

Empirical Characteristic Function Estimation and Its Applications

Jun Yu*

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

ABSTRACT

This paper reviews the method of model-fitting via the empirical characteristic function. The advantage of using this procedure is that one can avoid difficulties inherent in calculating or maximizing the likelihood function. Thus it is a desirable estimation method when the maximum likelihood approach encounters difficulties but the characteristic function has a tractable expression. The basic idea of the empirical characteristic function method is to match the characteristic function derived from the model and the empirical characteristic function obtained from data. Ideas are illustrated by using the methodology to estimate a diffusion model that includes a self-exciting jump component. A Monte Carlo study shows that the finite sample performance of the proposed procedure offers an improvement over a GMM procedure. An application using over 72 years of DJIA daily returns reveals evidence of jump clustering.

Key Words: Diffusion process; Poisson jump; Self-exciting; GMM; Jump clustering.

JEL Classification: C13; C15; C22; G10.

*Correspondence: Jun Yu, Department of Economics, University of Auckland, Private Bag 92019, 3A Symonds St., Auckland, New Zealand; Fax: 64-9-373-7427; E-mail: j.yu@auckland.ac.nz.

1. INTRODUCTION

Traditionally the maximum likelihood (ML) approach is widely favored in economic and financial applications due to its generality and asymptotic efficiency. In a variety of applications in economics and finance, the ML method can be difficult. The difficulties arise when the likelihood function is not tractable or not bounded over the parameter space or does not have a closed form expression in the sense that it is not expressible in terms of known elementary functions.

Although the likelihood function can be unbounded, its Fourier transform is always bounded. Moreover, while the likelihood function is not tractable or has no closed form solution, the Fourier transform can have a closed form expression. Since the Fourier transform of the density function is the characteristic function (CF), one can exploit the empirical characteristic function (ECF) to estimate the system parameters.

A main purpose of this paper is to explain the estimation method via the ECF to applied researchers. The paper also summarizes the models for which the ML approach encounters difficulties but the CF has a closed form expression and hence the ECF method can be a viable estimation method. The statistical properties of the ECF estimators are also discussed.

Work in this area has been initiated by Parzen (1962), and can be dichotomized according to whether we are dealing with independent, identically distributed (iid) or dependent stationary stochastic processes. Section 2 reviews various ECF procedures both in the iid and non-iid cases. Section 3 discusses important assumptions for the ECF procedures that applied researchers should be aware of, together with some asymptotic properties for the ECF estimators. Section 4 lists some examples for which the likelihood function is not bounded over the parameter space or does not have a closed form expression. In Sec. 5, I illustrate the ECF procedure to estimate a self-exciting jump diffusion process in a Monte Carlo study and in an empirical study. Section 6 concludes.

2. ECF PROCEDURES

2.1. IID Case

The ECF procedure in the iid case has been previously investigated by Paulson et al. (1975), Heathcote (1977), Feuerverger and Mureika (1977), Bryant and Paulson (1983), Feuerverger and McDunnough (1981b,c), Koutrouvelis (1980), and more recently by Tran (1998) and Carrasco and Florens (2002). The justification for the ECF method is that the CF is the Fourier–Stieltjes transform of the cumulative distribution function (CDF) and hence there is a one-one correspondence between the CF and CDF. As a consequence, the ECF retains all information in the sample. This observation suggests that estimation and inference via the ECF should work as efficiently as the likelihood-based approaches.



Suppose the CDF of X is $F(x; \theta)$ which depends on a K -dimensional vector of parameters θ . The CF is defined by

$$c(r; \theta) = E[\exp(ir X)] = \int \exp(ir x) dF(x; \theta),$$

and the ECF is the sample counterpart of the CF defined by

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

where $i = \sqrt{-1}$, $\{X_j\}_{j=1}^n$ is an iid sequence, $F_n(x)$ is the empirical CDF, and r is the transform variable. Also assume the true value of θ is θ_0 . Note that the CF is a deterministic function of θ while ECF depends on θ_0 only through the observations $\{X_j\}$. θ is suppressed in $c(r; \theta)$ when there is no confusion.

Since the ECF estimator can be treated as a generalized method of moment (GMM) estimator of Hansen (1982), it is worth briefly reviewing GMM first. Suppose one has the following l moment conditions:

$$E(f(X_j; \theta_0)) = 0,$$

where $f : R \times R^K \rightarrow R^l$. Further assume that the strong law of large numbers is invoked so that we have the following result for the sample moments:

$$\frac{1}{n} \sum_{j=1}^n f(X_j; \theta) \xrightarrow{a.s.} E(f(X_j; \theta)).$$

The basic idea of GMM estimation is to minimize a distance measure between the sample moments and the population moments, that is,

$$\min_{\theta} \frac{1}{n} \sum_{j=1}^n f(X_j; \theta)' W_n \frac{1}{n} \sum_{j=1}^n f(X_j; \theta),$$

where W_n is a positive semidefinite weighting matrix which converges to a positive definite matrix W_0 almost surely. Under some regularity conditions, the GMM estimator is consistent and asymptotically normally distributed for arbitrary weighting matrices. When the system is just identified ($K = l$), the GMM estimator does not depend on the choice of W_n and basically solves the estimation equation: $(1/n) \sum_{j=1}^n f(X_j; \theta) = 0$. As a result, this is the method of moment estimation. When the system is over identified ($K < l$), Hansen (1982) shows that if $W_0 = \Sigma^{-1}$, the GMM estimator is asymptotically efficient in the sense that the covariance matrix of the GMM estimator is minimized, where Σ is the long run covariance matrix of $f(X_j; \theta_0)$. It should be pointed out that in general GMM efficiency is different from ML efficiency and the GMM estimator is optimal only for the given moment conditions $f(X_j; \theta)$. When moment conditions are different, GMM efficiency can vary. Hence GMM is sub-optimal relative to ML.



Motivated from the recognition that two distribution functions are equal if and only if their CFs agree on $-\infty < r < \infty$ (Lukacs, 1970, p. 28), the general idea for ECF estimation is to minimize various distance measures between the ECF and CF.

To link the ECF method to GMM, define the following function based on the ECF,

$$h(r, X_j; \theta) = \exp(ir X_j) - c(r; \theta). \tag{2.1}$$

Obviously $E(h(r, X_j; \theta_0)) = 0, \forall r$. Consequently, a finite set of moment conditions or a continuum of moment conditions can be constructed, depending how the transform variable r is chosen.

If r is chosen to be a set of discrete points, the procedure is called the discrete ECF method and is used by Tran (1998) to estimate the mixtures of normal distributions, following the suggestion made by Quandt and Ramsey (1978) and Schmidt (1982).

Suppose q discrete points r_1, \dots, r_q are used and define

$$f(X_j; \theta) = (Re[h(r_1, X_j; \theta)], \dots, Re[h(r_q, X_j; \theta)], Im[h(r_1, X_j; \theta)], \dots, Im[h(r_q, X_j; \theta)])',$$

where $Re[\cdot]$ and $Im[\cdot]$ are the real and imaginary parts of a complex number. By construction $E(f(X_j; \theta_0)) = 0$. This forms $2q$ (usually larger than l) moment conditions. Also note that the strong law of large numbers applies here (see, for example, Feuerverger and Mureika, 1977).

When $W_n = I$, this discrete ECF estimator is basically the first stage GMM estimator and can also be thought of as the nonlinear OLS regression of V_n on V_θ , where

$$V_n = (Re[c_n(r_1)], \dots, Re[c_n(r_q)], Im[c_n(r_1)], \dots, Im[c_n(r_q)])'$$

and

$$V_\theta = (Re[c(r_1; \theta)], \dots, Re[c(r_q; \theta)], Im[c(r_1; \theta)], \dots, Im[c(r_q; \theta)])'$$

Obviously, $(1/n)\sum_{j=1}^n f(X_j; \theta) = V_n - V_\theta$.^a However, the resulting estimator cannot attain GMM efficiency since by construction the covariance matrix of V_n is not diagonal. Denote the covariance matrix of V_n by Ω and it has been shown that (see, for example, Feuerverger and Mureika, 1977)

$$\Omega = \begin{bmatrix} \Omega_{RR} & \Omega_{RI} \\ \Omega_{IR} & \Omega_{II} \end{bmatrix},$$

^aAlthough we separate the real and imaginary parts for ease of understanding, Carrasco and Florens (2002) argue that this separation is not needed as most software packages allow for operations of complex numbers. In this case, the moment conditions are $f(X_j; \theta) = (h(r_1, X_j; \theta), \dots, h(r_q, X_j; \theta))'$.



where the elements in the partitions associated with r_j and r_k are given by

$$\begin{aligned}
 (\Omega_{RR})_{jk} &= \frac{1}{2}(Re[c(r_j + r_k)] + Re[c(r_j - r_k)]) - Re[c(r_j)]Re[c(r_k)], \\
 (\Omega_{RI})_{jk} &= \frac{1}{2}(Im[c(r_j + r_k)] - Im[c(r_j - r_k)]) - Re[c(r_j)]Im[c(r_k)], \\
 (\Omega_{II})_{jk} &= \frac{1}{2}(Re[c(r_j + r_k)] - Re[c(r_j - r_k)]) - Im[c(r_j)]Im[c(r_k)].
 \end{aligned}$$

Using the covariance matrix, Tran (1998) estimates θ by finding the minimizer of $(V_n - V_\theta)' \hat{\Omega}^{-1} (V_n - V_\theta)$, where $\hat{\Omega}$ is a consistent estimate of Ω . The procedure can be thought of as the second stage GMM estimation or the non-linear GLS regression of V_n on V_θ and hence yields GMM efficient estimators.

Just like how GMM depends on the choice of moment conditions, the above ECF procedure hinges on the choice of a grid of discrete points. To select the optimal discrete points, two choices have to be made: how many discrete points (i.e., q) and which discrete points should be used. These correspond, respectively, to how many and which moment conditions should be used for GMM. For a given q , Schmidt (1982) suggests selecting the grid that minimizes the determinant of the asymptotic covariance matrix and has found it is best to select all the points close together. Feuerverger and McDunnough (1981c) show that the asymptotic matrix can be made arbitrarily close to the Cramér–Rao lower bound (i.e., ML efficiency) provided that q is sufficiently large and the grid is sufficiently fine and extended. They further suggest that the grid should be chosen to be equally spaced, i.e., $r_j = \tau j$ for $j = 1, \dots, q$. This suggestion will ease the computational burden but whether or not the ML efficiency is warranted is still an open question. Furthermore, as noted by Carrasco and Florens (2002), when the grid is too fine, the covariance matrix $\hat{\Omega}$ becomes singular and hence the ECF estimator can not be computed. The same problem is also identified in Madan and Seneta (1990) in the context of the variance gamma distribution.

When r is chosen continuously, one can minimize

$$\int_{-\infty}^{\infty} |c_n(r) - c(r; \theta)|^2 g(r) dr, \tag{2.2}$$

with $g(r)$ being a continuous weighting function. Or equivalently one can minimize

$$\int_{-\infty}^{\infty} |c_n(r) - c(r; \theta)|^2 dG(r),$$

or solve the following estimation equation

$$\int_{-\infty}^{\infty} w(r)(c_n(r) - c(r; \theta)) dr = 0, \tag{2.3}$$

where $G(r)$ and $w(r)$ are weighting functions.



Since $g(r)$ is a continuous function, the procedure (2.2) basically matches the ECF and CF continuously over an interval and hence can be viewed as a special class of GMM on a continuum of moment conditions given by Carrasco and Florens (2000). To see this, consider the objective function of the GMM procedure based on a continuum of moment conditions defined in Carrasco and Florens (2000),

$$\iint h_n(r; \theta) g_n(r, s) \overline{h_n(s; \theta)} dr ds, \tag{2.4}$$

where \bar{h} is the conjugate of h . If we choose $g_n(r, s) = g(r)I(r - s)$, $h_n(r; \theta) = (1/n) \sum h(r, X_j; \theta)$, (2.4) is equivalent to (2.2).

The above continuous ECF procedure has been used in Press (1972), Paulson et al. (1975), Thornton and Paulson (1977), and more recently in Carrasco and Florens (2002). The advantage of using a continuum of moment conditions is that in theory with a judiciously chosen weighting function it results in full ML efficiency (Carrasco and Florens, 2002). While an arbitrary continuous function with bounded total variation for $g(r)$ can guarantee consistency, in practice an exponential weighting function is often used. Although the exponential weight has the numerical advantage associated with quadratures, in general, the resulting ECF estimator from the exponential weight, say $\exp(-r^2)$, is less efficient than the ML estimator.

To see this, suppose the random sample X_1, \dots, X_n is from $N(\mu, \sigma^2)$, where σ^2 is known, and we want to estimate μ . It is easy to show that

$$\begin{aligned} \int_{-\infty}^{\infty} |c_n(r) - c(r)|^2 \exp(-r^2) dr &= \int_{-\infty}^{\infty} \left| \frac{1}{n} \sum_{j=1}^n e^{irX_j} - e^{ir\mu - \frac{\sigma^2 r^2}{2}} \right|^2 e^{-r^2} dr \\ &= \frac{\pi^{1/2}}{n^2} \sum_{i=1}^n \sum_{j=1}^n e^{-\frac{1}{4}(X_i - X_j)^2} + \left(\frac{\pi}{1 + \sigma^2} \right)^{1/2} \\ &\quad - \left(\frac{2}{n} \right) \left(\frac{\pi}{1 + \sigma^2/2} \right)^{1/2} \sum_{j=1}^n \exp \left(-\frac{(X_j - \mu)^2}{4 + 2\sigma^2} \right). \end{aligned}$$

The first order condition gives the following estimating equation

$$\sum_{j=1}^n (X_j - \mu) \exp \left(-\frac{(X_j - \mu)^2}{4 + 2\sigma^2} \right) = 0.$$

The asymptotic relative efficiency of the ECF estimator of μ is

$$\left\{ \frac{1 + 2\sigma^2 + \frac{3}{4}\sigma^4}{1 + 2\sigma^2 + \sigma^4} \right\}^{3/2},$$

and is generally less than 1. When $\sigma^2 = 1$, for instance, the asymptotic relative efficiency is about 95%; as σ^2 tends to 0 it tends to 100% but as σ^2 tends to ∞ it tends to about 65%.



The optimal weight obtained by Feuerverger and McDunnough (1981b) using the Parsaval identity is given by

$$w^*(r) = \left(\frac{1}{2\pi}\right) \int \exp(-irx) \frac{\partial \log f_{\theta}(x)}{\partial \theta} dx. \tag{2.5}$$

The weight is optimal in the sense that the resulting estimator from Eq. (2.3) attains ML efficiency. Obviously when the likelihood function has no closed form expression, the optimal weight is unknown.

Based on the results obtained in Carrasco and Florens (2000), Carrasco and Florens (2002) provide a solution to this dilemma which also avoids the singularity problem discussed in the discrete case. According to Carrasco and Florens (2002), a covariance operator associated with a continuum of moments (i.e., $h(r, X_j; \theta)$), perturbed by a regularization parameter α_n , is used in the second stage estimation. The perturbation guarantees that the inverse of the covariance operator always exists. Denoting this covariance operator by Ω , Carrasco and Florens (2002) obtain the expression for the kernel of Ω (called $g^*(r, s)$)

$$g^*(r, s) = c(r - s) - c(r)c(-s).$$

The asymptotic variance of the resulting estimator is shown to reach the Cramér–Rao lower bound when $n\alpha_n \rightarrow \infty$ and $\alpha_n \rightarrow 0$. The intuitions for this ML efficiency are as follows. First, relative to the optimal scheme of the ECF approach based on a grid of discrete points, more moment conditions and hence more information are used here. As a result, the estimator should be more efficient. Second, relative to the non-optimal continuous ECF approach which uses the full continuum but a sub-optimal weight, it provides an optimal GMM scheme by using the information in the covariance.

2.2. Non-iid Stationary Case

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. Like the marginal empirical CDF, the marginal ECF may not identify all the parameters in the case of dependent data or may result in a loss in efficiency. Consequently, approaches based on the joint CF and conditional CF have been used in the literature.

2.2.1. Joint ECF

The approach via the joint CF is used in Feuerverger (1990), Knight and Satchell (1996), Yu (1998), Knight and Yu (2002), Carrasco et al. (2002), and Jiang and Knight (2002). The procedures involve moving blocks of data. Denote the moving blocks for X_1, X_2, \dots, X_T as $\mathbf{Z}_j = (X_j, \dots, X_{j+p})'$, $j = 1, \dots, T - p$. Thus



each block has $p + 1$ observations and p overlapping periods with its adjacent blocks. The characteristic function of each block is defined as

$$c(\mathbf{r}; \boldsymbol{\theta}) = E(\exp(i\mathbf{r}'\mathbf{Z}_j)),$$

where $\mathbf{r} = (r_1, \dots, r_{p+1})'$ and hence the transform variable is of $p + 1$ dimensions. The joint ECF is defined as

$$c_n(\mathbf{r}) = \frac{1}{n} \sum_{j=1}^n \exp(i\mathbf{r}'\mathbf{Z}_j),$$

where $n = T - p$.

To estimate the parameter via the joint ECF one can minimize a distance measure between the joint CF and joint ECF,

$$\int \cdots \int |c(\mathbf{r}; \boldsymbol{\theta}) - c_n(\mathbf{r})|^2 g(\mathbf{r}) d\mathbf{r}, \tag{2.6}$$

or

$$\int \cdots \int |c(\mathbf{r}; \boldsymbol{\theta}) - c_n(\mathbf{r})|^2 dG(\mathbf{r}), \tag{2.7}$$

or solve the following estimating equation

$$\int \cdots \int (c(\mathbf{r}; \boldsymbol{\theta}) - c_n(\mathbf{r}))w(\mathbf{r})d\mathbf{r} = 0, \tag{2.8}$$

where $g(\mathbf{r})$, $G(\mathbf{r})$ and $w(\mathbf{r})$ are weighting functions. Under suitable conditions Eqs. (2.6)–(2.8) are equivalent.

As in the iid environment, the ECF estimator is a special case of GMM where the moment conditions are $\exp(i\mathbf{r}'\mathbf{Z}_j) - c(\mathbf{r}; \boldsymbol{\theta}), \forall \mathbf{r} \in R^{p+1}$. Since the transform variable \mathbf{r} is a vector, the moment conditions include both marginal and joint moments.

The discrete ECF procedure is advocated in Feuerverger (1990), Knight and Satchell (1996, 1997) and further discussed in Yu (1998). It corresponds to minimizing Eq. (2.6) with $g(\mathbf{r})$ being a function which takes a finite number of non-zero values. Compared to the iid case, the situation is more complicated since a set of $p + 1$ dimensional vectors must be selected. Feuerverger (1990) argues that under some regularity conditions, if p is sufficiently large and the discrete vectors are sufficiently fine and extended, the resulting estimators can be made arbitrarily close to the Cramér–Rao lower bound. The result is of theoretical interest but offers no guidance as to the practical choice of an optimal set of vectors.

Defining V_n and V_θ in the same way as in the iid case but based on the moving blocks, Knight and Satchell (1997) suggest a multi-step procedure on the implementation of the discrete ECF method which is basically an optimal GMM scheme. The main idea is as follows. Firstly, choose p and q and an arbitrary set of vectors, $(\mathbf{r}_1, \dots, \mathbf{r}_q)$. Secondly, choose $\boldsymbol{\theta}$ to minimize $(V_n - V_\theta)'(V_n - V_\theta)$ to obtain a consistent estimate for Ω , say $\hat{\Omega}$. Thirdly, choose elements in $(\mathbf{r}_1, \dots, \mathbf{r}_q)$ to minimize some measure of the asymptotic covariance matrix of the nonlinear GLS estimator.



Fourthly, based on the resulting (r_1, \dots, r_q) from Step 3, we repeat Step 1 to obtain another consistent estimate for Ω , say $\hat{\Omega}$. Finally, choose θ to minimize $(V_n - V)' \hat{\Omega}^{-1} (V_n - V)$. The minimizer is the desirable estimator and should be efficient in the GMM sense. Knight and Satchell (1996) give the expression of the covariance matrix Ω of V_n for stationary processes and subsequently implement the procedure for a Gaussian MA(1) model with $p = 2, q = 5$ but ignore Step 3.

To improve the GMM efficiency, Yu (1998) implements the above procedure without missing any step. Apart from the well-known difficulties associated with the choice of p and q , Yu (1998) also identifies several numerical difficulties. In particular, in Step 3 it is not clear how many elements in (r_1, \dots, r_q) should be chosen. That is, should one choose the entire set of vectors or should one choose a set of elements in the vectors with some pre-specified restrictions, such as even spacing? Clearly the choice of the entire set of vectors would generally gain in asymptotic efficiency but would also increase the computational burden for numerical optimization in Step 3.^b Moreover, in Step 4 the estimated covariance matrix Ω often becomes singular when too many elements are chosen in Step 3. It seems that this singularity problem would be worse as p or q or both increase.

Alternatively one can match the joint CF and joint ECF continuously. In this continuous ECF procedure the weighting function is a continuous function and hence the transform variable is integrated out. As in the iid case, this procedure can be treated as GMM based on a continuum of moment conditions. Yu (1998) and Knight and Yu (2002) consider two continuous procedures. When an unequal weight is used, the procedure is referred to as the WLS-ECF method. The procedure is referred to as the GLS-ECF method when the weighting function in (2.8), $w(r)$, is given by

$$w^*(r) = \int \dots \int \exp(-ir'Z_j) \frac{\partial \log f(X_{j+p}|X_j, \dots, X_{j+p-1})}{\partial \theta} dX_j \dots dX_{j+p}, \quad (2.9)$$

where $f(X_{j+p}|X_j, \dots, X_{j+p-1})$ is the conditional score function. This weight is optimal in the sense that the asymptotic variance of the GLS-ECF estimator can be made arbitrarily close to the Cramér–Rao lower bound when p is large enough. Knight and Yu (2002) derive the expressions for $w^*(r)$ for the Gaussian ARMA models in which the conditional score is known, and show that in finite samples the GLS-ECF estimator has reasonably good efficiency in comparison with ML. However, this quantity is not calculable if the conditional score is unknown. In a recent study in progress Jiang and Knight (2003) suggest approximating the optimal weight using the Edgeworth expansion to approximate the conditional score in (2.9) for Markov processes. As an alternative, Carrasco et al. (2002) propose to use an optimal GMM scheme based on a continuum of moment conditions. It is interesting to compare the performances of these two alternative approaches.

It is very important to recognize that when using the joint ECF, an additional choice needs to be made, which is that of the overlapping size of the moving

^bFor example, when $p = 2, q = 5$, the optimization problem in Step 3 is of 15 dimensions if the entire set of vectors is chosen.



blocks, p . The choice of p in the context of the ECF estimation is closely related to the well-known difficult choice of the block size in the moving block bootstrap method of Künsch (1989), as well as to the choice of bandwidth in the nonparametric setting. See Politis and White (2004) for an overview of block bootstrap methods and the block size selection problem, and Härdle and Linton (1994) for an overview of the bandwidth selection problem in the nonparametric setting. As a result, one should expect that the ECF estimator can be sensitive to p , similar to the case in the block bootstrap and nonparametric methods.

Ideally an optimal p is selected to minimize the mean square error (MSE) of the ECF estimator. However, the form of the MSE expansion for the ECF estimator has not yet been developed. Knight and Yu (2002, Remark 3.7) point out that the choice of p is related to the dimension of the minimal sufficient statistics. In particular, for a Markov process of order 1, the overlapping moving blocks with block size of 2 form a set of sufficient statistics and hence $p = 1$ is enough. This result is identical to that reached in Bühlmann (1994) in the context of the moving block bootstrap for the AR(1) model. For non-Markov processes, however, since any statistics of dimension less than the sample size is not sufficient (see Arato, 1961), the blocks with a larger p (as long as $p \rightarrow \infty$ as $n \rightarrow \infty$ but $p = o(n)$) will always improve asymptotic efficiency. It seems reasonable to believe that when a non-Markov process can be well approximated by a Markov process of order l , $p = l$ should work well. Furthermore, while for a general non-Markov process a larger p can increase asymptotic efficiency, it will lead to a higher computational inefficiency and hence in practice there is always a trade-off between a large p and a small p .

2.2.2. Empirical Conditional CF

For ease of exposition, I first restrict my attention to Markov processes in this section. The conditional CF (CCF) for a Markov process $\{X_t\}$ is defined by

$$c_{X_t}(r; \theta) = E[\exp(ir X_{t+1}) | X_t, \dots, X_1] = E[\exp(ir X_{t+1}) | X_t], \quad (2.10)$$

where the second equality follows the Markov property. Note that the transform variable is a scalar in the CCF, in contrast to that in the joint CF.

As in the case of the unconditional CF, the estimation based on the empirical CCF (ECCF) can be motivated from GMM which is based on a set of moment conditions but is conditional this time:

$$E(\exp(ir X_{t+1}) - c_{X_t}(r; \theta_0) | X_t) = 0, \quad \forall r.$$

This implies that for any weighting function, $w(\cdot, \cdot)$ (often termed instruments in the GMM literature), we have

$$E(\exp(ir X_{t+1}) - c_{X_t}(r; \theta_0)) = 0, \quad \forall r, \quad (2.11)$$

$$E\left\{ \int (\exp(ir X_{t+1}) - c_{X_t}(r; \theta_0)) w(X_t, r) dr \right\} = 0, \quad (2.12)$$



and

$$E((\exp(ir X_{t+1}) - c_{X_t}(r; \theta_0))w(X_t, s)) = 0, \forall r, s. \quad (2.13)$$

Obviously (2.13) implies (2.12) which, in turn, implies (2.11). As a result, resulting estimators based on (2.11) should be generally less efficient than those based on (2.12) and (2.13).

Using (2.11) to form moment conditions, Chacko and Viceira (2003) consider an optimal GMM scheme which is basically a discrete ECF procedure based on the CCF with a set of integer values assigned to the transform variable. A drawback of their procedure lies in the obvious loss in efficiency.

Singleton (2001) makes use of (2.12) to construct the ECF procedure which solves the estimation equation:

$$\frac{1}{T-1} \int \sum_{t=1}^{T-1} (\exp(ir X_{t+1}) - c_{X_t}(r; \theta))w(X_t, r) = 0, \quad (2.14)$$

where $w(X_t, r)$ is a set of K functions. He further shows that if $w(X_t, r)$ is chosen as

$$w^*(X_t, r) = \left(\frac{1}{2\pi}\right) \int \exp(-ir X_{t+1}) \frac{\partial \log f_{\theta}(X_{t+1} | X_t)}{\partial \theta} dX_{t+1}, \quad (2.15)$$

the resulting estimator reaches the ML efficiency although the actual implementation is infeasible when the conditional score cannot be computed.

To overcome this problem, Singleton suggests approximating the integral in (2.14) with the sum over a finite number of discrete points. In particular, he proposes to fix the interval $[-q\tau, q\tau]$, divide it into $2q + 1$ equally spaced sub-intervals of width τ , and then solve

$$\frac{1}{T-1} \sum_{t=1}^{T-1} \tau \sum_{j=-q}^q w(X_t, j\tau) (\exp(ij\tau X_{t+1}) - c_{X_t}(j\tau; \theta)) = 0. \quad (2.16)$$

This discrete ECF procedure results in a consistent estimator, albeit generally inefficient both in the GMM sense and in the ML sense, for any $q \geq 1$ and w . To improve efficiency, Singleton (2001) chooses $w(X_t, j\tau)$ to be the optimal instruments given in Hansen (1985), and argues that when $q \rightarrow \infty, \tau \rightarrow 0$, this optimal discrete ECF procedure can be made arbitrarily close to the Cramér–Rao lower bound. Unfortunately, as Carrasco et al. (2002) point out, when $q \rightarrow \infty, \tau \rightarrow 0$, the covariance matrix used for obtaining the optimal instruments tends to be singular.

By using (2.13) Carrasco et al. (2002) construct an optimal GMM scheme based on a continuum of moment conditions with $w(X_t, s)$ set to be $\exp(is X_t)$, and show that the resulting estimator can achieve the ML efficiency. ML efficiency is ensured by the fact that a full continuum of moments and the corresponding covariance operator are used. The singularity problem is overcome by attaching a regularization parameter to the covariance operator, as in Carrasco and Florens (2002).



All the estimation procedures discussed above apply to non-Markov processes but the condition has to be made on the whole history of X_{t+1} (i.e., the CCF of X_{t+1} is $C_{X_1, \dots, X_t}(r; \theta)$). Moreover, these procedures apply more generally to multivariate processes, including those involved with state variables. As long as all the state variables are observed, the ECCF method is used in the same way. However, when a state variable is not observable, the CCF of X_{t+1} cannot be calculated and hence the ECCF methods are not directly applicable unless the latent state variable can be integrated out from the joint CCF of observables and unobservables. Such an example is the stochastic volatility model where $X_t = (Y_t, \sigma_t^2)'$. In this case the return process Y_t , but not the volatility process σ_t^2 is observed. As a result, Singleton (2001) discusses how the ECCF method can be used in combination with simulations whereas Chacko and Viceira (2003) explain how to integrate out volatility from the joint CCF of (Y_t, σ_t^2) .

3. IMPORTANT ASSUMPTIONS AND ASYMPTOTIC PROPERTIES

3.1. Important Assumptions

In this subsection, I discuss assumptions under which asymptotic properties of the ECF estimators are derived. As non-iid processes include iid processes as special cases, I restrict my attention to the non-iid case in this section. It was shown that the discrete ECF procedure is the special case of the GMM procedure of Hansen (1982) while the continuous ECF procedure is the special case of the GMM procedure of Carrasco and Florens (2000). Not surprisingly, the assumptions adopted are closely related to those used in Hansen (1982) and Carrasco and Florens (2000). In particular, to develop the asymptotic properties of the ECCF estimator, Singleton (2001) makes use of the same set of assumptions as in Hansen (1982). Observing that the continuous ECF estimator is a class of extremum estimators, Knight and Yu (2002) impose a set of assumptions to ensure sufficient conditions of an extremum estimator listed by Newey and McFadden (1994, p. 2121). There are three common assumptions in Hansen (1982) and Knight and Yu (2002) which are important to applied researchers. We review them in detail.

Stationarity. X_t is assumed to be strictly stationary. A time series X_t is said to be strictly stationary if the joint distribution of $\{X_t, \dots, X_{t+\tau}\}$ is identical to that of $\{X_{t+s}, \dots, X_{t+s+\tau}\}$ for any t, s, τ . This rules out unit root processes, deterministic trend models, and unconditional heteroskedasticity. The stationarity assumption, however, does not rule out the possibility of conditional heteroskedasticity.

Weak Dependence. X_t is assumed to follow a form of weak dependence. The weak dependence holds true under suitable mixing (for example, α -mixing and β -mixing) conditions. For processes which do not have this form of weak dependence, such as long memory processes, we do not know if the ECF method is applicable.

Identification. This puts restrictions on the model and is related to the GMM identification restriction that $f(X_j; \theta)' W_0 f(X_j; \theta)$ has the unique minimizer at θ_0 .



In the ECF context, this is equivalent to requiring θ_0 to be the unique minimizer of (2.6) or the estimation equation (such as (2.12) or (2.8)) to have a unique solution at θ_0 . In the context of the joint CF, the restriction necessitates a careful choice of p . For example, for an MA(10) process, $X_t = \varepsilon_t - \phi\varepsilon_{t-10}, \varepsilon_t \sim \text{iid}N(0, \sigma^2)$, the moving blocks with $p < 10$ (say $p = 0$) cannot identify all the parameters. To see this we have

$$E(\exp(ir X_j) - c(r, \theta)) = \exp\left(-\frac{r^2 \sigma_0^2 (1 + \phi_0^2)}{2}\right) - \exp\left(-\frac{r^2 \sigma^2 (1 + \phi^2)}{2}\right).$$

Hence only $\sigma^2(1 + \phi^2)$ is identified.

3.2. Asymptotic Properties

The asymptotic properties of the ECF estimator in the iid case is established in Heathcote (1977). Since the ECCF estimator proposed by Singleton (2001) is treated as a GMM estimator, the asymptotic properties are the same as those of the GMM estimator. The asymptotic properties of the estimator based on the joint CF are established in Knight and Yu (2002). In all cases, the resulting ECF estimator is strongly consistent and asymptotically normally distributed. More interestingly, the convergence rate for the ECF estimator is \sqrt{n} . This is true for the processes with stable noise. This result is remarkably different from the ML estimator, the least absolute deviation (LAD) estimator, and the least square (LS) estimator where the rates of convergence are $n^{1/\alpha}$ for LAD/ML and $(n/\log n)^{1/\alpha}$ for LS with α being the index parameter in the stable distribution (see, for example, Calder and Davis, 1998 and references therein).

4. EXAMPLES

There are many models used in economics and finance for which the likelihood function is not bounded over the parameter space or has no closed form expression, while the CF and CCF have a closed form solution and hence the estimation method via the ECF or ECCF can be used. In this section we list some of these models which are used widely in practice.

4.1. Mixtures of Normal Distributions

Titterton et al. (1985, Chapter 2) list many applications of mixtures of normal distributions. Mixtures of k normal distributions are defined by a random variable X such that

$$X \sim N(\mu_i, \sigma_i^2) \quad \text{with probability } \lambda_i, \quad i = 1, \dots, k,$$

where $\sum_{i=1}^k \lambda_i = 1$, and $(\lambda_i, \mu_i, \sigma_i^2)_{i=1, \dots, k}$ are $(3k - 1)$ unknown parameters. The likelihood is unbounded when one of the above distributions is imputed to have a mean exactly equal to one of the observations with the corresponding variance going to zero. Consequently, a global maximum fails to exist.



The CF of X is

$$c(r) = \lambda_1 \exp\left(i\mu_1 r - \frac{1}{2}\sigma_1^2 r^2\right) + \dots + \lambda_k \exp\left(i\mu_k r - \frac{1}{2}\sigma_k^2 r^2\right).$$

Since the CF of the mixtures of normal distributions has a closed form expression and is uniformly bounded, an estimation method suggested in the literature is via the ECF. References include Bryant and Paulson (1983) and Tran (1998) and the references contained.

4.2. Switching/Dis-equilibrium Models

Mixtures of normal distributions often appear in economics, in switching regressions introduced by Quandt (1958), in dis-equilibrium models introduced by Goldfeld and Quandt (1973), and in regime switching models introduced by Hamilton (1989, 1990). Not surprisingly, estimation of these models has the same problems as for the mixtures of normal distributions and hence the ECF method is a viable estimation method.

4.3. Variance Gamma Distribution

The variance gamma (VG) distribution is proposed by Madan and Seneta (1990) to model share market returns. The VG distribution assumes that the conditional variance is distributed as a gamma variate. Formally, $X|V \sim N(0, V\sigma^2)$ and $V \sim \Gamma(c, \gamma)$, where Γ is the Gamma distribution. The density is given by

$$f(x) = \int_0^\infty \frac{\exp(-x^2/(2v\sigma^2))c^\gamma v^{\gamma-1} \exp(-cv)}{\sigma\sqrt{2\pi v}\Gamma(\gamma)} dv.$$

To calculate this density, one has to evaluate the above integral numerically. On the other hand, the CF of the VG distribution is given by

$$c(r) = [1 + \sigma^2 \gamma r^2 / (2c^2)]^{-c^2/\gamma},$$

and hence is very easy to calculate.

4.4. Stable Distribution

The stable distribution was first proposed by Mandelbrot (1963) and Fama (1965) to model stock returns. It is usually characterized by the CF given by

$$c(r) = \begin{cases} \exp\{i\mu r - \sigma|r|^\alpha [1 - i\beta \text{sign}(r) \tan(\pi\alpha/2)]\} & \text{if } \alpha \neq 1 \\ \exp\{i\mu r - \sigma|r|^\alpha [1 + i\beta(2/\pi)\text{sign}(r) \log(|r|)]\} & \text{if } \alpha = 1 \end{cases} \quad (4.17)$$



The stable distribution with the above CF is said to follow $S_\alpha(\sigma, \beta, \mu)$, where α , β , σ , and μ are, respectively, index, skewness, scale, and location parameters. Analytic forms for the density in terms of elementary functions are known for three cases, $\alpha = 1/2, 1, 2$. For any other value of α , the density function has to be calculated numerically by Fourier inverting (4.17).

A widely used method for estimating a stable distribution is based on the ECF. References include Press (1972), Paulson et al. (1975), Arad (1980), Koutrouvelis (1980, 1981), Feuerverger and McDunnough (1981a), Brockwell and Brown (1981), Paulson and Delehanty (1985) and Kogon and Williams (1998).

4.5. Stable ARMA Process

The stable ARMA model has been used to model financial time series, including returns in stock markets, commodity markets and foreign exchange markets; see Mittnik et al. (1998). The ARMA(l, m) model is of the form

$$Y_t = c + \rho_1 Y_{t-1} + \dots + \rho_l Y_{t-1-l} + \varepsilon_t - \phi_1 \varepsilon_{t-1} - \dots - \phi_m \varepsilon_{t-1-m},$$

where $\varepsilon_t \sim \text{iid } S_\alpha(\sigma, \beta, \mu)$.

Although the ML estimation for the stable ARMA model is notoriously difficult, since the CF of the error term has a closed form expression, it can be shown that the joint CF of the stable ARMA model has a closed form expression. For example, a two dimensional joint CF of the stable ARMA(1, 1) model, $Y_t = \rho Y_{t-1} + \varepsilon_t - \phi \varepsilon_{t-1}$ where $\varepsilon_t \sim S_\alpha(1, \beta, 0)$, is given by

$$\begin{aligned}
 c(r_1, r_2; \theta) = & \exp \left\{ -|r_2|^\alpha - |r_1 + (\rho - \phi)r_2|^\alpha - \frac{|r_1 + \rho r_2|^\alpha |\rho - \phi|^\alpha}{1 - |\rho|^\alpha} \right\} \\
 & \times \exp \left\{ i\beta \tan \frac{\pi\alpha}{2} \left[|r_2|^\alpha \text{sign}(r_2) + |r_1 + (\rho - \phi)r_2|^\alpha \text{sign}(r_1 + (\rho - \phi)r_2) \right. \right. \\
 & \left. \left. + \frac{|r_1 + \rho r_2|^\alpha \text{sign}(r_1 + \rho r_2) |\rho - \phi|^\alpha \text{sign}(\rho - \phi)}{1 - \text{sign}(\rho) |\rho|^\alpha} \right] \right\}.
 \end{aligned}$$

Based on the above expression of the joint CF, Knight and Yu (2002) estimate various stable ARMA models.

4.6. Discrete Time Stochastic Volatility Model

The discrete time stochastic volatility (SV) model has been used to model stock returns, interest rates, exchange rates; see Ghysels et al. (1996) and references therein. The basic SV model is of the form,

$$\begin{aligned}
 X_t = \sigma_t e_t = & \exp(0.5h_t) e_t, \quad e_t \sim \text{iid}N(0, 1), \quad t = 1, 2, \dots, T, \\
 h_t = & \lambda + \alpha h_{t-1} + v_t, \quad v_t \sim \text{iid}N(0, \sigma^2),
 \end{aligned}$$

where $\text{cov}(e_t, v_{t+1}) = 0$.



Since X_t is a non-linear function of the latent AR(1) process, h_t , it is difficult to work with. Defining Y_t to be the logarithm of X_t^2 , we have

$$Y_t = \log \sigma_t^2 + \log e_t^2 = h_t + \epsilon_t, \quad t = 1, 2, \dots, T,$$

where ϵ_t is the logarithm of the χ_1^2 random variable. Hence, Y_t depends on the latent process h_t in a linear form.

It is known that the SV model offers a powerful alternative to more widely used ARCH-type models (Kim et al., 1998). Unfortunately, neither Y_t nor X_t has a closed form expression for the likelihood function. This property makes the likelihood-based estimation extremely difficult to implement since it requires that the latent process be integrated out of the joint density for the observed and latent processes. As the convolution of an AR(1) process and an iid $\log \chi_1^2$ sequence, Y_t has a closed form expression of the joint CF. Hence the ECF method is a viable alternative. The joint CF of $Y_t, Y_{t+1}, \dots, Y_{t+k-1}$ is first derived in Yu (1998) and given by

$$c(r_1, \dots, r_k; \theta) = \exp \left[\frac{i\lambda}{1-\alpha} \sum_{j=1}^k r_j - \frac{\sigma^2}{2(1-\alpha^2)} \left(\sum_{j=1}^k r_j^2 + 2\alpha \sum_{l=1}^k \sum_{j=l+1}^k \alpha^{j-l-1} r_l r_j \right) \right] \\ \times \prod_{j=1}^k \frac{\Gamma(\frac{1}{2} + i r_j)}{\Gamma^k(\frac{1}{2})} 2^i \sum_{j=1}^k r_j,$$

where $\Gamma(\cdot)$ is the Gamma function.

A special case of the above SV model is the subordinated stochastic process proposed in Clark (1973) where α is set to be 0. Furthermore, it can be shown that more general SV models can also have a closed form joint CF and hence the ECF procedures are applicable. For example, for the SV model with the leverage effect defined by

$$Y_t = \log \sigma_t^2 + \log e_t^2, \quad e_t \sim \text{iid}N(0, 1), \quad t = 1, \dots, T \\ \log \sigma_t^2 = \lambda + \alpha \log \sigma_{t-1}^2 + v_t, \quad v_t \sim \text{iid}N(0, \sigma^2)$$

where $\text{cov}(e_t, v_{t+1}) = \rho\sigma^2$, the joint CF of $Y_t, Y_{t+1}, \dots, Y_{t+k-1}$ is given by Knight et al. (2002)

$$c(r_1, \dots, r_k; \theta) = \exp \left[\frac{i\lambda}{1-\alpha} \sum_{j=1}^k r_j \right] \exp \left[-\frac{\sigma^2}{2(1-\alpha^2)} \left(\sum_{j=1}^k r_j \alpha^{k-j} \right)^2 \right] \\ \times \exp \left[-\frac{\sigma^2(1-\rho^2)}{2} \sum_{l=2}^k \left(\sum_{j=l}^k \alpha^{j-l} r_{k+1-j} \right)^2 \right] \frac{\prod_{j=1}^k \Gamma(\frac{1}{2} + i r_j)}{\Gamma^k(\frac{1}{2})} 2^i \sum_{j=1}^k r_j \\ \times \prod_{j=2}^k {}_1F_1 \left(r_j + \frac{1}{2}, \frac{1}{2}, -\frac{\sigma^2 \rho^2}{2} \left(\sum_{l=1}^{j-1} r_j \alpha^{j-1-l} \right)^2 \right),$$

where ${}_1F_1$ is the hypergeometric function.



4.7. Affine Jump Diffusion Models

Affine jump diffusion models have been used extensively to describe the dynamics of asset prices in finance. A general affine model is taken from Duffie et al. (2000) and is of the form

$$dY_t = \mu(Y_t)dt + \sigma(Y_t)dB_t + dZ_t,$$

where B_t is a standard Brownian motion, and Z_t is a pure jump process with intensity $\{\lambda(Y_t)\}$ and jump size ν . The process is affine if

$$\begin{aligned} \mu(y) &= K_0 + K_1y, \\ \sigma(y)\sigma'(y) &= (H_0)_{ij} + (H_1)_{ij}y, \\ \lambda(y) &= l_0 + l_1y. \end{aligned}$$

Duffie et al. (2000) derive the CCF of Y_{t+1} conditional on Y_t which is given by

$$c_{Y_t}(r; Y_{t+1}) = E[\exp(irY_{t+1})|Y_t] = \exp(C(1) + D(1)'Y_t),$$

where $D(\cdot)$ and $C(\cdot)$ satisfy the following complex-valued Ricatti equations:

$$\begin{aligned} \frac{\partial D(\tau)}{\partial \tau} &= K_1'D(\tau) + \frac{1}{2}D(\tau)'H_1D(\tau) + l_1(g(D(\tau)) - 1), \\ \frac{\partial C(\tau)}{\partial \tau} &= K_0'D(\tau) + \frac{1}{2}D(\tau)'H_0D(\tau) + l_0(g(D(\tau)) - 1), \end{aligned}$$

with boundary conditions: $D(0) = ir, C(0) = 0, g(\cdot)$ being the moment generating function of ν . With certain specifications of the coefficient functions (K, H, L) and g , explicit solutions of $D(\cdot)$ and $C(\cdot)$ can be found. Two papers which have estimated the above model via the ECCF are Singleton (2001) and Chacko and Viceira (2003).

Jiang and Knight (2002) derive the joint CF of a particular class of affine jump diffusion models, where some of the state variables are unobserved. It includes as a special case the following continuous time square-root SV model

$$\begin{aligned} dY_t &= \mu dt + h_t^{1/2} dB_{1t}, \\ dh_t &= \beta(\alpha - h_t)dt + \sigma h_t^{1/2} dB_{2t}, \\ dB_{1t}dB_{2t} &= \rho dt. \end{aligned} \tag{4.18}$$

Suppose Y_t is observed at equi-spaced intervals on the model defined by (4.18) with initial condition $Y_0 = y_0$. The joint CF of (Y_1, \dots, Y_{p+1}) is given by,

$$c(r_1, \dots, r_{p+1}; Y_1, \dots, Y_{p+1}|y_0) = \exp\left(\sum_{k=1}^{p+1} C(1, r_k^*)\right) \exp\left(D(1, r_1^*)'y_0\right), \tag{4.19}$$



where $r_{p+1}^* = r_{p+1}$ and $r_k^* = r_k - iD(1, r_{k+1}^*)$, with $k = 1, \dots, p$. If the CF of Y_0 is $\phi(r; y_0)$, then the unconditional joint CF of (Y_1, \dots, Y_{p+1}) is given by,

$$c(r_1, \dots, r_{p+1}; Y_1, \dots, Y_{p+1}) = \exp\left(\sum_{k=1}^{p+1} C(1, r_k^*)\right) \phi(D(1, r_1^*); y_0).$$

Based on the joint CF, Jiang and Knight (2002) estimate the continuous time SV model by applying the continuous ECF procedure proposed in Yu (1998) and Knight and Yu (2002).

5. ECF ESTIMATION FOR A SELF-EXCITING JUMP DIFFUSION MODEL

To illustrate the ECF procedure I now consider a self-exciting jump diffusion model which is first proposed by Knight and Satchell (1998). A small scale Monte Carlo study and an empirical study are performed.

5.1. 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)

$$dP(t) = \gamma P(t)dt + \sigma P(t)dB(t),$$

where $B(t)$ is a standard BM, γ is the instantaneous return and σ^2 is the instantaneous variance. By including the jump component, Merton (1976) assumes that the price follows the mixed Brownian–Poisson process

$$dP(t) = \gamma P(t)dt + \sigma P(t)dB(t) + P(t)(\exp(Q) - 1)dN(t), \tag{5.20}$$

where $B(t)$ is a standard Brownian motion; $N(t)$ is Poisson process with intensity parameter λ ; $B(t)$ and $N(t)$ are assumed to be independent; Q is an independent normal variable with mean μ_Q and variance σ_Q^2 . Using Ito's lemma, we solve the stochastic differential Eq. (5.20) for the log return $X(t)(= \log(P(t)/P(t - 1)))$,

$$\begin{aligned} X(t) &= \left(\gamma - \frac{\sigma^2}{2}\right) + \sigma(B(t) - B(t - 1)) + \sum_{n=1}^{\Delta N(t)} Q(n) \\ &= \mu + \sigma(B(t) - B(t - 1)) + \sum_{n=1}^{\Delta N(t)} Q(n), \end{aligned} \tag{5.21}$$

where $Q(n) = \sum_{i=1}^n Q_i$ if $n \geq 1$, and $\mu = \gamma - \sigma^2/2$. Hence, the behavior of $X(t)$ depends not only on the continuous diffusion part $\mu + \sigma(B(t) - B(t - 1))$, but also a dis-continuous jump part $\sum_{n=1}^{\Delta N(t)} Q(n)$.

Knight and Satchell (1998) extend the Merton model by assuming that the Poisson process $N(t)$ has a stochastic intensity function $\lambda(t)$ which is self-exciting



as follows:

$$\lambda(t) = \beta\sigma^2 + \phi\lambda(t-1) + \alpha\nu^2(t-1), \tag{5.22}$$

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 . In Appendix A, we show that Eq. (5.22) is equivalent to

$$\lambda(t) = \beta\text{Var}(X(t-1)|I(t-2)) + \alpha\nu^2(t-1). \tag{5.23}$$

and

$$\text{Var}(X(t)|I(t-1)) = \sigma^2 + \phi\text{Var}(X(t-1)|I(t-2)) + \alpha(\mu_Q^2 + \sigma_Q^2)\nu^2(t-1), \tag{5.24}$$

where $\phi = \beta(\mu_Q^2 + \sigma_Q^2)$.

According to the specification, the jump component's arrival time is endogenously determined, reflecting past volatility in the data and deviations from economic fundamentals. It offers a more general specification than models with a constant jump intensity (Jorion, 1988; Merton, 1976). It can be regarded as an alternative way to model the time varying jump intensity via exogenous variables (Bekaert and Gray, 1998; Das, 1999). A similar jump specification is incorporated in models proposed by Chan and Maheu (2002) and Maheu and McCurdy (2004) where the jump intensive is assumed to follow a Gaussian autoregressive structure.^c As noted in Maheu and McCurdy (2004), such models allow jumps to cluster. This feature is termed "jump clustering" which is analogous to "volatility clustering" in the GARCH and SV literature. In Eq. (5.22) jump clustering is captured by parameter ϕ .

Obviously, the model defined by (5.21) and (5.22) can be regarded as the discrete version of bivariate diffusion process for $(X_t, \lambda_t)'$ with one of the state variables, λ_t , latent. Unfortunately, the likelihood function has no closed form and hence the ML method is infeasible. We show in the theorem below, however, that the joint CF has a closed form expression and thereby facilitates the use of the GMM and ECF procedures.

Theorem 5.1. *If a random process $\{X(t)\}_{t=1}^T$ is a self-exciting Poisson jump diffusion model which is defined by Eqs. (5.21) and (5.23), then the joint CF of $X(t), \dots, X(t-k)$ is,*

$$\begin{aligned} c(r_1, \dots, r_{k+1}; \theta) &= \exp\left(i\mu \sum_{j=1}^{k+1} r_j - \frac{1}{2}\sigma^2 \sum_{j=1}^{k+1} r_j^2\right) \exp\left\{\frac{\beta\sigma^2}{1-\phi} \sum_{j=1}^{k+1} G(r_j)\right\} \\ &\times \prod_{l=0}^{\infty} \left\{1 - 2\alpha\phi^l \sum_{j=1}^{k+1} \phi^{k+1-j} G(r_j)\right\}^{-1/2} \\ &\times \prod_{j=2}^{k+1} \left\{1 - 2\alpha \sum_{l=1}^j \phi^{j-l} G(r_l)\right\}^{-1/2}, \end{aligned} \tag{5.25}$$

^cMaheu and McCurdy (2004, Sec. 4) give several interesting examples in which one should expect that jumps tend to be followed by jumps.



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

Proof. See Appendix A. □

It is well known that the joint CF can be used, as well, to derive closed form expressions for marginal and joint moments of the model, by evaluating derivatives of the joint CF at zero. In particular, the unconditional moments can be derived from the marginal CF while the joint moments can be derived from the joint CF. For example, $\text{cov}(X(t), X(t-s)) = 2\mu_Q^2\alpha^2\phi^s/(1-\phi^2)$. Hence, $X(t)$ and $X(t-s)$ are uncorrelated when $\mu_Q = 0$. However, $X(t)$ and $X(t-1)$ are not independent since $c(r_1, r_2; \theta) \neq c(r_1; \theta)c(r_2; \theta)$. Appendix B presents the analytic expressions for some moments of the model, including four unconditional moments and three autocovariances. These moment conditions form the basis of the GMM procedure.

As for the ECF method, we use the continuous ECF procedure based on the joint CF. The ECCF procedure is not used here since the state variables λ_t is not observable. To use the ECF method, we have to choose a value for p . It is easy to see that the model does not have a Markov property, and hence a larger p works better than a smaller p in terms of asymptotic efficiency. In the Monte Carlo study we only choose $p = 1$ to ease the computational burden while in the empirical study we choose several values for p to examine the effect of p on the estimates. Note that with $p = 1$

$$c(r_1, r_2; \theta) = \exp\left\{\frac{\beta\sigma^2}{1-\phi}(G(r_1) + G(r_2)) + i\mu(r_1 + r_2) - \frac{1}{2}\sigma^2(r_1^2 + r_2^2)\right\} \times \prod_{l=0}^{\infty} \{1 - 2\alpha\phi^l(\phi G(r_1) + G(r_2))\}^{-1/2} (1 - 2\alpha G(r_1))^{-\frac{1}{2}}, \quad (5.26)$$

and

$$c_n(r_1, r_2) = \frac{1}{n} \sum_{j=1}^n \exp(ir_1x_j + ir_2x_{j+1}). \quad (5.27)$$

Although the optimal ECF procedure proposed in a recent study by Carrasco et al. (2002) should lead to an estimator with ML efficiency, in this paper I use the sub-optimal WLS-ECF method with an exponential weighting function. As a result, the procedure is to choose $(\hat{\mu}, \hat{\sigma}^2, \hat{\alpha}, \hat{\beta}, \hat{\mu}_Q, \hat{\sigma}_Q^2)$ to minimize

$$\iint |c(r_1, r_2; \theta) - c_n(r_1, r_2)|^2 \exp(-r_1^2 - r_2^2) dr_1 dr_2, \quad (5.28)$$

where $c(r_1, r_2; \theta)$ and $c_n(r_1, r_2)$ are given by (5.26) and (5.27).

The implementation of the ECF method essentially requires minimizing (5.28), 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 (5.28), followed by numerical



minimization of (5.28) with respect to θ . The numerical solutions are the desired estimators.

A 96-point Hermite quadrature is used to approximate the two dimensional integral in (5.28). 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 procedure, we examine the performance of the ECF method in the estimation of the self-exciting jump diffusion model in a Monte Carlo study. We also fit the model to a real data set.

5.2. Monte Carlo Experiments

The Monte Carlo study is designed to check the viability of the ECF method in comparison with a GMM procedure. For simplicity, in the model defined by Eqs. (5.21) and (5.23), we let $\mu = \mu_Q = \sigma = 0, \alpha = 1$ and assume them to be known quantities. Therefore, the model can be represented by

$$X(t) = \sum_{n=1}^{\Delta N(t)} Q(n), \tag{5.29}$$

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

$$\lambda(t) = \phi\lambda(t - 1) + \nu^2(t - 1). \tag{5.30}$$

I choose parameters $\sigma_Q^2 = 1$ and $\beta = 0.5$, which imply $\phi = 0.5$. The number of observations is set at $T = 2000$ and the number of replications is 1000.

I propose two estimators. One is the GMM estimator based on ad hoc moment conditions, the other is the ECF estimator. The details of the GMM procedure are given by Hansen (1982) in a more general framework. In this specific situation, I 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). For the ECF, I used the GMM estimates as the starting point.

Table 1 shows the mean, median, minimum, maximum, MSE and root mean square error (RMSE) for both sets of estimates, and serves to illustrate that the ECF method outperforms the GMM procedure. For example, the ECF estimates have smaller biases than the GMM estimates. Moreover, the MSE’s of the GMM estimates are larger than those of the ECF estimates, suggesting the ECF method is more efficient than the GMM procedure. Of course, the GMM procedure adopted here may not be ideal since it is based on a set of ad hoc moments conditions. If one keeps adding moments and adopts the optimal scheme of Hansen (1982), asymptotic efficiency should improve (Liu, 1997). If the increased moment conditions are based on power functions of a random variable, the ML asymptotic efficiency can be



Table 1. Monte Carlo study to compare ECF with GMM for the self-exciting jump diffusion model.

	$\beta = 0.5$		$\sigma_Q^2 = 1.0$	
	ECF	GMM	ECF	GMM
Mean	0.5087	0.5132	0.9803	0.9630
Med	0.5067	0.5176	0.9757	0.9675
Min	0.3395	0.0965	0.6331	0.4116
Max	0.6492	0.847	1.340	1.643
Mse	0.00287	0.00485	0.0195	0.0394
RMSE	0.0536	0.0696	0.1396	0.1980

Note: The reported statistics are based on 1000 simulated replications each with sample size equal to 2000.

achieved (Gallant and Tauchen, 1999). However, the ECF procedure used here is also sub-optimal. In the context of the scalar diffusion, Carrasco et al. (2002) find that their optimal ECF procedure outperforms the efficient method of moment of Gallant and Tauchen (1996) where a judiciously chosen set of moment conditions is used.

5.3. An Empirical Application

The data used in the empirical study consists of more than 72 years (19, 302 observations) of daily geometric returns (defined as $100(\log P_{t+1} - \log P_t)$) for the Dow Jones Industrial Average (DJIA) index covering the period from October 2, 1928 to June 12, 2001. The series is plotted in Fig. 1.

In this empirical study, we fit the model defined by (5.21) and (5.23) to the data using the continuous ECF procedure with $p = 1, 2, 3$. The parameters of interest are $\theta = (\mu, \sigma^2, \alpha, \beta, \mu_Q, \sigma_Q^2)'$. The parameter estimates and corresponding asymptotic standard errors are presented in Table 2.

A few results emerge from the table. First and probably most importantly, the autoregressive coefficient ϕ in Eq. (5.22) is estimated to be 0.9512, 0.9340, 0.9426, for $p = 1, 2, 3$. Using the delta method, we can obtain the asymptotic standard error of $\hat{\phi}$, which is 0.2041, 0.1765, 0.1822 for $p = 1, 2, 3$. Obviously ϕ is statistically significant in all the three cases. The result suggests overwhelming evidence of jump clustering, consistent with the empirical findings recently documented in Chan and Maheu (2002) and Maheu and McCurdy (2004). Second, the ECF estimates of μ_Q, σ_Q^2 are highly significant and indicate that jumps are present in the sample path of DJIA returns over the period from 1928 to 2001. This result is consistent with the findings by Ball and Torous (1985) and Chib et al. (2002) in the sample paths of daily stock returns. Moreover, the estimate of μ_Q is significantly less than 0, suggesting that on average jumps lead to a negative return in the stock market. Third, the estimates clearly depend on but are not highly sensitive to the overlapping block size p . It is apparent that the parameters



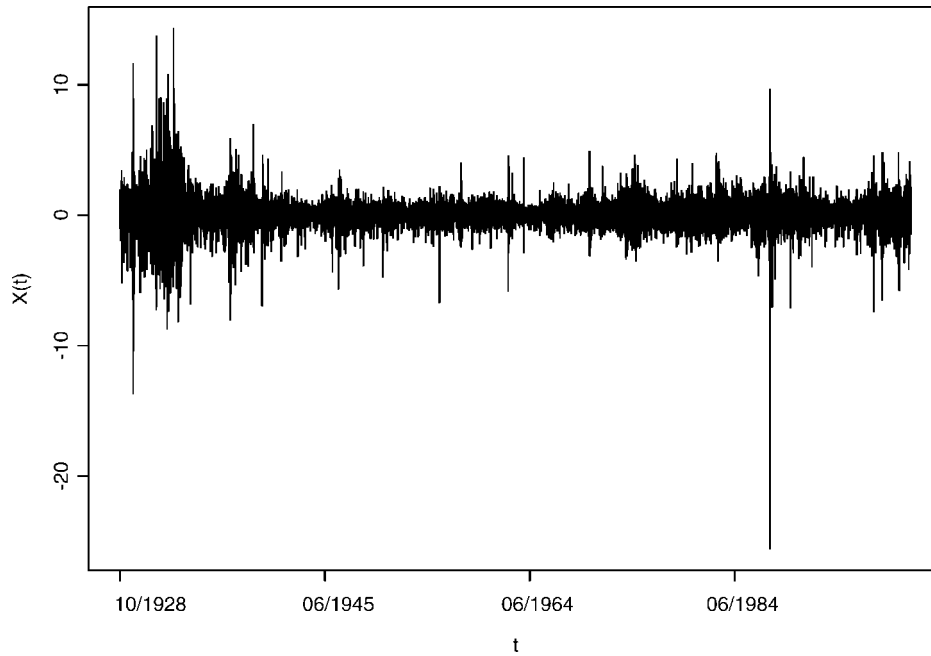


Figure 1. Daily observations of the Dow Jones industrial average index returns from October 2, 1928 to June 12, 2001.

associated with the jump component are relatively less stable than those in the diffusion component.

6. CONCLUSION

In this paper, the estimation method via the ECF has been described and illustrated. The estimation method requires no tractable form or property of the likelihood function. It is accomplished by matching the CF with ECF. We illustrate the ECF procedure by estimating a self-exciting jump diffusion model.

Table 2. Estimation of the self-exciting jump diffusion model via ECF using DJIA data.

Method	μ	σ^2	α	β	μ_Q	σ_Q^2
ECF ($p = 1$)	0.0712 (0.0091)	0.1687 (0.0009)	0.3065 (0.0315)	0.6053 (0.0954)	-0.4341 (0.0814)	1.3830 (0.1789)
ECF ($p = 2$)	0.0710 (0.0090)	0.1685 (0.0009)	0.3117 (0.0303)	0.6142 (0.0842)	-0.4122 (0.0758)	1.3507 (0.1655)
ECF ($p = 3$)	0.0711 (0.0089)	0.1686 (0.0009)	0.2986 (0.0290)	0.5979 (0.0801)	-0.4278 (0.0736)	1.3994 (0.1549)

Note: The numbers in brackets are standard errors.



Simulations demonstrate that the ECF method works well in comparison with an ad hoc GMM approach. An empirical application to DJIA index returns shows some interesting results, including evidence of jump clustering.

APPENDIX A

Proof of Theorem 5.1.

We start the proof by showing that Eqs. (5.22)–(5.24) are equivalent. From the definition of $I(t - 1)$, we get

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

Substituting out $\lambda(t)$ in Eq. (5.23), we then get

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

Furthermore, by substituting out $\text{Var}(X(t - 1)|I(t - 2))$ in Eq. (5.23), we get

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

which is Eq. (5.22). Since the intensity represents how fast new information arrives, Eq. (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 Eq. (A.3), we get

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

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

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

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

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



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(i r_l \sum_{n=1}^{\Delta N(t-l)} Q(n) \right) \right\} \\ &= E \left\{ E \left\{ E \left[\prod_{l=1}^{k+1} \exp \left(i r_l \sum_{n=1}^{\Delta N(t-l)} Q(n) \right) \middle| \Delta N(t-l), I(t-k) \right] \right\} \right\} \\ &= E \left\{ E \left[\prod_{l=1}^{k+1} \exp \left(i \mu_Q r_l \Delta N(t-l) - \frac{r_l^2}{2} \sigma_Q^2 \Delta N(t-l) \right) \middle| I(t-k) \right] \right\} \\ &= E \left\{ \prod_{l=1}^{k+1} \exp \left[\lambda(t-l) \left(\exp \left(i \mu_Q r_l - \frac{r_l^2}{2} \sigma^2 \right) - 1 \right) \right] \right\} \\ &= E \left\{ \prod_{l=1}^{k+1} \exp \left[\left(\frac{\beta \sigma^2}{1-\phi} (1 - \phi^{k-l}) + \phi^{k-l} \lambda(t-k) \right. \right. \right. \\ &\quad \left. \left. \left. + \alpha (\nu^2(t-j-1) + \dots + \phi^{k-l-1} \nu^2(t-k)) \right) \left(\exp \left(i \mu_Q r_l - \frac{r_l^2}{2} \sigma^2 \right) - 1 \right) \right] \right\} \\ &= \exp \left[\sum_{l=1}^{k+1} \frac{\beta \sigma^2}{1-\phi} (1 - \phi^{k-l}) \left(\exp \left(i \mu_Q r_l - \frac{r_l^2}{2} \sigma^2 \right) - 1 \right) \right] \\ &\quad \times E \left\{ \exp \left[\lambda(t-k) \sum_{l=1}^{k+1} \left(\phi^{k+1-l} \left(\exp \left(i \mu_Q r_l - \frac{r_l^2}{2} \sigma^2 \right) - 1 \right) \right) \right] \right\} \\ &\quad \times E \left\{ \prod_{l=1}^{k+1} \exp \left[\alpha (\nu^2(t-j-1) + \dots + \phi^{k-l-1} \nu^2(t-k)) \right. \right. \\ &\quad \left. \left. \times \left(\exp \left(i \mu_Q r_l - \frac{r_l^2}{2} \sigma^2 \right) - 1 \right) \right] \right\} \\ &= \exp \left[\sum_{l=1}^{k+1} \frac{\beta \sigma^2}{1-\phi} (1 - \phi^{k-l}) \left(\exp \left(i \mu_Q r_l - \frac{r_l^2}{2} \sigma^2 \right) - 1 \right) \right] \\ &\quad \times \exp \left\{ \frac{\beta \sigma^2}{1-\phi} \sum_{l=1}^{k+1} \left[\phi^{k+1-l} \left(\exp \left(i \mu_Q r_l - \frac{r_l^2}{2} \sigma^2 \right) - 1 \right) \right] \right\} \\ &\quad \times \prod_{j=0}^{\infty} \left\{ 1 - 2\alpha \phi^j \sum_{l=1}^{k+1} \left[\phi^{k+1-l} \left(\exp \left(i \mu_Q r_l - \frac{r_l^2}{2} \sigma^2 \right) - 1 \right) \right] \right\} \\ &\quad \times E \left\{ \exp \left[\sum_{j=2}^{k+1} \left(\alpha \sum_{l=1}^j \phi^{j-l} \left(\exp \left(i \mu_Q r_l - \frac{r_l^2}{2} \sigma^2 \right) - 1 \right) \right) \right] \right\} \end{aligned}$$



which equals

$$\begin{aligned}
 & \exp \left[\sum_{l=1}^{k+1} \frac{\beta \sigma^2}{1-\phi} (1-\phi^{k-l}) G(r_l) \right] \exp \left\{ \frac{\beta \sigma^2}{1-\phi} \sum_{l=1}^{k+1} [\phi^{k+1-l} G(r_l)] \right\} \\
 & \times \prod_{l=0}^{\infty} \left\{ 1 - 2\alpha_1 \phi^l \sum_{j=1}^{k+1} \phi^{k+1-j} G(r_j) \right\}^{-1/2} \prod_{j=2}^{k+1} \left\{ 1 - 2\alpha \sum_{l=1}^j \phi^{j-l} G(r_l) \right\}^{-1/2} \\
 & = \exp \left\{ \frac{\beta \sigma^2}{1-\phi} \sum_{j=1}^{k+1} G(r_j) \right\} \prod_{l=0}^{\infty} \left\{ 1 - 2\alpha \phi^l \sum_{j=1}^{k+1} \phi^{k+1-j} G(r_j) \right\}^{-1/2} \\
 & \times \prod_{j=2}^{k+1} \left\{ 1 - 2\alpha \sum_{l=1}^j \phi^{j-l} G(r_l) \right\}^{-1/2}.
 \end{aligned}$$

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

$$\begin{aligned}
 & c(r_1, \dots, r_{k+1}; \boldsymbol{\theta}) \\
 & = E\{\exp(ir_1 X(t) + \dots + ir_{k+1} X(t-k))\} \\
 & = E\left\{ \exp\left(ir_1 \left(\mu + \sigma B(1) + \sum_{n=1}^{\Delta N(t-1)} Q(n) \right) \right. \right. \\
 & \quad \left. \left. + \dots + ir_{k+1} \left(\mu + \sigma B(1) + \sum_{n=1}^{\Delta N(t-1)} Q(n) \right) \right) \right\} \\
 & = E\left\{ \exp\left(i\mu \sum_{l=1}^{k+1} r_l \right) \exp(i\sigma(r_1 B(1) + \dots + r_{k+1} B(1))) \exp\left(\sum_{l=1}^{k+1} ir_l \sum_{n=1}^{\Delta N(t-1)} Q(n) \right) \right\} \\
 & = \exp\left(i\mu \sum_{j=1}^{k+1} r_j - \frac{1}{2} \sigma^2 \sum_{j=1}^{k+1} r_j^2 \right) E\left\{ \prod_{l=1}^{k+1} \exp\left(ir_l \sum_{n=1}^{\Delta N(t-1)} Q(n) \right) \right\} \\
 & = \exp\left(i\mu \sum_{j=1}^{k+1} r_j - \frac{1}{2} \sigma^2 \sum_{j=1}^{k+1} r_j^2 \right) \exp\left\{ \frac{\beta \sigma^2}{1-\phi} \sum_{j=1}^{k+1} G(r_j) \right\} \\
 & \times \prod_{l=0}^{\infty} \left\{ 1 - 2\alpha \phi^l \sum_{j=1}^{k+1} \phi^{k+1-j} G(r_j) \right\}^{-1/2} \prod_{j=2}^{k+1} \left\{ 1 - 2\alpha \sum_{l=1}^j \phi^{j-l} G(r_l) \right\}^{-1/2}. \quad (\text{A.4})
 \end{aligned}$$

APPENDIX B

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

The model is defined by Eqs. (5.21) and (5.23). The joint cumulant generating function is obtained by applying the logarithmic operator to the joint characteristic



function (5.26). The coefficients on the Taylor series expansion of the joint cumulant generating function are the cumulants of the model (Kendall and Stuart, 1958, p. 83), that is

$$\log c(r_1, r_2; \boldsymbol{\theta}) = \sum_{j,k=1}^{\infty} \kappa_{jk} \frac{(ir_1)^j (ir_2)^k}{j! k!}.$$

Hence we have

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

The analytic expressions of corresponding moments can be then calculated using the relationship given by Kendall and Stuart (1958),

$$\begin{aligned} \mu_1 &= \kappa_1 \\ \mu_2 &= \kappa_2 \\ \mu_3 &= \kappa_3 \\ \mu_4 &= \kappa_4 + 3\mu_2^2 \\ \mu_{11} &= \kappa_{11} \\ \mu_{12} &= \kappa_{12} \\ \mu_{22} &= \kappa_{22} + \mu_2^2 + 2\mu_{11}^2, \end{aligned}$$

where $\mu_1 = \int_{-\infty}^{\infty} x dF(x)$, $\mu_j = \int_{-\infty}^{\infty} (x - \mu_1)^j dF(x)$, $\forall j > 1$, and $\mu_{jk} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (X_t - \mu_1)^j (X_{t+1} - \mu_1)^k dF(X_t, X_{t+1})$.



ACKNOWLEDGMENT

I would like to thank Essie Maasoumi (the editor) and four anonymous referees for comments that have substantially improved the paper. I also thank Viv Hall, John Knight, Peter Phillips, Hailong Qian, Alan Rogers, Julian Wright, and seminar participants at the Midwest Econometric Study Group Meeting in Chicago for their suggestions and comments. I am grateful to A.R. Bergstrom for his financial support via the Bergstrom Prize in Econometrics. An earlier version of this paper was circulated under the title “Estimation of a Self-Exciting Poisson Jump Diffusion Model by the Empirical Characteristic Function Method.” This research is supported by the Royal Society of New Zealand Marsden Fund under No. 01-UOA-015.

REFERENCES

- Andersen, T., Sorensen, B. (1996). GMM estimation of a stochastic volatility model: A Monte Carlo study. *J. Business Economic Statist.* 14:329–352.
- Arad, R. W. (1980). Parameter estimation for symmetric stable distribution. *Int. Economic Rev.* 21:209–220.
- Arato, M. (1961). On the sufficient statistics for Gaussian random processes. *Theory Probability Appl.* 4:199–201.
- Ball, C., Torous, W. (1985). On jumps in common stock prices and their impact on call option pricing. *J. Finance* 40:155–173.
- Bekaert, G., Gray, S. (1998). Target zones and exchange rates: An empirical investigation. *J. Int. Economics* 45:1–35.
- Brockwell, P. J., Brown, B. M. (1981). High-efficient estimation for the positive stable laws. *J. Am. Statist. Assoc.* 76:622–631.
- Bryant, J. L., Paulson, A. S. (1983). Estimation of mixing properties via distance between characteristic functions. *Commun. Statist. Theory Methods* 12: 1009–1029.
- Bühlmann, P. (1994). Blockwise bootstrapped empirical process for stationary sequences. *Ann. Statist.* 22:995–1012.
- Calder, M., Davis, R. A. (1998). Inferences for linear processes with stable noise. In: Adler, R. J., Feldman, R. E., Taqqu, M. S., eds. *A Practical Guide to Heavy Tails*. Birkhäuser, pp. 159–176.
- Carrasco, M., Chernov, M., Florens, J., Ghysels, E. (2002). Efficient estimation of jump diffusions and general dynamic models with a continuum of moment conditions. Working Paper; Department of Economics: University of Rochester.
- Carrasco, M., Florens, J. (2000). Generalization of GMM to a continuum of moment conditions. *Econometric Theory* 16:797–834.
- Carrasco, M., Florens, J. (2002). Efficient GMM estimation using the empirical characteristic function. Working Paper; Department of Economics: University of Rochester.
- Chacko, G., Viceira, L. M. (2003). Spectral GMM estimation of continuous-time processes. *J. Econometrics* 116:259–292.



- Chan, W., Maheu, J. (1994). Conditional jump dynamics in stock market returns. *J. Business Economic Statist.* 20:377–389.
- Chib, S., Nardari, F., Shephard, N. (2002). Markov chain Monte Carlo methods for generalized stochastic volatility models. *J. Econometrics* 108:281–316.
- Clark, P. K. (1973). A subordinated stochastic process model with finite variance for speculative price. *Econometrica* 41:135–155.
- Das, S. R. (1999). Poisson Gaussian processes and the bond market. National Bureau of Economic Research, Working Paper 6631.
- Duffie, D., Pan, J., Singleton, K. J. (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica* 68:1343–1376.
- Fama, E. (1965). The behavior of stock prices. *J. Business* 47:244–280.
- Feuerverger, A. (1990). An efficiency result for the empirical characteristic function in stationary time-series models. *Canadian J. Statist.* 18:155–161.
- Feuerverger, A., McDunnough, P. (1981a). On efficient inference in symmetric stable laws and processes. In: Csorgo, M., ed. *Statistics and Related Topics*. New York: North-Holland, pp. 109–122.
- Feuerverger, A., McDunnough, P. (1981b). On some Fourier methods for inference. *J. Am. Statist. Assoc.* 76:379–387.
- Feuerverger, A., McDunnough, P. (1981c). On the efficiency of empirical characteristic function procedures. *J. Roy. Statist. Soc. Series B, Methodological* 43:20–27.
- Feuerverger, A., Mureika, R. A. (1977). The empirical characteristic function and its applications. *Ann. Statist.* 5:88–97.
- Gallant, A. R., Tauchen, G. (1996). Which moments to match? *Econometric Theory* 12:657–681.
- Gallant, A. R., Tauchen, G. (1999). The relative efficiency of method of moments estimators. *J. Econometrics* 92:149–172.
- Goldfeld, S. M., Quandt, R. E. (1973). A Markov model for switching regression. *J. Econometrics* 1:3–15.
- Ghysels, E., Harvey, A. C., Renault, E. (1996). Stochastic volatility. In: Maddala, G. S., Rao, C. R., eds. *Handbook of Statistics, Vol 14, Statistical Methods in Finance*. Amsterdam: North-Holland, pp. 119–191.
- Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57:357–384.
- Hamilton, J. D. (1990). Analysis of time series subject to changes in regime. *J. Econometrics* 45:39–70.
- Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* 50:1029–1054.
- Hansen, L. P. (1985). A method for calculating bounds on the asymptotic covariance matrices of generalized method of moments estimators. *J. Econometrics* 30:203–238.
- Härdle, W., Linton, O. (1994). Applied nonparametric methods. In: Engle, R. F., McFadden, D., eds. *Handbook of Econometrics*. Vol. 4. Amsterdam: North-Holland, pp. 2295–2339.
- Heathcote, C. R. (1977). The integrated squared error estimation of parameters. *Biometrika* 64:255–264.
- Jiang, G., Knight, J. L. (2002). Estimation of continuous time processes via the empirical characteristic function. *J. Business Economic Statist.* 20:198–212.



- Jiang, G., Knight, J. L. (2003). Efficient estimation in Markov models where the transition density is unknown. Manuscript, Department of Economics: University of Western Ontario.
- Jorion, P. (1988). On jump processes in the foreign exchange and stock markets. *Rev. Financial Stud.* 1:427–445.
- Kendall, M., Stuart, A. (1958). *The Advanced Theory of Statistics*. London: Charles Griffin and Company.
- Kim, S., Shephard, N., Chib, S. (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *Rev. Economic Stud.* 65:361–393.
- Knight, J. L., Satchell, S. E. (1996). Estimation of stationary stochastic processes via the empirical characteristic function. Manuscript, Department of Economics, University of Western Ontario.
- Knight, J. L., Satchell, S. E. (1997). The cumulant generating function estimation method. *Econometric Theory* 13:170–184.
- Knight, J. L., Satchell, S. E. (1998). GARCH processes – some exact results, some difficulties and a suggested remedy. In: Knight, J. L., Satchell, S. E., eds. *Forecasting Volatility in the Financial Markets*. Oxford: Butterworth-Heinemann, pp. 321–346.
- Knight, J. L., Yu, J. (2002). The empirical characteristic function in time series estimation. *Econometric Theory* 18:691–721.
- Knight, J. L., Satchell, S. E., Yu, J. (2002). Estimation of the stochastic volatility model by the empirical characteristic function method. *Australian New Zealand J. Statist.* 44:319–335.
- Kogon, S. M., Williams, D. B. (1998). Characteristic function based estimation of stable distribution parameters. In: Adler, R. J., Feldman, R. E., Taquq, M. S., eds. *A Practical Guide to Heavy Tails*. Birkhäuser, pp. 311–335.
- Koutrouvelis, I. A. (1980). Regression-type estimation of the parameters of stable laws. *J. Am. Statist. Assoc.* 75:918–928.
- Koutrouvelis, I. A. (1981). An iterative procedure for estimation of the parameters of stable laws. *Commun. Statist. Simulation Comput.* 10:17–28.
- Künsch, H. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.* 17:1217–1241.
- Liu, J. (1997). Generalized method of moments estimation of affine diffusion processes. Working Paper, Graduate School of Business: Stanford University.
- Lukacs, E. (1970). *Characteristic Functions*. New York: Griffin.
- Madan, D. B., Seneta, E. (1990). The variance gamma (V.G.) model from share market returns. *J. Business* 63:511–524.
- Maheu, J., McCurdy, T. (2004). News arrival, jump dynamics and volatility components for individual stock returns. *J. Finance* 59:755–793.
- Mandelbrot, B. (1963). The variations of certain speculative prices. *J. Business* 36:394–419.
- Merton, R. (1976). Option pricing when underlying stock returns are discontinuous. *J. Financial Economics* 3:125–144.
- Mittnik, S., Rachev, S. T., Paoletta, M. S. (1998). Stable paretian modeling in finance: Some empirical and theoretical aspects. In: Adler, R. J., Feldman, R. E., Taquq, M. S., eds. *A Practical Guide to Heavy Tails*. Birkhäuser, pp. 79–110.



- Newey, W. K., McFadden, D. (1994). Large sample estimation and hypothesis testing. In: Engle, R. F., McFadden, D., eds. *Handbook of Econometrics*. Vol. 4. Amsterdam: North-Holland, pp. 2111–2245.
- Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* 33:1065–1076.
- Paulson, A. S., Delehanty, T. A. (1985). Modified weighted squared error estimation procedures with special emphasis on the stable laws. *Commun. Statist. Simulation Comput.* 14:927–972.
- Paulson, A. S., Holcomb, E. W., Leitch, R. A. (1975). The estimation of the parameters of the stable laws. *Biometrika* 62:163–170.
- Politis, D. N., White, H. (2004). Automatic block-length selection for dependent bootstrap. *Econometric Rev.* 23:53–70.
- Powell, M. (1964). An efficient method for finding the minimum of a function of several variables without calculating derivatives. *Comput. J.* 7:155–162.
- Press, S. J. (1972). Estimation in univariate and multivariate stable distributions. *J. Am. Statist. Assoc.* 67:842–846.
- Quandt, R. E. (1958). The estimation of parameters of a linear regression system obeying two separate regimes. *J. Am. Statist. Assoc.* 55:873–880.
- Quandt, R. E., Ramsey, J. B. (1978). Estimating mixtures of normal distributions and switching regressions. *J. Am. Statist. Assoc.* 73:730–738.
- Schmidt, P. (1982). An improved version of the Quandt-Ramsey MGF estimator for mixtures of normal distributions and switching regressions. *Econometrica* 50:501–524.
- Singleton, K. J. (2001). Estimation of affine asset pricing models using the empirical characteristic function. *J. Econometrics* 102:111–141.
- Thorton, J. C., Paulson, A. S. (1977). Asymptotic distribution of CF-based estimator for the stable laws. *Sankhya* 39:341–354.
- Titterton, D. M., Smith, A. F. M., Markov, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions*. Toronto, Chichester: John Wiley and Sons.
- Tran, K. C. (1998). Estimating mixtures of normal distributions via empirical characteristic function. *Econometric Rev.* 17:167–83.
- Yu, J. (1998). *The Empirical Characteristic Function in Time-Series Estimation*. Ph.D. thesis, University of Western Ontario.

