boldsymbol{r},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\cdot(\Omega_{RI})_{\boldsymbol{r},\boldsymbol{s}} \\
& +\frac{\partial Im\ c(\boldsymbol{r},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial Im\ c(\boldsymbol{s},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\cdot(\Omega_{II})_{\boldsymbol{r},\boldsymbol{s}}\Big]\Big\}\,dr_1\,dr_2\,ds_1\,ds_2
\end{aligned}$$

Also we note that

$$E\left[\frac{\partial^2 s(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right] \;=\; -\frac{2}{n}\sum_{j=1}^n E\left[\frac{\partial k_j(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right]$$

$$
\begin{aligned}
&= \frac{2}{n}\sum_{j=1}^{n}\int_0^1\int_0^1\Big[\frac{\partial Re\ c(\boldsymbol{r},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial Re\ c(\boldsymbol{r},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} \\
&\quad +\frac{\partial Im\ c(\boldsymbol{r},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial Im\ c(\boldsymbol{r},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\Big]\,dr_1\,dr_2 \\
&= -2\int_0^1\int_0^1\Big[\frac{\partial Re\ c(\boldsymbol{r},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial Re\ c(\boldsymbol{r},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} \\
&\quad +\frac{\partial Im\ c(\boldsymbol{r},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial Im\ c(\boldsymbol{r},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\Big]\,dr_1\,dr_2 \\
&= -2B(\boldsymbol{\theta})
\end{aligned}
$$

Thus standard asymptotic theory results in

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

References

[1] Andersen, T. and B. Sorensen. (1996). GMM estimation of a stochastic volatility model: A Monte Carlo study. *Journal of Business and Economic Statistics.* **14** 329–352.

[2] Ball C. and W. Torous. (1985). On jumps in common stock prices and their impact on call option pricing. *Journal of Finance.* **40** 155–173.

[3] Berntsen, J., T.O. Espelid, and Genz, A. (1991). Algorithm 698: DCUHRE-an adaptive multidimensional integration routine for a vector of integrals. *ACM Transactions on Mathematical Software.* **17** 452–456.

[4] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics.* **31** 307–327.

[5] Clark, P.K. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica.* **41** 135–155.

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

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

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

[9] Engle, R.F., D.M. Lilien, and R.P. Robins. (1987). Estimating time varying risk premia in the term structure: The ARCH-M model. *Econometrica.* **55** 391–407.

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

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

[12] Gallant, A.R., P.E. Rossi, and G. Tauchen. (1992). Stock prices and volumn. *Review of Financial Studies.* **5** 199–242.

[13] Hansen, L.P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica.* **50** 1029–1054.

[14] Hsieh, D. (1991). Chaos and nonlinear dynamics: Applications to financial markets. *Journal of Finance.* **46** 1839–1877.

[15] 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.

[16] Jarrow, B. and E. Rosenfeld. (1984). Jump risks and the intertemporal capital asset pricing model. *Journal of Business.* **57** 337–351.

[17] Jorion, P. (1988). On jump processes in the foreign exchange and stock markets. *Review of Financial Studies.* **1** 427–445.

[18] Kendall, M. and A. Stuart. (1974). *The Advanced Theory of Statistics.* Charles Griffin and Company, London.

[19] Knight, J.L. and J. Yu. (1999). The empirical characteristic function in time series estimation. Submitted.

[20] Knight, J.L. and S.E. Satchell. (1997). The cumulant generating function estimation method. *Econometric Theory.* **13** 170–184.

[21] Knight, J.L. and S.E. Satchell. (1998). GARCH processes – some exact results, some difficulties and a suggested remedy. In Knight, J.L. and S.E. Satchell, editors, *Forecasting Volatility in the Financial Markets.* Oxford: Butterworth-Heineman.

[22] Knight, J.L., S.E. Satchell, and Y. Yoon. (1993). A model of a continuous time GARCH process applied to UK stocks. The University of Western Ontario.

[23] Lin, S., J.L. Knight, and S.E. Satchell. (1999). Modelling intraday equity prices and volatility using information arrivals - a comparative study of different choices of informational proxies. In Lequeux, P., editor, *Financial Markets Tick by Tick,* pages 27–64. John Wiley and Sons.

[24] Merton, R. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics.* **3** 125–144.

[25] Nelson, D. (1990). ARCH models as diffusion approximations. *Journal of Econometrics.* **45** 7–38.

[26] Powell, M.J.D. (1964). An efficient method for finding the minimum of a function several variables without calculating derivatives. *The Computer Journal.* **7** 155–162.

[27] Press, S.J. (1967). A compound events models of security prices. *Journal of Business.* **40** 317–335.

[28] 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.

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

[30] Schwartz, G. (1978). Estimating the dimension of a model. *Annals of Statistics.* **6** 461–464.

[31] Singleton K.J. (1999). Estimation of affine asset pricing models using the empirical characteristic function. Stanford University.

[32] Tauchen, G. and M. Pitts. (1983). The price variability-volume relationship on speculative markets. *Econometrica.* **51** 485–505.

[33] Taylor, S. (1986). *Modelling Financial Time Series.* John Wiley, New York.

[34] Tran, K.C. (1998). Estimating mixtures of normal distributions via empirical characteristic function. *Econometric Review.* **17** 167–83.

Table 1

Monte Carlo Study with True Values of Parameters $\beta = 0.5$, $\sigma_Q^2 = 1.0$

	$\beta = 0.5$		$\sigma_Q^2 = 1.0$	
	ECF	GMM	ECF	GMM
MEAN	.5098	.5732	.9792	.9630
MED	.5078	.5576	.9746	.9375
MIN	.3408	.0965	.6342	.4116
MAX	.6481	1.847	1.340	1.643
MSE	.00276	.0485	.0184	.0394
RMSE	.0525	.220	.1358	.198

Note: The reported statistics are based on 500 simulated replications each with sample size equal to 2,000. The jump diffusion model is defined by $X(t) = \mu + \sigma B(1) + \sum_{n=1}^{\Delta N(t)} Q(n)$ with $\lambda(t) = \alpha\nu^2(t-1) + \beta Var(X(t-1)|I(t-2))$ where $\lambda(t)$ is the intensity of the Poisson process $N(t)$.

Table 2

Descriptive Statistics of S&P 500 Returns (1980-1987)

	Mean	SD	Skewness	Kurtosis	Minimum	Maximum
$X(t)$	0.041	1.11	-2.56	51.68	-19.35	8.247

Note: $X(t)$ denotes the daily S&P 500 index return over 1980-1987.

Table 3

Descriptive Statistics of S&P 500 Returns (1980-1987)

	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6	ρ_7	ρ_8
$X(t)$	0.083	-0.043	-0.024	-0.046	0.049	0.029	0.012	-0.020
$X^2(t)$	0.125	0.183	0.094	0.019	0.132	0.030	0.011	0.054

Note: $X(t)$ denotes the daily S&P 500 index return over 1980-1987. ρ_j denotes the autocorrelation coefficient of order j.

Table 4

Estimation of the Jump Diffusion Model via ECF and GMM

Method	μ	σ^2	α	β	μ_Q	σ_Q^2
ECF	0.055	.79	0.00016	0.047	-0.16	6.18
	(0.026)	(2.050)	(0.291)	(0.551)	(0.249)	(1.587)
GMM	0.089	.065	0.0637	.00059	-0.913	17.56
	(0.097)	(2.063)	(0.335)	(0.360)	(0.259)	(1.541)

Note: The numbers in brackets are standard errors of the estimates. The jump diffusion model is defined by $X(t) = \mu + \sigma B(1) + \sum_{n=1}^{\Delta N(t)} Q(n)$ with $\lambda(t) = \alpha\nu^2(t-1) + \beta Var(X(t-1)|I(t-2))$ where $\lambda(t)$ is the intensity of the Poisson process $N(t)$.

Table 5

Fitting a GARCH(1,1) Model and an ARCH-M to S&P 500 Returns

	μ	ω	α	β	δ
GARCH(1,1)	0.0618	0.035	0.096	0.876	·
	(0.020)	(0.009)	(0.004)	(0.011)	
ARCH-M	-0.078	0.034	0.095	0.878	0.152
	(0.115)	(0.009)	(0.004)	(0.011)	(0.125)

Note: The ML estimates are obtained when the models are fitted into S&P 500 daily returns for the period 1980-1987. The numbers in brackets are standard errors of the estimates. The GARCH(1,1) model is defined by $x_t = \mu + \sigma_t\epsilon_t$ with $\sigma_t^2 = \omega + \alpha(x_{t-1} - \mu)^2 + \beta\sigma_{t-1}^2$. The ARCH-M model is defined by $x_t = \mu + \delta\sigma_t + \sigma_t\epsilon_t$ with $\sigma_t^2 = \omega + \alpha(x_{t-1} - \mu - \delta\sigma_{t-1})^2 + \beta\sigma_{t-1}^2$.

Table 6

Comparison between Data Moments and Model Moments Calculated from Different Models

	Standard Deviation	Skewness	Kurtosis
Data	1.11	-2.56	51.68
Jump Diffusion	1.06	-1.33	7.88
GARCH	1.12	0.00	4.50
ARCH-M	1.14	0.00	4.54
SV	1.10	0.00	4.39

Note: The data moments of S&P 500 returns are calculated from daily log returns ($\times 100$) over the sample period from 1980 through 1987. The jump diffusion model is defined by $X(t) = \mu + \sigma B(1) + \sum_{n=1}^{\Delta N(t)} Q(n)$ with $\lambda(t) = \alpha\nu^2(t-1) + \beta Var(X(t-1)|I(t-2))$ where $\lambda(t)$ is the intensity of the Poisson process $N(t)$. The GARCH(1,1) model is defined by $x_t = \mu + \sigma_t\epsilon_t$ with $\sigma_t^2 = \omega + \alpha(x_{t-1} - \mu)^2 + \beta\sigma_{t-1}^2$. The variance and kurtosis for the GARCH(1,1) model are $\frac{\omega}{1-\alpha-\beta}$ and $3\frac{(1-\alpha-\beta)^2+2(\alpha+\beta)(1-\alpha-\beta)}{1-\beta^2-3\alpha^2-2\alpha\beta}$ respectively. The ARCH-M model is defined by $x_t = \mu+\delta\sigma_t+\sigma_t\epsilon_t$ with $\sigma_t^2 = \omega+\alpha(x_{t-1}-\mu-\delta\sigma_{t-1})^2+\beta\sigma_{t-1}^2$. The variance and kurtosis for the ARCH-M model are $\frac{(\delta^2+1)\omega}{1-\alpha-\beta}$ and $\frac{\delta^4+6\delta^2+3}{(\delta^2+1)^2} \times \frac{(1-\alpha-\beta)^2+2(\alpha+\beta)(1-\alpha-\beta)}{1-\beta^2-3\alpha^2-2\alpha\beta}$ respectively. The SV model is defined by $x_t = \exp(0.5h_t)e_t$ with $h_t = \lambda + \alpha h_{t-1} + v_t$. The variance and kurtosis for the SV model are $\exp(\frac{\sigma^2}{2(1-\alpha^2)})$ and $3\exp(\frac{\sigma^2}{1-\alpha^2})$ respectively.

Figure 1

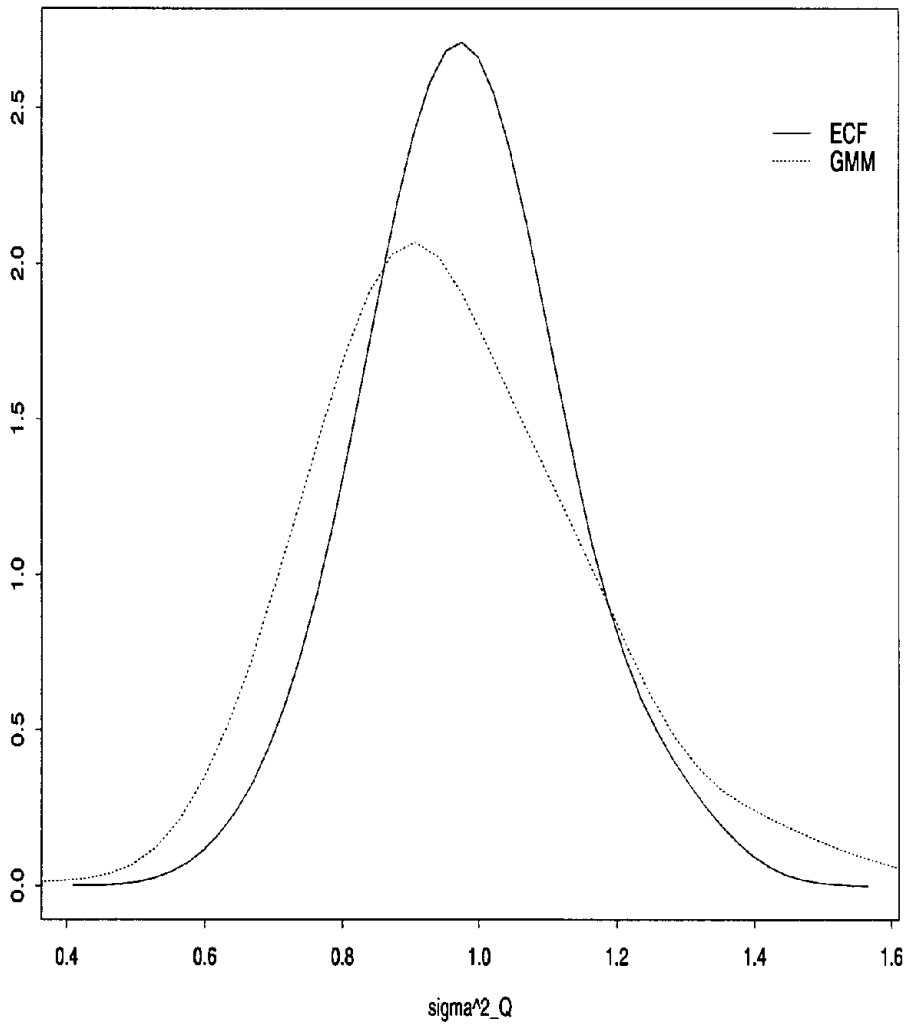

Figure 2

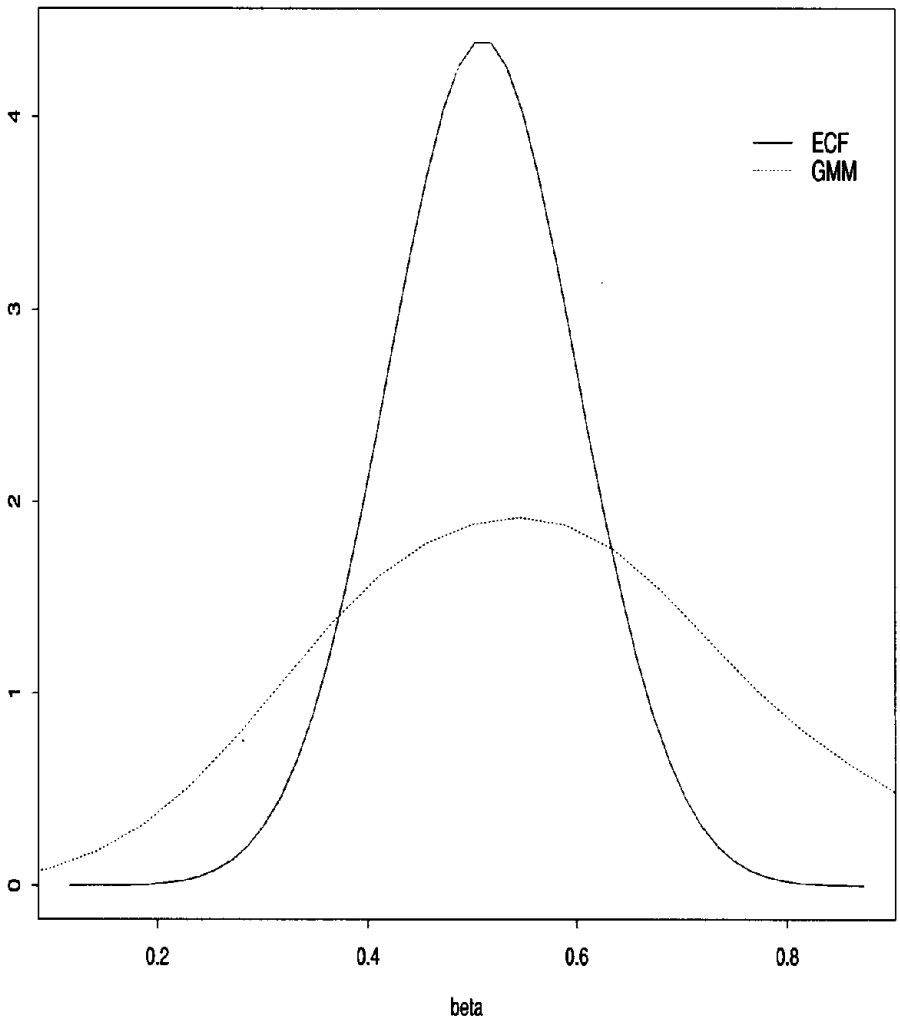

Figure 3

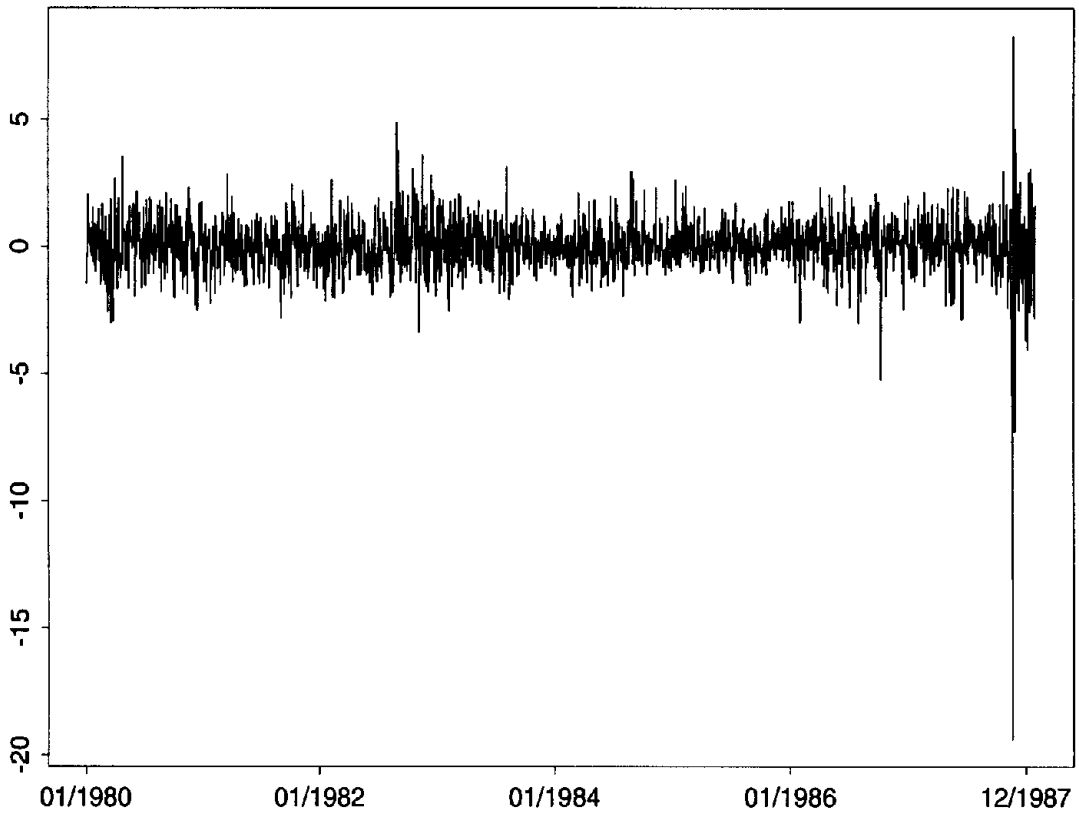

Figure 4