Empirical Likelihood Methods Based on Characteristic Functions With Applications to Lévy Processes

Ngai Hang CHAN, Song Xi CHEN, Liang PENG, and Cindy L. YU

Lévy processes have been receiving increasing attention in financial modeling. One distinctive feature of such models is that their characteristic functions are readily available. Inference based on characteristic functions is very useful for studying Lévy processes. By incorporating the recent advances in nonparametric approaches, empirical likelihood methods based on characteristic functions are developed in this paper for parameter estimation, testing a particular parametric class including the presence of a jump component in the Lévy process and testing for symmetry of a distribution. Simulation and case studies confirm the effectiveness of the proposed method.

KEY WORDS: Characteristic function; Empirical likelihood; Goodness-of-fit test; Lévy processes.

1. INTRODUCTION

Lévy processes are popular models for pricing the dynamics of financial securities due to their ability to capture jumps. There is an increasing set of empirical evidence that the value of a financial security may evolve in a discontinuous manner with jumps. And yet, whether there is really a jump in the underlying process is always an intensively contested hypothesis among financial practitioners; see, for example, the monograph of Cont and Tankov (2004) for background discussions.

Let $\{Y_s\}_{s\geq 0}$ be a continuous-time Lévy process, representing a financial security in a continuous fashion over time. The financial securities can be an interest rate or a stock price. Despite the process evolves continuously over time, it is only observed over a set of discrete time points, say $\delta, 2\delta, \dots, (n+1)\delta$, over a time span $[0,T]$ where $T=(n+1)\delta$ is the total amount of time the process is observed and δ is the unit sampling interval. Since Lévy processes have independent stationary increments, the increment $\{X_j = Y_{(j+1)\delta} - Y_{j\delta}\}_{j=1}^n$ is a set of independent observations with the same distribution as Y_δ, say F, and its characteristic function is

$$\phi(t) = E\{\exp(itX_j)\} = \int \exp(itx)F(dx).$$

Fitting a parametric class of characteristic functions to $\{Y_s\}_{s\geq 0}$ becomes how to fit a parametric class of characteristic functions based on observations $X_1, \dots, X_n$.

For parameter estimation based on the characteristic functions, several methods have been proposed in the literature; see Heathcote (1977) and Wells (1992). For testing, let $\phi_n(t) = \frac{1}{n}\sum_{j=1}^n \exp\{itX_j\}$ and consider

$$\beta_n(t;\boldsymbol{\theta}) = \sqrt{n}\{\phi_n(t) - \phi(t;\boldsymbol{\theta})\}, \tag{1.1}$$

where $\phi(t;\boldsymbol{\theta})$ denotes the characteristic function of X_j, and $\boldsymbol{\theta}$ is the vector of parameters for the process $\{Y_s\}_{s\geq 0}$. It is known that, in general, $\beta_n(t;\boldsymbol{\theta})$ does not converge weakly to a Gaussian process in $C([-a,a])$ for a positive constant a, although the finite dimensional distributions of $\beta_n(t;\boldsymbol{\theta})$ do, where $C([-a,a])$ denotes the space of continuous, complex-valued functions on $[-a,a]$ with the supremum norm. The reason is that the limiting process may have discontinuous sample paths. Therefore, under some extra conditions to ensure that the limit has continuous sample paths, one can show that

$$\beta_n(t;\boldsymbol{\theta}) \xrightarrow{D} \beta(t;\boldsymbol{\theta}) \tag{1.2}$$

in $C[-a,a]$, where $\beta(t;\boldsymbol{\theta})$ is a complex-valued Gaussian process with zero mean and covariance structure

$$\mathrm{E}\{\beta(t_1;\boldsymbol{\theta})\overline{\beta(t_2;\boldsymbol{\theta})}\} = \phi(t_1 - t_2;\boldsymbol{\theta}) - \phi(t_1;\boldsymbol{\theta})\phi(t_2;\boldsymbol{\theta}),$$

see Marcus (1981), Wells (1992), and Yukich (1985).

Since (1.2) is not always true, two methods for conducting goodness-of-fit tests via characteristic functions exist: one is to study the distance between the empirical characteristic function and the true characteristic function at a finite number of points $t_1, \dots, t_m$ with $m = m(n) \to \infty$ as $n \to \infty$; the other is to study an integrated distance by imposing extra conditions on $\phi(t;\boldsymbol{\theta})$ such that (1.2) holds. In modeling independent data, Fan (1997), Koutrouvelis (1980), and Koutrouvelis and Kellermeier (1981) employed the former approach to fit a parametric class to characteristic functions; Bilodeau and de Micheaux (2005), Epps (1983), Klar and Meintanis (2005), and Wells (1992) employed the latter approach to testing normality and fitted a parametric class to characteristic functions; Feuerverger and Mureika (1977), Henze, Klar, and Meintanis (2003), and Meintanis (2005) applied the latter approach to testing symmetry of a distribution. For the study of using characteristic functions to model dependent data, we refer to Jiang and Knight (2002) and Singleton (2001). A good review on the applications of empirical characteristic functions is Yu (2004).

In this paper, we propose to use the empirical likelihood (Owen 1988) to formulate the test statistic for the specification of characteristic functions and to conduct parameter estimation.

Ngai Hang Chan is Chair Professor of Statistics and Head, Department of Statistics, The Chinese University of Hong Kong, Shatin NT, Hong Kong (E-mail: *nhchan@sta.cuhk.edu.hk*). Song Xi Chen is Professor, Department of Statistics, Iowa State University, Ames, IA 50011-1210 (E-mail: *songchen@iastate.edu*) and Professor and Chairman, Department of Business Statistics and Econometrics, Guanghua School of Management, Peking University, Beijing 100651, China (E-mail: *csx@gsm.pku.edu.cn*). Liang Peng is Professor, School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332-0160 (E-mail: *peng@math.gatech.edu*). Cindy Yu is Assistant Professor, Department of Statistics, Iowa State University, Ames, IA 50011-1210 (E-mail: *cindyyu@iastate.edu*). This research was supported in part by grants from HKSAR-RGC-GRF 400306 and 400408, NSF-DMS 0604563 and 0714978, and NSF-SES 0631608. The authors would like to thank the editor, an associate editor, and two referees for helpful comments and suggestions.

Journal of the American Statistical Association
December 2009, Vol. 104, No. 488, Theory and Methods
DOI: 10.1198/jasa.2009.tm08349

Note that a characteristic function and its corresponding probability density function contain the same amount of model information as they are related to each other via a Fourier transform. Although it is more direct and efficient in formulating parameter estimators based on probability density functions, it is unclear if a test procedure based on density functions has any advantage over a test based on characteristic functions. Using the empirical likelihood to formulate the test statistic is tantamount to utilizing empirical likelihood's ability to self-studentize a raw discrepancy metric by its variance without explicitly estimating it.

We propose using the empirical likelihood method based on the integrated distance of characteristic functions (i) to estimate parameters, (ii) to test whether a characteristic function belongs to a particular parametric class, and (iii) to test whether the underlying distribution is symmetric around an unknown mean in this paper; see Section 2 for details. How to apply the proposed empirical likelihood test for testing various features of a Lévy process is discussed in Section 3. A simulation study and a case study are given in Sections 4 and 5, respectively. All proofs are presented in Section 6.

2. METHODOLOGY

Throughout this section we assume that $X_1, \dots, X_n$ are iid with distribution function F and characteristic function $\phi(t; \boldsymbol{\theta})$. This assumption encompasses Lévy processes as a Lévy process has independent stationary increments. Likewise, the proposed methods are also applicable to dependent data in view of the blockwise empirical likelihood method of Kitamura (1997). This issue for dependent data will be dealt with in a separate paper.

2.1 MELE for Parameters

As in Qin and Lawless (1994), we first study the maximum empirical likelihood estimator (MELE) based on the characteristic function as follows. Since $\phi(t; \boldsymbol{\theta}) = \mathrm{E} e^{itX_j}$, the standard empirical likelihood method for a mean can be employed, which defines the empirical likelihood function as

$$R_1(t; \boldsymbol{\theta}) = \sup\left\{\prod_{j=1}^n (np_j) : p_j \ge 0, \sum_{j=1}^n p_j = 1, \right.$$
$$\sum_{j=1}^n p_j \cos(tX_j) = \phi^R(t; \boldsymbol{\theta}),$$
$$\left. \sum_{j=1}^n p_j \sin(tX_j) = \phi^I(t; \boldsymbol{\theta})\right\},$$

where ϕ^R and ϕ^I denote the real and imaginary parts of ϕ, respectively. By the standard Lagrange multiplier procedure, the log-likelihood ratio becomes

$$l_1(t; \boldsymbol{\theta}) = -2\log R_1(t; \boldsymbol{\theta}) = \sum_{j=1}^n 2\log\{1 + \boldsymbol{\lambda}_1^{\mathrm{T}} \mathbf{Y}_j(t; \boldsymbol{\theta})\}, \quad (2.1)$$

where

$$\begin{aligned} \mathbf{Y}_j(t; \boldsymbol{\theta}) &= (\cos(tX_j) - \phi^R(t; \boldsymbol{\theta}), \sin(tX_j) - \phi^I(t; \boldsymbol{\theta}))^{\mathrm{T}} \\ &:= \mathbf{g}(t, X_j; \boldsymbol{\theta}) \end{aligned}$$

and $\boldsymbol{\lambda}_1 = \boldsymbol{\lambda}_1(t; \boldsymbol{\theta})$ satisfies

$$\sum_{j=1}^n \frac{\mathbf{Y}_j(t; \boldsymbol{\theta})}{1 + \boldsymbol{\lambda}_1^{\mathrm{T}} \mathbf{Y}_j(t; \boldsymbol{\theta})} = 0 \quad \text{for all } t \in [-a, a], \quad (2.2)$$

with some $a > 0$ given in the regularity conditions below. Then, the MELE is defined as

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta} \in \Omega} T_1(\boldsymbol{\theta}), \quad (2.3)$$

where $T_1(\boldsymbol{\theta}) = \int_{-a}^a l_1(t; \boldsymbol{\theta})\, dG_1(t)$ and G_1 is a given smooth distribution function.

Before deriving the asymptotic limit of $\hat{\boldsymbol{\theta}}$, we need the following regularity conditions.

A1. $\mathrm{E}\{\mathbf{g}(t, X_1; \boldsymbol{\theta}_0)\mathbf{g}^{\mathrm{T}}(t, X_1; \boldsymbol{\theta}_0)\}$ is positive definite for $t \in [-a, a]$ with some $a > 0$;

A2. $\frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{g}(t, x; \boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$ in a neighborhood of $\boldsymbol{\theta}_0$, say Ω_0, for $t \in [-a, a]$ and $x \in R$;

A3. $\sup_{\boldsymbol{\theta} \in \Omega_0} \|\frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{g}(t, x; \boldsymbol{\theta})\| \le H(t, x)$, where

$$\int_{t=-a}^a \int_{x=-\infty}^{\infty} H(t, x)\, dF(x)\, dG_1(t) < \infty;$$

A4. The rank of $\mathrm{E}\{\frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{g}(t, X_1; \boldsymbol{\theta}_0)\}$ is $\min\{2, d\}$ for all $t \in [-a, a]$, where d is the dimension of $\boldsymbol{\theta}$;

A5. $\frac{\partial^2}{\partial \boldsymbol{\theta}\, \partial \boldsymbol{\theta}^{\mathrm{T}}} \mathbf{g}(t, x; \boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$ for $\boldsymbol{\theta} \in \Omega_0$, $t \in [-a, a]$ and $x \in R$;

A6. $\sup_{\boldsymbol{\theta} \in \Omega_0} \|\frac{\partial^2}{\partial \boldsymbol{\theta}\, \partial \boldsymbol{\theta}^{\mathrm{T}}} \mathbf{g}(t, x; \boldsymbol{\theta})\| \le H(t, x)$, where H is given in A3.

Theorem 1. Under conditions A1–A4, with probability one, $T_1(\boldsymbol{\theta})$ attains its minimum at $\hat{\boldsymbol{\theta}}$, which satisfies $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| \le n^{-1/3}$ and

$$\begin{cases} \mathbf{Q}_{1n}(t; \hat{\boldsymbol{\theta}}, \boldsymbol{\lambda}_1(t; \hat{\boldsymbol{\theta}})) = 0 & \text{for all } t \in [-a, a] \\ \displaystyle\int_{-a}^a \mathbf{Q}_{2n}(t; \hat{\boldsymbol{\theta}}, \boldsymbol{\lambda}_1(t; \hat{\boldsymbol{\theta}}))\, dG_1(t) = 0, & \end{cases} \quad (2.4)$$

where

$$\begin{cases} \mathbf{Q}_{1n}(t; \boldsymbol{\theta}, \boldsymbol{\lambda}) = \dfrac{1}{n} \displaystyle\sum_{j=1}^n \frac{1}{1 + \boldsymbol{\lambda}^{\mathrm{T}} \mathbf{Y}_j(t; \boldsymbol{\theta})} \mathbf{Y}_j(t; \boldsymbol{\theta}), \\ \mathbf{Q}_{2n}(t; \boldsymbol{\theta}, \boldsymbol{\lambda}) = \dfrac{1}{n} \displaystyle\sum_{j=1}^n \frac{1}{1 + \boldsymbol{\lambda}^{\mathrm{T}} \mathbf{Y}_j(t; \boldsymbol{\theta})} \frac{\partial \mathbf{Y}_j^{\mathrm{T}}(t; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \boldsymbol{\lambda}. \end{cases} \quad (2.5)$$

Theorem 2. Under conditions A1–A6, for the estimator $\hat{\boldsymbol{\theta}}$ given in Theorem 1, we have as $n \to \infty$

$$\begin{aligned} \sqrt{n}\{\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\} &= -\left\{\int_{-a}^a \mathbf{s}_{21}(t)\mathbf{s}_{11}^{-1}(t)\mathbf{s}_{12}(t)\, dG_1(t)\right\}^{-1} \\ &\quad \times \left\{\int_{-a}^a \mathbf{s}_{21}(t)\mathbf{s}_{11}^{-1}(t)\sqrt{n}\mathbf{Q}_{1n}(t; \boldsymbol{\theta}_0, 0)\, dG_1(t)\right\} \\ &\quad + o_p(1) \\ &\xrightarrow{d} \mathrm{N}(0, \boldsymbol{\Sigma}) \end{aligned}$$

and

$$\sup_{-a\le t\le a}\left|\sqrt{n}\boldsymbol{\lambda}_1(t;\hat{\boldsymbol{\theta}})+\mathbf{s}_{11}^{-1}(t)\sqrt{n}\mathbf{Q}_{1n}(t;\boldsymbol{\theta}_0,0)\right.$$
$$\left.+\mathbf{s}_{11}^{-1}(t)\mathbf{s}_{12}(t)\sqrt{n}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)\right|=o_p(1),$$

where

$$\mathbf{s}_{11}(t)=-\mathrm{E}\{\mathbf{Y}_1(t;\boldsymbol{\theta}_0)\mathbf{Y}_1^{\mathrm{T}}(t;\boldsymbol{\theta}_0)\},$$
$$\mathbf{s}_{12}(t)=\mathbf{s}_{21}^{\mathrm{T}}(t)=\mathrm{E}\left\{\frac{\partial}{\partial\boldsymbol{\theta}}\mathbf{Y}_1(t;\boldsymbol{\theta}_0)\right\},$$
$$\boldsymbol{\Sigma}=\left\{\int_{-a}^{a}\mathbf{s}_{21}(t)\mathbf{s}_{11}^{-1}(t)\mathbf{s}_{12}(t)\,dG_1(t)\right\}^{-1}$$
$$\times\left\{\int_{-a}^{a}\int_{-a}^{a}\mathbf{s}_{21}(t_1)\mathbf{s}_{11}^{-1}(t_1)\boldsymbol{\Gamma}(t_1,t_2)\mathbf{s}_{11}^{-1}(t_2)\right.$$
$$\left.\times\,\mathbf{s}_{12}(t_2)\,dG_1(t_1)\,dG_1(t_2)\right\}$$
$$\times\left\{\int_{-a}^{a}\mathbf{s}_{21}(t)\mathbf{s}_{11}^{-1}(t)\mathbf{s}_{12}(t)\,dG_1(t)\right\}^{-1},$$

and

$$\boldsymbol{\Gamma}(t_1,t_2)=\mathrm{E}\{\mathbf{g}(t_1,X_t;\boldsymbol{\theta}_0)\mathbf{g}^{\mathrm{T}}(t_2;X_t;\boldsymbol{\theta}_0)\}.$$

Next we show how the MELE becomes asymptotically efficient by choosing a suitable weight function G_1. For simplicity we assume $\theta\in R$, that is, $d=1$. Motivated by Feuerverger and McDunnough (1981a, 1981b), define

$$\pi(t)=\frac{1}{2\pi}\int e^{-itx}\frac{\partial\log f(x;\theta_0)}{\partial\theta}\,dx \tag{2.6}$$

and

$$G_1'(t)=\{\mathbf{s}_{21}(t)\mathbf{s}_{11}^{-1}(t)\mathbf{s}_{12}(t)\}^{-1}\frac{\partial\phi(t;\theta_0)}{\partial\theta}\pi(t), \tag{2.7}$$

where f denotes the density function of X_t. By replacing the domain of integration $[-a,a]$ in Theorem 2 with $(-\infty,\infty)$, it can be shown that

$$\Sigma=\left[\mathrm{E}\left\{\frac{\partial\log f(X_t;\theta_0)}{\partial\theta}\right\}^2\right]^{-1}, \tag{2.8}$$

where Σ is given in Theorem 2 (the derivation is given in next section). Hence the MELE is asymptotically efficient if the weight function $\pi(t)$ is chosen according to (2.6). In practice, this choice may not be available owing to the lack of knowledge of the underlying density function. Simulation study suggests that a uniform weight function works well.

2.2 Empirical Likelihood Test

In this subsection, we give an empirical likelihood test procedure for a general model specification defined by a parametric characteristic function $\phi(t;\boldsymbol{\theta})$. Here the hypotheses are

$$H_0:\phi(t)\in\{\phi(t;\boldsymbol{\theta}):\exists\boldsymbol{\theta}\in\Omega\}$$

against

$$H_a:\phi(t)\notin\{\phi(t;\boldsymbol{\theta}):\forall\boldsymbol{\theta}\in\Omega\},$$

where $\Omega\subset R^d$ is a given set. Like Einmahl and McKeague (2003) and Li (2003), we consider the integrated version of test statistic, that is, $T_1(\hat{\boldsymbol{\theta}})$, where $\hat{\boldsymbol{\theta}}$ is the parameter estimator proposed in the previous section. Note that any other estimator for $\boldsymbol{\theta}$ with rate $n^{-1/2}$ can also be employed.

Theorem 3. Under conditions A1–A6, we have as $n\to\infty$,

$$\left|T_1(\hat{\boldsymbol{\theta}})-\left\{-W_2\right.\right.$$
$$\left.\left.+\mathbf{W}_1^{\mathrm{T}}\left\{\int_{-a}^{a}\mathbf{s}_{21}(t)\mathbf{s}_{11}^{-1}(t)\mathbf{s}_{12}(t)\,dG_1(t)\right\}^{-1}\mathbf{W}_1\right\}\right|=o_p(1),$$

where

$$\mathbf{W}_1=\int_{-\infty}^{\infty}\left\{\int_{-a}^{a}\mathbf{s}_{21}(t)\mathbf{s}_{11}^{-1}(t)\right.$$
$$\left.\times\,(\cos(tx),\sin(tx))^{\mathrm{T}}\,dG_1(t)\right\}dB_n(F(x)),$$
$$W_2=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left\{\int_{-a}^{a}(\cos(tx),\sin(tx))\mathbf{s}_{11}^{-1}(t)\right.$$
$$\left.\times\,(\cos(ty),\sin(ty))^{\mathrm{T}}\,dG_1(t)\right\}dB_n(F(x))\,dB_n(F(y)),$$

and $\{B_n(y):0\le y\le 1\}$ is a sequence of Brownian bridges.

Note that W_2 is contributed by the plug-in estimator $\hat{\boldsymbol{\theta}}$, and so the above limit is not a chi-squared distribution. As the convergence in the above theorem may be slow in finite samples and simulating the limiting distribution is not straightforward, a bootstrap method can be employed to obtain critical values of the test based on T_1. If we only know that $X_1,\dots,X_n$ are iid random vectors with characteristic function $\phi(t;\boldsymbol{\theta})$, then we simply resample from $X_1,\dots,X_n$ and compute the test statistic $T_1(\hat{\boldsymbol{\theta}})$ based on the resample. However, in case we know that $X_1,\dots,X_n$ are δ increments of a Lévy process, which satisfies some stochastic equation with unknown parameters depending on H_0 such as examples given in the simulation study below, then the key of the bootstrap algorithm is to reproduce a sample path under H_0. Specifically, we apply the following parametric bootstrap procedures to approximate the distribution of the test statistic T_1. An underlying assumption is that the Lévy process $\{Y_s\}_{s\ge 0}$ is fully determined by a set of finite dimensional parameter $\boldsymbol{\theta}$ under H_0, so that we can write Y_s as $Y_s(\boldsymbol{\theta})$ to highlight such a determination. Note that $X_1,\dots,X_n$ are the δ increments of Y_s.

1. Generate a resample path $\{Y_\delta^\star,\dots,Y_{(n+1)\delta}^\star\}$ according to the Lévy process $\{Y_s(\hat{\boldsymbol{\theta}})\}_{s\ge 0}$ at the same sampling frequency, and reestimate the parameter $\boldsymbol{\theta}$ based on the δ increments of the resample path and denote it as $\hat{\boldsymbol{\theta}}^\star$.
2. Compute a version of the test statistic $T_1(\hat{\boldsymbol{\theta}})$, say $T_1^\star(\hat{\boldsymbol{\theta}}^\star)$, based on the δ increments of the resample path $\{Y_\delta^\star,\dots,Y_{(n+1)\delta}^\star\}$.
3. Repeat steps 1 and 2 B times where B is a large number and evaluate $T_1^{\star,1}(\hat{\boldsymbol{\theta}}^{\star,1})\le T_1^{\star,2}(\hat{\boldsymbol{\theta}}^{\star,2})\le\cdots\le T_1^{\star,B}(\hat{\boldsymbol{\theta}}^{\star,B})$.

Then, the bootstrap approximation to the α-upper quantile of $T_1(\hat{\boldsymbol{\theta}})$ is $T_1^{\star,[B(1-\alpha)]+1}$, and a test at α significant level rejects H_0 if $T_1(\hat{\boldsymbol{\theta}})\ge T_1^{\star,[B(1-\alpha)]+1}$.

2.3 Test for Symmetry

We now investigate the possibility of employing empirical likelihood methods via characteristic functions to test for symmetry. Testings for symmetry and conditional symmetry are important for improving inference and predictions. For example, Fan and Gencay (1995) applied their test for symmetry to a linear regression model

$$\Delta P_i = \beta_0 + \beta_1 \Delta M_i + \epsilon_i,$$

where ΔP_i is the average annual growth rate of consumer prices and ΔM_i is the growth rate of the stock of currency across 83 countries, and confirmed the symmetry of ϵ_i.

Assume that $F(x)$ is symmetric around the unknown mean μ. Hence, $\mathrm{E}\sin(t(X_j-\mu)) = 0$. As before, an empirical likelihood function is defined as

$$R_2(t;\mu) = \sup\left\{\prod_{j=1}^n (np_j) : p_j \ge 0, \sum_{j=1}^n p_j = 1, \sum_{j=1}^n p_j \sin(t(X_j-\mu)) = 0\right\},$$

and the log-likelihood ratio can be obtained as

$$l_2(t;\mu) = \sum_{j=1}^n 2\log\{1 + \lambda_2 \sin(t(X_j-\mu))\},$$

where $\lambda_2 = \lambda_2(t;\mu)$ satisfies

$$\sum_{j=1}^n \frac{\sin(t(X_j-\mu))}{1+\lambda_2\sin(t(X_j-\mu))} = 0. \tag{2.9}$$

Our test statistic is defined as

$$T_2 = \int_{-a}^a l_2(t;\hat\mu)\,dG_2(t),$$

where G_2 is a given symmetric distribution function around zero and $\hat\mu = \frac{1}{n}\sum_{j=1}^n X_j$. Note that we could replace $\hat\mu$ by the maximum empirical likelihood estimator as in the above section.

Theorem 4. Assume $\mathrm{E}(|X_1|^3) < \infty$ and F is symmetric around the true mean μ_0. Then as $n \to \infty$

$$\left|T_2 - \left\{\int_{-\infty}^\infty \int_{-\infty}^\infty \left\{\int_{-a}^a \frac{A(t,x)A(t,y)}{\mathrm{E}\{\sin^2(t(X_1-\mu_0))\}}\,dG_1(t)\right\} dB_n(F(x))\,dB_n(F(y))\right\}\right| = o_p(1),$$

where

$$A(t,x) = \sin(t(x-\mu_0)) - \mathrm{E}\{\cos(t(X_1-\mu_0))\}t\int_{-\infty}^\infty s\,dB_n(F(s))$$

and B_n is defined in Theorem 3.

We remark that a bootstrap method can be employed to obtain critical values for the preceding limit, that is, resample from $X_1,\ldots,X_n$ and then recalculate the test statistic T_2 based on the resample. Note that both MELE and the goodness-of-fit test based on $T_n(\boldsymbol\theta)$ and $T_1(\hat{\boldsymbol\theta})$ require specifying the tuning parameter a. In general, if the underlying distribution is sufficiently smooth such that it has an infinite number of moments, then the value of a may be chosen to be small since the characteristic function near the origin $t=0$ contains enough information about the underlying model. Under such circumstances, a moderate value of a would be sufficient when the sample size n is large. For heavier tail distributions, a larger value of a would be needed to capture more information for the underlying model. Our simulation study shows that the value $a = 1/2$ and a uniform weight function G_1 work reasonably well for the Lévy models under consideration.

3. TESTING FOR LÉVY PROCESSES

The proposed estimation and testing methods as outlined in Sections 2.1 and 2.2 can be used to test various hypotheses with regard to Lévy processes. According to the Lévy–Khintchine formula, the characteristic function $\phi(t)$ of X_j has the form

$$\phi(t) = \exp\left(\delta\left[it\theta_1 - \theta_2 t^2/2 + \int J(dx)\{\exp(itx) - 1 - itxI(|x|\le 1)\}\right]\right), \tag{3.1}$$

where (θ_1, θ_2, J) is the triplet of the Lévy process and $J(dx)$ is the Lévy jump measure such that $\int \min\{1, x^2\}J(dx) < \infty$. If we use $\vec{\boldsymbol\theta} = (\theta_1,\theta_2)^{\mathrm{T}}$ to denote the parameters in the continuous part of the Lévy process, we can write $\phi(t)$ as $\phi(t;\vec{\boldsymbol\theta}, J)$ to highlight the role of the triple. Let

$$\phi_c(t;\vec{\boldsymbol\theta}) = \exp(\delta[it\theta_1 - \theta_2 t^2/2]) \quad \text{and}$$
$$\phi_j(t) = \exp\left(\delta\left[\int J(dx) \times \{\exp(itx) - 1 - itxI(|x|\le 1)\}\right]\right) \tag{3.2}$$

be, respectively, the characteristic functions of the continuous and jump components. Then $\phi(t) = \phi(t;\vec{\boldsymbol\theta}, J) = \phi_c(t;\vec{\boldsymbol\theta}) \times \phi_j(t;J)$. Note that $\phi_c(t;\vec{\boldsymbol\theta})$ is in fact the characteristic function of $\mathrm{N}(\theta_1,\theta_2)$.

From the Lévy–Itò decomposition (Sato 1999), the Lévy process Y_s can be written as

$$Y_s = \theta_1 s + \theta_2 W_s + Y_s^{(3)} + Y_s^{(4)}, \tag{3.3}$$

where W_s is a standard Brownian Motion, $Y_s^{(3)}$ is a compound Poisson process, $Y_s^{(4)}$ is a pure jump martingale, and W_s, $Y_s^{(3)}$, and $Y_s^{(4)}$ are mutually independent. Further, the compound Poisson process $Y_s^{(3)}$ can be written as $Y_s^{(3)} =: \sum_{i=1}^{N_s} J_i$, where N_s is a Poisson process with intensity λ, $\{J_i\}$ are independent and identically distributed random variables with distribution F_J, and $\{J_i\}$ is independent of N_s. The characteristic function of the increment of $Y_s^{(3)}$ is

$$\phi_{cp}(t;\lambda,F_J) = \exp\left\{\delta\int \exp(itx - 1 - itx)\lambda F_J(dx)\right\}. \tag{3.4}$$

Denote the characteristic function of the increment of the pure jump process $Y_s^{(4)}$ as $\phi_{pj}(t)$. Then, the characteristic function due to jump $\phi_j(t)$ as given in (3.2) can be written as

$$\phi_j(t) = \phi_{cp}(t;\lambda,F_J)\phi_{pj}(t).$$

Since characteristic functions are generally available for Lévy processes as given above and the underlying distribution function or probability density function may not admit an analytic expression for a Lévy distribution, fitting a parametric class of characteristic functions and conducting inference (testing and estimation) based on the characteristic function constitute important tools for the applications of Lévy processes in finance and mathematical physics (see Barndorff-Nielsen, Cox, and Klüppelberg 2001). If data comes from a Lévy process, a relevant question is whether the data fits a class of characteristic functions with or without jumps. To answer this question, we could first test

$$H_0: \phi(t) = \phi_c(t; \vec{\boldsymbol{\theta}}) \quad \text{for all } t \in R \text{ and a specific } \vec{\boldsymbol{\theta}} \in \Theta \tag{3.5}$$

against

$$H_1: \phi(t) \neq \phi_c(t; \vec{\boldsymbol{\theta}}) \quad \text{for some } t \in R \text{ and for all } \vec{\boldsymbol{\theta}} \in \Theta. \tag{3.6}$$

If H_0 is rejected, we would entertain a Lévy process with jumps. Further, we can also test if the jump is a compound Poisson with a specific jump size distribution $F_J = F_{J\xi}$ (e.g., the Gaussian distribution), where ξ represents extra parameters that defines F_J. In this case, we test

$$H_0: \phi(t) = \phi_c(t; \vec{\boldsymbol{\theta}})\phi_{cp}(t; \lambda, F_{J\xi}) \quad \text{for all } t \in R \text{ and a specific } (\vec{\boldsymbol{\theta}}, \lambda, \xi) \in \Theta \tag{3.7}$$

against

$$H_1: \phi(t) \neq \phi_c(t; \vec{\boldsymbol{\theta}})\phi_{cp}(t; \lambda, F_{J\xi}) \quad \text{for some } t \in R \text{ and for all } (\vec{\boldsymbol{\theta}}, \lambda, \xi) \in \Theta. \tag{3.8}$$

Besides the compound Poisson setting, the proposed set up can also be used to test for a specific type of jump. If $\{X_j\}_{j=1}^n$ are iid but not necessarily the increments of a Lévy process, we can test for the appropriateness of the characteristic function of a particular parametrization. Specific examples of these tests are illustrated in Section 5.

Note that characteristic functions of Lévy processes are implicitly determined by the sampling interval δ as indicated in (3.1) and (3.2). The proposed empirical likelihood based estimation and testing for Lévy processes require the sampling interval δ to be fixed. Nevertheless, δ can be chosen to be a fixed small number when dealing with high frequency data.

4. SIMULATION STUDIES

4.1 Simulated Models

Using discretely observed data, simulation studies were conducted to examine if the proposed empirical likelihood (EL) estimator can effectively identify the unknown parameters of continuous time models and if the proposed test procedure can adequately test for the appropriateness of a hypothesized parametric process. We consider the following three commonly used models with known characteristic functions.

Black–Scholes Model (BS). Suppose that the log stock price, Y_s, satisfies the following stochastic differential equation:

$$dY_s = \mu\, ds + \sigma\, dW_s, \tag{4.1}$$

where the parameter μ reflects the drift and σ models the volatility of the log stock prices. It is a Black–Scholes type of model since μ and σ are assumed to be constant over time. The Lévy triplet of this process is $(\mu, \sigma^2, J(dx) = 0)$, where zero Lévy measure indicates that the sample path is continuous. The increment $X_j = Y_{(j+1)\delta} - Y_{j\delta}$ has a normal distribution $N(\mu\delta, \sigma^2\delta)$, and its characteristic function is

$$\phi(t; \boldsymbol{\theta}) = \exp\{\delta(it\mu - \sigma^2 t^2/2)\}.$$

Despite the popularity of the celebrated Black–Scholes model, one of the main problems with the Black–Scholes model is that the log returns of stocks are not normally distributed. They are skewed and have kurtosis higher than that of a normal distribution. Recently, models with jumps that are able to capture skewness and excess kurtosis have been proposed. Two most commonly used models are: finite-activity jump processes and infinite-activity jump processes.

Black–Scholes Model With Merton Jumps (BS-MJ). Consider the model

$$dY_s = \mu\, ds + \sigma\, dW_s + J_s\, dN_s, \tag{4.2}$$

where N_s is a Poisson process with intensity parameter $\lambda > 0$ and J_s is the jump size following a normal distribution $N(\mu_J, \sigma_J^2)$, independent of W_s. The jump part in (4.2) is a compound Poisson process proposed in Merton (1976), hence it is also called Merton jumps. The Lévy triplet of BS-MJ process is $(\mu, \sigma^2, J(dx) = \lambda f_J(x)\, dx)$ where $f_J(x)$ is the density of $N(\mu_J, \sigma_J^2)$. Since the Lévy measure $J(dx)$ is integrable, that is, $J(R) = \int_{-\infty}^{\infty} \lambda f_J(x)\, dx = \lambda < \infty$, the process of (4.2) is called a finite activity jump process which allows a finite number of jumps within any finite time interval. Although the increment X_j does not have closed form density, its characteristic function is given by $\phi(t; \boldsymbol{\theta}) = \exp\{\delta[it\mu - \sigma^2 t^2/2 + \lambda(\exp(i\mu_J t - \sigma_J^2 t^2/2) - 1)]\}$.

Motivated by the fact that the Poisson processes can be well approximated by Bernoulli processes if the interval is small, we simulated the sample paths of (4.2) by the following discretization:

$$Y_{s+\delta} - Y_s = \mu\delta + \sigma\sqrt{\delta} Z + \sum_{k=1}^{200} w_k,$$

where Z is a standard $N(0, 1)$, $w_k = N_k J_k$, $P(N_k = 1) = \lambda \frac{\delta}{200} \exp(-\lambda \frac{\delta}{200})$, and $J_k \sim N(\mu_J, \sigma_J^2)$.

Variance Gamma Model (VG). The class of Variance Gamma (VG) distributions was introduced by Madan and Seneta (1987) as a model for stock returns. One way of defining a VG process is by means of a (Gamma) time-changed Brownian motion with drift. More precisely, let G_s be a Gamma process with the unit mean rate and the variance rate of ν, that is, with parameters $(1/\nu, \nu)$ where $\nu > 0$. Let W_s denote a standard Brownian motion, $\sigma > 0$ and $\mu \in R$, then the VG process Y_s with parameters $\mu, \sigma > 0$, and $\nu > 0$ is defined as

$$dY_s = \mu\, dG_s + \sigma\, dW_{G_s}, \tag{4.3}$$

where the calender time s in (4.1) is replaced by a random time G_s. The Lévy triplet of (4.3) is $(\gamma, 0, J(dx))$, where

$$\gamma = C(MG)^{-1}\big(G(\exp(-M)-1) - M(\exp(-G)-1)\big),$$

$$J(dx) = \begin{cases} C\exp(Gx)|x|^{-1}, & x<0 \\ C\exp(-Mx)x^{-1}, & x>0, \end{cases}$$

$C = 1/\nu$, $M = (\sqrt{2\sigma^2/\nu+\mu^2} - \mu)/\sigma^2$, and $G = (\sqrt{2\sigma^2/\nu+\mu^2}+\mu)/\sigma^2$.

This Lévy measure has infinite mass, hence a VG process has infinitely many jumps in any finite time interval. The density form of X_j is not known, but the characteristic function of X_j is $\phi(t;\boldsymbol{\theta}) = (1 - it\mu\nu + \sigma^2\nu t^2/2)^{-\delta/\nu}$. Note that when $\mu = 0$, the characteristic function becomes a real-valued function, indicating that the density of X_j is symmetric under this situation. Negative (positive) values of μ result in negative (positive) skewness and the parameter ν primarily controls the kurtosis. Explicit expressions for the first three central moments of the return distribution over an interval of length δ were derived in Madan, Carr, and Chang (1998) and are given as follows:

$$\mathrm{E}[X_j] = \mu\delta, \tag{4.4}$$

$$\mathrm{E}\big[(X_j - \mathrm{E}(X_j))^2\big] = (\mu^2\nu + \sigma^2)\delta, \tag{4.5}$$

$$\mathrm{E}\big[(X_j - \mathrm{E}(X_j))^3\big] = (2\mu^3\nu^2 + 3\sigma^2\mu\nu)\delta. \tag{4.6}$$

We simulated the sample path from this process by

$$Y_{s+\delta} - Y_s = \mu g + \sigma\sqrt{g}Z,$$

where Z is a standard N(0, 1), g is distributed as $\Gamma(\delta/\nu, \nu)$ and is independent of Z.

For each model, we simulated 500 sample paths each of size $n = 125$, 250, 500, and 1000 starting at the initial value $Y_0 = 0$ with weekly frequency ($\delta = 1/52$). By choosing the uniform weight function $G_1(t)$, $T_1(\boldsymbol{\theta})$ in (2.3) can be approximated by the Riemann sum of $l_1(t;\boldsymbol{\theta})$ defined in (2.1) evaluated at a number of equally spaced grid values of t in $[-a, a]$, where $a = 1/2$. We have also conducted simulation studies for other value of a (not reported here). The results were very similar to $a = 1/2$, indicating that the estimates are not sensitive to the choice of a for the Lévy models considered. In our simulation and case study, we set the number of grids equal to 100.

We first evaluate the quality of the proposed maximal EL estimator by estimating the parameters of these models using all simulated sample paths. We then construct three examples to evaluate the performance of the proposed tests in model validation. The three hypotheses are:

Test 1. H_0 : the process is the BS model (4.1). For computing powers of the proposed empirical likelihood test, we simulated data from the BS model with Merton jumps (4.2).

Test 2. H_0 : the process is the VG model (4.3) with $\mu = 0$. For computing powers of the test, we simulated data from the VG model with nonzero drift μ.

Test 3. H_0 : the process is the BS-MJ model (4.2). For power calculation, we simulated data from the VG model (4.3) with infinity-activity jumps.

4.2 Simulation Results

Table 1 reports the empirical means and standard errors of the parameter estimates obtained from the 500 simulated sample paths, as well as the true parameter values used for generating the data. When the sample size increases, standard errors of the estimates decreases, indicating the consistency of the estimators. We note that from Table 1(a) for the BS model where the maximum likelihood estimator (MLE) is available, the proposed maximal EL estimates are very close to the MLEs. For VG model, although the density is not available in closed form, the moments exist as explicit functions of unknown parameters as shown in (4.4)–(4.6). We also applied the method of moments (MME) to estimate the parameters so that they can be compared with the EL estimates [Table 1(c)]. It is interesting to find that for parameters μ and σ the maximal EL estimates are consistently more efficient than the MMEs over all the sample sizes considered. For the parameter ν, the maximal EL estimate has smaller standard errors than the MMEs for small sizes of $n = 125$ and 250, while the MMEs performed better for larger sample sizes of $n = 500$ and 1000. For BS-MJ model, where no MLE or MME to compare with, the proposed estimates as reported in Table 1(b) are reasonably close to the true values and the standard errors decrease as the sample size increases.

Table 2 reports the empirical sizes and powers of the proposed tests based on $B = 250$ bootstrap resample paths for each simulated data used in the estimation for $n = 125$, 250, and 500. We observe that the tests gave satisfactory sizes under BS model. The sizes under BS-MJ and VG ($\mu = 0$) models were slightly underestimated. But when the sample sizes increase, the rejection rates get closer to the nominal level 0.05. In the first test where we used data with jumps (BS-MJ) to test the continuous diffusion model BS, the powers are very high, achieving 97% even for the smallest sample size. For the second example where we used data from the asymmetric distribution (VG with $\mu \neq 0$) to test the symmetric structure (VG with $\mu = 0$) and the third example where we used the infinity-activity jump data (VG) to test the finite activity jump process (BS-MJ), the powers are lower than those of the first test, ranging from 17% to 70%, across different sample sizes. This might be due to the fact that the parameters can be more precisely estimated under BS model than under BS-MJ and VG models (as seen in Table 1) since the latter two models are more complex with more parameters. These simulation results suggest that the proposed maximal EL estimator accurately estimates model parameters and the proposed testing procedure adequately examines the validity of a specific parametric setup.

5. CASE STUDY

It is now a well known fact that Lévy models provide a better fit to many financial time series; see Cont and Tankov (2004) for details. In this section, we examine empirically whether our proposed testing procedure can adequately check for the validity of the Lévy jump models in modeling the return dynamics of the S&P500 index.

We computed summary statistics of the continuously compounded returns of the S&P500 index (the log difference of index levels) between January 2, 1987 and December 31, 2007. The index is sampled at weekly frequency, and in total we have 1059 observations. The mean weekly rate of return is 0.002 and the volatility is 0.021. Figure 1 plots the level and log difference of the S&P500 index. The sample period covers some major events in the history of the U.S. stock market, such as the stock market crash of 1987, the long boom in the late 1990 and the 9/11 crash in 2001.

Table 1. Empirical averages and their standard errors (in parentheses) of the maximum likelihood estimates (MLE) or the method of moments estimates (MME) and the proposed empirical likelihood estimates (EL) under the three models

(a) Black–Scholes model (BS)

n		$\mu = 0.05$	$\sigma = 0.3$
125	MLE	0.044 (0.192)	0.298 (0.019)
	EL	0.043 (0.192)	0.297 (0.019)
250	MLE	0.047 (0.133)	0.299 (0.014)
	EL	0.048 (0.133)	0.299 (0.014)
500	MLE	0.051 (0.097)	0.300 (0.010)
	EL	0.051 (0.097)	0.299 (0.010)
1000	MLE	0.047 (0.068)	0.300 (0.007)
	EL	0.047 (0.068)	0.300 (0.007)

(b) Black–Scholes model with Merton jumps (BS-MJ)

n		$\mu = 0.05$	$\sigma = 0.3$	$\lambda = 2.0$	$\mu_J = -2.0$	$\sigma_J = 6.0$
125	EL	0.223 (1.465)	0.522 (0.976)	1.913 (1.078)	−2.508 (3.619)	4.197 (2.458)
250	EL	0.120 (1.055)	0.506 (0.975)	1.838 (0.755)	−2.285 (2.519)	5.153 (1.750)
500	EL	0.060 (0.695)	0.443 (0.798)	1.942 (0.551)	−2.008 (1.574)	5.644 (1.140)
1000	EL	0.096 (0.472)	0.428 (0.746)	1.936 (0.380)	−2.095 (1.052)	5.945 (0.743)

(c) Variance Gamma model (VG)

n		$\mu = -0.3$	$\sigma = 0.3$	$\nu = 2.0$
125	MME	−0.283 (0.315)	0.289 (0.237)	0.770 (2.283)
	EL	−0.258 (0.256)	0.231 (0.170)	0.841 (1.310)
250	MME	−0.293 (0.222)	0.333 (0.204)	1.230 (3.115)
	EL	−0.270 (0.195)	0.279 (0.147)	0.897 (1.114)
500	MME	−0.292 (0.159)	0.344 (0.140)	1.131 (0.618)
	EL	−0.283 (0.148)	0.303 (0.112)	1.102 (0.958)
1000	MME	−0.299 (0.118)	0.342 (0.100)	1.387 (0.651)
	EL	−0.297 (0.117)	0.305 (0.075)	1.450 (0.779)

Table 2. Size and power evaluations (in percentage) under the three test settings

	Size	Power
(a) H_0 : BS model versus H_1 : BS model with Merton jumps		
$n = 125$	4.2	97
$n = 250$	5.0	100
$n = 500$	4.6	100
(b) H_0 : Symmetric VG model ($\mu = 0$) versus H_1 : Asymmetric VG model $\mu \neq 0$		
$n = 125$	3.4	17
$n = 250$	4.0	52
$n = 500$	4.6	70
(c) H_0 : BS model with Merton jumps versus H_1 : VG model		
$n = 125$	3.8	17
$n = 250$	4.0	44
$n = 500$	4.0	71

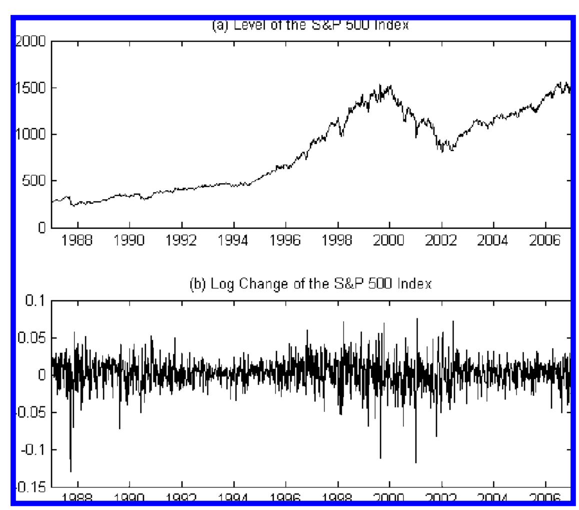

Figure 1. The weekly time series of the level and log change of the S&P500 index from January 2, 1987 to December 31, 2007.

Table 3. Empirical estimation for the S&P500 index between January 2, 1987 and December 31, 2007

(a) Black–Scholes model (BS)

	μ	σ
MLE	0.086 (0.032)	0.154 (0.003)
EL	0.086 (0.032)	0.154 (0.003)

(b) BS model with Merton jumps (BS-MJ)

	μ	σ	λ	μ_J	σ_J
EL	0.121 (0.026)	0.125 (0.015)	1.495 (0.395)	−0.026 (0.015)	0.064 (0.013)

(c) Variance Gamma model (VG)

	μ	σ	ν
EL	0.080 (0.017)	0.120 (0.038)	1.056 (0.884)

The proposed parameter estimates for the three models considered are reported in Table 3. For BS model, the maximal EL estimates are very close to the MLEs. The maximal EL estimates of the drift μ is 0.085 for both BS and VG models, which lead to the estimated mean of log returns, $\hat{\mu}\delta = 0.085/52 = 0.002$, which is equal to the summary statistic of the mean log returns (0.002). The estimate of the drift μ under BS-MJ model is 0.110, higher than those of other two models. This is because the mean of log returns under BS-MJ is $(\mu + \mu_J\lambda)\delta$ and the model is estimated with a negative mean jump size ($\mu_J < 0$). As a result, μ needs to be bigger to revert the overall mean return level as reflected in the data. The estimated volatility of the log returns under the BS model is $\hat{\sigma}\sqrt{\delta} = 0.154\sqrt{1/52} = 0.021$, which is equal to the summary statistic of the volatility for the log returns (0.021). The maximal EL estimates of σ for BS-MJ and VG are 0.129 and 0.126, respectively, lower than that of BS model (0.154). This indicates that including jumps in the process capture large movements and consequently the volatility of the Brownian motion part does not have to be as big as before. In BS-MJ model, our estimates of jump parameters suggest that about 1.5 jumps occur per year with a negative average jump size (−0.017) and a standard deviation of 0.065, which is consistent with the empirical finding that there are always bigger negative spikes observed than the positive ones in the S&P500. Under VG model, the estimated standard deviation of the log returns is $\sqrt{(\hat{\mu}^2\hat{\nu} + \hat{\sigma}^2)\delta} = \sqrt{(0.085^2 \times 1.059 + 0.126^2)\delta} = 0.021$, which is the same as the summary statistic of the volatility for log returns. Overall, the maximal EL estimator provides reasonable parameter estimates.

We also applied the proposed test to test for the appropriateness of the three parametric models considered. We tested the VG model with zero drift, that is, the symmetric VG model. Table 4 reports the p-values of the tests, as well as the test statistics and the approximated critical value of the tests based on 500 bootstrap resamples. The significant level of tests is 0.05. There is no empirical support for BS model. This is not surprising since many empirical evidence revealed that the classical Black–Scholes model does not constitute a suitable model for financial time series. We rejected the symmetric VG model due to the very low p-values (0.01), which is consistent with the fact that significant skewness are observed for log returns of indexes. By allowing a nonzero drift in VG model, we fail to reject the asymmetric VG model (p-value = 0.328). By including the Merton jumps into the BS model, the p-value changed from 0 for BS model to 0.736 for BS-MJ model. This implies that adding jumps (finite or infinite activity rate) does help capturing the underlying dynamics of the S&P500 index. In conclusion, our analysis demonstrates the empirical relevance of the Lévy jump models in modeling the S&P500 index returns.

Table 4. p-values for the S&P500 data under four models

	Test stats	$l^*_{0.05}$	p-values
BS	3.002	1.493	0.0
BS-MJ	5.410	17.668	0.736
VG ($\mu = 0$)	6.412	3.891	0.010
VG ($\mu \neq 0$)	3.477	6.288	0.328

6. PROOFS

Proof of Theorem 1. It can be shown in the same way as in the proof of lemma 1 of Qin and Lawless (1994).

Proof of Theorem 2. It is easy to check that

$$\begin{cases} \dfrac{\partial}{\partial\boldsymbol{\theta}}\mathbf{Q}_{1n}(t;\boldsymbol{\theta}_0,0) = \dfrac{1}{n}\displaystyle\sum_{j=1}^{n}\dfrac{\partial}{\partial\boldsymbol{\theta}}\mathbf{Y}_j(t;\boldsymbol{\theta}_0) \xrightarrow{p} \mathbf{s}_{12}(t), \\ \dfrac{\partial}{\partial\boldsymbol{\lambda}^{\mathrm{T}}}\mathbf{Q}_{1n}(t;\boldsymbol{\theta}_0,0) = -\dfrac{1}{n}\displaystyle\sum_{j=1}^{n}\mathbf{Y}_j(t;\boldsymbol{\theta}_0)\mathbf{Y}_j^{\mathrm{T}}(t;\boldsymbol{\theta}_0) \xrightarrow{p} \mathbf{s}_{11}(t), \\ \dfrac{\partial}{\partial\boldsymbol{\theta}}\mathbf{Q}_{2n}(t;\boldsymbol{\theta}_0,0) = 0, \\ \dfrac{\partial}{\partial\boldsymbol{\lambda}^{\mathrm{T}}}\mathbf{Q}_{2n}(t;\boldsymbol{\theta}_0,0) = \dfrac{1}{n}\displaystyle\sum_{j=1}^{n}\dfrac{\partial}{\partial\boldsymbol{\theta}}\mathbf{Y}_j^{\mathrm{T}}(t;\boldsymbol{\theta}_0) \xrightarrow{p} \mathbf{s}_{21}(t) \end{cases} \tag{6.1}$$

uniformly in $t \in [-a,a]$. Put $\delta_n = \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| + \sup_{-a\le t\le a}\|\boldsymbol{\lambda}_1(t;\hat{\boldsymbol{\theta}})\|$. Then, by Taylor expansions, we have

$$\begin{aligned} 0 &= \mathbf{Q}_{1n}(t;\hat{\boldsymbol{\theta}},\boldsymbol{\lambda}_1(t;\hat{\boldsymbol{\theta}})) \\ &= \mathbf{Q}_{1n}(t;\boldsymbol{\theta}_0,0) + \frac{\partial\mathbf{Q}_{1n}(t;\boldsymbol{\theta}_0,0)}{\partial\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\ &\quad + \frac{\partial\mathbf{Q}_{1n}(t;\boldsymbol{\theta}_0,0)}{\partial\boldsymbol{\lambda}^{\mathrm{T}}}\boldsymbol{\lambda}_1(t;\hat{\boldsymbol{\theta}}) + o_p(\delta_n) \end{aligned} \tag{6.2}$$

uniformly in $t \in [-a,a]$, and

$$\begin{aligned} 0 &= \int_{-a}^{a}\mathbf{Q}_{2n}(t;\hat{\boldsymbol{\theta}},\boldsymbol{\lambda}_1(t;\hat{\boldsymbol{\theta}}))\,dG_1(t) \\ &= \int_{-a}^{a}\bigg\{\mathbf{Q}_{2n}(t;\boldsymbol{\theta}_0,0) + \frac{\partial\mathbf{Q}_{2n}(t;\boldsymbol{\theta}_0,0)}{\partial\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\ &\quad + \frac{\partial\mathbf{Q}_{2n}(t;\boldsymbol{\theta}_0,0)}{\partial\boldsymbol{\lambda}^{\mathrm{T}}}\boldsymbol{\lambda}_1(t;\hat{\boldsymbol{\theta}})\bigg\}\,dG_1(t) + o_p(\delta_n). \end{aligned} \tag{6.3}$$

It follows from (6.2) that

$$\begin{aligned} \boldsymbol{\lambda}_1(t;\hat{\boldsymbol{\theta}}) &= -\mathbf{s}_{11}^{-1}(t)\mathbf{Q}_{1n}(t;\boldsymbol{\theta}_0,0) \\ &\quad - \mathbf{s}_{11}^{-1}(t)\mathbf{s}_{12}(t)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + o_p(\delta_n). \end{aligned} \tag{6.4}$$

Substituting (6.4) into (6.3), we get

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = -\left\{\int_{-a}^{a} \mathbf{s}_{21}(t)\mathbf{s}_{11}^{-1}(t)\mathbf{s}_{12}(t)\,dG_1(t)\right\}^{-1} \times \left\{\int_{-a}^{a} \mathbf{s}_{21}(t)\mathbf{s}_{11}^{-1}(t)\mathbf{Q}_{1n}(t;\boldsymbol{\theta}_0,0)\,dG_1(t)\right\} + o_p(\delta_n). \tag{6.5}$$

Hence the theorem follows from (6.4) and (6.5).

Proof of (2.8). Define $\mathbf{M} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix}$, $\epsilon(t, X_j; \boldsymbol{\theta}) = \exp\{itX_j\} - \phi(t;\boldsymbol{\theta})$, and let $\bar{\epsilon}(t, X_j; \boldsymbol{\theta})$ denotes the conjugate of $\epsilon(t, X_j; \boldsymbol{\theta})$. Then

$$\mathbf{g}(t, X_j; \boldsymbol{\theta}) = \mathbf{M}\begin{pmatrix} \epsilon(t, X_j; \boldsymbol{\theta}) \\ \bar{\epsilon}(t, X_j; \boldsymbol{\theta}) \end{pmatrix}. \tag{6.6}$$

It is easy to check that

$$\mathrm{E}\left\{\frac{\partial}{\partial\boldsymbol{\theta}}\epsilon(t, X_j; \boldsymbol{\theta}_0)\right\} = -\frac{\partial}{\partial\boldsymbol{\theta}}\phi(t;\boldsymbol{\theta}_0) \tag{6.7}$$

and

$$\mathrm{E}\left\{\frac{\partial}{\partial\boldsymbol{\theta}}\bar{\epsilon}(t, X_j; \boldsymbol{\theta}_0)\right\} = -\frac{\partial}{\partial\boldsymbol{\theta}}\bar{\phi}(t;\boldsymbol{\theta}_0). \tag{6.8}$$

Hence

$$\begin{aligned}
&\int \mathbf{s}_{21}(t)\mathbf{s}_{11}^{-1}(t)\mathbf{s}_{12}(t)\,dG_1(t) \\
&= \int \pi(t)\frac{\partial\phi(t;\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\,dt \\
&= \frac{\partial}{\partial\boldsymbol{\theta}}\left\{\int\int \pi(t) f(x;\boldsymbol{\theta})e^{itx}\,dx\,dt\right\}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\
&= \frac{\partial}{\partial\boldsymbol{\theta}}\left\{\int \frac{\partial \log f(x;\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} f(x;\boldsymbol{\theta})\,dx\right\}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\
&= \int \left\{\frac{\partial \log f(x;\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\right\}^2 f(x;\boldsymbol{\theta}_0)\,dx\,dy \\
&= \mathrm{E}\left\{\frac{\partial \log f(X_j;\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\right\}^2.
\end{aligned} \tag{6.9}$$

It follows from (2.6), (2.7), (6.6)–(6.8) that

$$\begin{cases} \mathbf{s}_{21}(t)\mathbf{s}_{11}^{-1}(t)\mathbf{M}G_1'(t) = \pi(t)(-1,0), \\ \mathbf{M}^{\mathrm{T}}\mathbf{s}_{11}^{-1}(t)\mathbf{s}_{12}(t)G_1'(t) = \pi(t)(-1,0)^{\mathrm{T}}. \end{cases} \tag{6.10}$$

It is easy to check that

$$\begin{aligned}
&\mathrm{E}\{\epsilon(t_1, X_j;\boldsymbol{\theta}_0)\epsilon(t_2, X_j;\boldsymbol{\theta}_0)\} \\
&\quad = \phi(t_1+t_2;\boldsymbol{\theta}_0) - \phi(t_1;\boldsymbol{\theta}_0)\phi(t_2;\boldsymbol{\theta}_0) \\
&\quad := a_{11}(t_1,t_2),
\end{aligned} \tag{6.11}$$

$$\begin{aligned}
&\mathrm{E}\{\epsilon(t_1, X_j;\boldsymbol{\theta}_0)\bar{\epsilon}(t_2, , X_j;\boldsymbol{\theta}_0)\} \\
&\quad = \phi(t_1-t_2;\boldsymbol{\theta}_0) - \phi(t_1;\boldsymbol{\theta}_0)\bar{\phi}(t_2;\boldsymbol{\theta}_0) \\
&\quad := a_{12}(t_1,t_2),
\end{aligned} \tag{6.12}$$

$$\begin{aligned}
&\mathrm{E}\{\bar{\epsilon}(t_1, X_j;\boldsymbol{\theta}_0)\epsilon(t_2, X_j;\boldsymbol{\theta}_0)\} \\
&\quad = \phi(-t_1+t_2;\boldsymbol{\theta}_0) - \bar{\phi}(t_1;\boldsymbol{\theta}_0)\phi(t_2;\boldsymbol{\theta}_0) \\
&\quad := a_{21}(t_1,t_2),
\end{aligned} \tag{6.13}$$

$$\begin{aligned}
&\mathrm{E}\{\bar{\epsilon}(t_1, X_j;\boldsymbol{\theta}_0)\bar{\epsilon}(t_2, X_j;\boldsymbol{\theta}_0)\} \\
&\quad = \phi(-t_1-t_2;\boldsymbol{\theta}_0) - \bar{\phi}(t_1;\boldsymbol{\theta}_0)\bar{\phi}(t_2;\boldsymbol{\theta}_0) \\
&\quad := a_{22}(t_1,t_2),
\end{aligned} \tag{6.14}$$

and

$$\boldsymbol{\Gamma}(t_1,t_2) = \mathbf{M}\begin{pmatrix} a_{11}(t_1,t_2) & a_{12}(t_1,t_2) \\ a_{21}(t_1,t_2) & a_{22}(t_1,t_2) \end{pmatrix}\mathbf{M}^{\mathrm{T}}. \tag{6.15}$$

By (6.11)–(6.15), we have

$$\int\int \mathbf{s}_{21}(t_1)\mathbf{s}_{11}^{-1}(t_1)\boldsymbol{\Gamma}(t_1,t_2)\mathbf{s}_{11}^{-1}(t_2)\mathbf{s}_{12}(t_2)\,dG_1(t_1)\,dG_1(t_2) = \int\int a_{11}(t_1,t_2)\pi(t_1)\pi(t_2)\,dt_1\,dt_2. \tag{6.16}$$

It is straightforward to check that

$$\begin{aligned}
&\int \phi(t_1+t_2;\boldsymbol{\theta}_0)\pi(t_1)\,dt_1 \\
&= \frac{1}{2\pi}\int\int \phi(t_1+t_2;\boldsymbol{\theta}_0)\frac{\partial \log f(x;\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\exp\{-ixt_1\}\,dx\,dt_1 \\
&= \int f(x;\boldsymbol{\theta}_0)\exp\{it_2x\}\frac{\partial \log f(x;\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\,dx
\end{aligned} \tag{6.17}$$

and

$$\begin{aligned}
&\int \phi(t_1;\boldsymbol{\theta}_0)\phi(t_2;\boldsymbol{\theta}_0)\pi(t_1)\,dt_1 \\
&= \phi(t_2;\boldsymbol{\theta}_0)\frac{1}{2\pi} \times \int\int \phi(t_1;\boldsymbol{\theta}_0)\frac{\partial \log f(x;\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\exp\{-ixt_1\}\,dx\,dt_1 \\
&= \phi(t_2;\boldsymbol{\theta}_0)\int f(x;\boldsymbol{\theta}_0)\frac{\partial \log f(x;\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\,dx \\
&= 0.
\end{aligned} \tag{6.18}$$

Hence (2.8) follows from (6.16) and (6.18).

Proof of Theorem 3. It follows from standard arguments in empirical likelihood methods and Theorem 2 that

$$\begin{aligned}
T_1(\hat{\boldsymbol{\theta}}) &= \int_{-a}^{a}\Bigg\{2n\boldsymbol{\lambda}_1^{\mathrm{T}}(t;\hat{\boldsymbol{\theta}})\mathbf{Q}_{1n}(t;\hat{\boldsymbol{\theta}},0) \\
&\qquad - \boldsymbol{\lambda}_1^{\mathrm{T}}(t;\hat{\boldsymbol{\theta}})\left\{\sum_{j=1}^{n}\mathbf{Y}_j(t;\hat{\boldsymbol{\theta}})\mathbf{Y}_j^{\mathrm{T}}(t;\hat{\boldsymbol{\theta}})\right\}\boldsymbol{\lambda}_1(t;\hat{\boldsymbol{\theta}})\Bigg\}\,dG_1(t) \\
&\qquad + o_p(1) \\
&= \int_{-a}^{a}\Bigg\{2n\boldsymbol{\lambda}_1^{\mathrm{T}}(t;\hat{\boldsymbol{\theta}})\mathbf{Q}_{1n}(t;\boldsymbol{\theta}_0,0) \\
&\qquad + 2n\boldsymbol{\lambda}_1^{\mathrm{T}}(t;\hat{\boldsymbol{\theta}})\frac{\partial\mathbf{Q}_{1n}(t;\boldsymbol{\theta}_0,0)}{\partial\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) \\
&\qquad + n\boldsymbol{\lambda}_1^{\mathrm{T}}(t;\hat{\boldsymbol{\theta}})\mathbf{s}_{11}(t)\boldsymbol{\lambda}_1(t;\hat{\boldsymbol{\theta}})\Bigg\}\,dG_1(t) + o_p(1) \\
&= \int_{-a}^{a}\{-n\mathbf{Q}_{1n}^{\mathrm{T}}(t;\boldsymbol{\theta}_0,0)\mathbf{s}_{11}^{-1}(t)\mathbf{Q}_{1n}(t;\boldsymbol{\theta}_0,0) \\
&\qquad - n(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)^{\mathrm{T}}\mathbf{s}_{21}(t)\mathbf{s}_{11}^{-1}(t)\mathbf{Q}_{1n}(t;\boldsymbol{\theta}_0,0)\}\,dG_1(t) \\
&\qquad + o_p(1).
\end{aligned} \tag{6.19}$$

Define $F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \le x)$ and $\alpha_n(x) = \sqrt{n}\{F_n(x) - F(x)\}$. Then,

$$\sup_{x\in R}\bigl|\alpha_n(x) - B_n(F(x))\bigr| = O\bigl(n^{-1/2}\log n\bigr) \quad \text{a.s.} \tag{6.20}$$

(see Kómlos, Major, and Túsnady 1975). By (6.20),

$$\begin{aligned}
&\sqrt{n}\int_{-a}^{a} \mathbf{s}_{21}(t)\mathbf{s}_{11}^{-1}(t)\mathbf{Q}_{1n}(t;\boldsymbol{\theta}_0, 0)\,dG_1(t) \\
&\quad = \int_{-\infty}^{\infty}\left\{\int_{-a}^{a} \mathbf{s}_{21}(t)\mathbf{s}_{11}^{-1}(t)(\cos(tx), \sin(tx))^{\mathrm{T}}\,dG_1(t)\right\} d\alpha_n(x) \\
&\quad = \int_{-\infty}^{\infty}\left\{\int_{-a}^{a} \mathbf{s}_{21}(t)\mathbf{s}_{11}^{-1}(t)\right. \\
&\qquad \left. \times (\cos(tx), \sin(tx))^{\mathrm{T}}\,dG_1(t)\right\} dB_n(F(x)) + o_p(1)
\end{aligned} \tag{6.21}$$

and

$$\begin{aligned}
&n\int_{-a}^{a} \mathbf{Q}_{1n}^{\mathrm{T}}(t;\boldsymbol{\theta}_0, 0)\mathbf{s}_{11}^{-1}(t)\mathbf{Q}_{1n}(t;\boldsymbol{\theta}_0, 0)\,dG_1(t) \\
&\quad = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left\{\int_{-a}^{a}(\cos(tx), \sin(tx))\mathbf{s}_{11}^{-1}(t)\right. \\
&\qquad \left. \times (\cos(tx), \sin(tx))^{\mathrm{T}}\,dG_1(t)\right\} d\alpha_n(x)\,d\alpha_n(y) \\
&\quad = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left\{\int_{-a}^{a}(\cos(tx), \sin(tx))\mathbf{s}_{11}^{-1}(t)\right. \\
&\qquad \left. \times (\cos(ty), \sin(ty))^{\mathrm{T}}\,dG_1(t)\right\} dB_n(F(x))\,dB_n(F(y)) \\
&\qquad + o_p(1).
\end{aligned} \tag{6.22}$$

Hence, the theorem follows from (6.19), (6.21), and (6.22).

Proof of Theorem 4. By (6.20), we have

$$\sqrt{n}\{\hat{\mu} - \mu_0\} = \int_{-\infty}^{\infty} x\,dB_n(F(x)) + o_p(1). \tag{6.23}$$

By standard arguments in empirical likelihood methods, it can be shown that

$$\sup_{-a\le t\le a}\left| l_2(t;\hat{\mu}) - \frac{\{\sum_{j=1}^{n}\sin(t(X_j - \hat{\mu}))\}^2}{\sum_{j=1}^{n}\{\sin(t(X_j - \hat{\mu}))\}^2}\right| = o_p(1). \tag{6.24}$$

Write

$$\begin{aligned}
&\sin(t(X_j - \hat{\mu})) \\
&\quad = \sin(t(X_j - \mu_0)) \\
&\qquad - \cos(t(X_j - \mu_0))t(\hat{\mu} - \mu_0)(1 + o_p(1)).
\end{aligned} \tag{6.25}$$

Hence, Theorem 4 follows from (6.23), (6.24), (6.25), and (6.20).

[Received June 2008. Revised February 2009.]

REFERENCES

Barndorff-Nielsen, O. E., Cox, D. R., and Klüppelberg, C. (2001), *Complex Stochastic Systems*, New York: Chapman & Hall.

Bilodeau, M., and de Micheaux, P. L. (2005), "A Multivariate Empirical Characteristic Function Test of Independence With Normal Marginals," *Journal of Multivariate Analysis*, 95, 345–369.

Cont, R., and Tankov, P. (2004), *Financial Modeling With Jump Processes*, New York: Chapman & Hall/CRC.

Einmahl, J. H. J., and McKeague, I. W. (2003), "Empirical Likelihood Based Hypothesis Testing," *Bernoulli*, 9, 267–290.

Epps, T. W. (1983), "A Test for Normality Based on the Empirical Characteristic Function," *Biometrika*, 70, 723–726.

Fan, Y. (1997), "Goodness-of-Fit Tests for a Multivariate Distribution by the Empirical Characteristic Function," *Journal of Multivariate Analysis*, 62, 36–63.

Fan, Y., and Gencay, R. (1995), "A Consistent Nonparametric Test of Symmetry in Linear Regression Models," *Journal of the American Statistical Association*, 90, 551–557.

Feuerverger, A., and McDunnough, P. (1981a), "On the Efficiency of Empirical Characteristic Function Procedures," *Journal of the Royal Statistical Society, Ser. B*, 43, 20–27.

——— (1981b), "On Some Fourier Methods for Inference," *Journal of the American Statistical Association*, 76, 379–387.

Feuerverger, A., and Mureika R. (1977), "The Empirical Characteristic Function and Its Applications," *The Annals of Statistics*, 5, 88–97.

Heathcote, C. R. (1977), "The Integrated Squared Error Estimation of Parameters," *Biometrika*, 64, 255–264.

Henze, N., Klar, B., and Meintanis, S. G. (2003), "Invariant Tests for Symmetry About an Unspecified Point Based on the Empirical Characteristic Function," *Journal of Multivariate Analysis*, 87, 275–297.

Jiang, G. J., and Knight, J. L. (2002), "Estimation of Continuous-Time Processes via the Empirical Characteristic Function," *Journal of Business & Economic Statistics*, 20, 198–212.

Kitamura, Y. (1997), "Empirical Likelihood Methods With Weakly Dependent Processes," *The Annals of Statistics*, 25, 2084–2102.

Klar, B., and Meintanis, S. G. (2005), "Tests for Normal Mixtures Based on the Empirical Characteristic Function," *Computational Statistics & Data Analysis*, 49, 227–242.

Kómlos, J., Major, P., and Túsnady, G. (1975), "An Approximation of Partial Sums of Independent Random Variables and the Sample Distribution Function," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 32, 111–131.

Koutrouvelis, I. A. (1980), "A Goodness-of-Fit Test of Simple Hypotheses Based on the Empirical Characteristic Function," *Biometrika*, 67, 238–240.

Koutrouvelis, I. A., and Kellermeier, J. (1981), "A Goodness-of-Fit Test Based on the Empirical Characteristic Function When Parameters Must Be Estimated," *Journal of the Royal Statistical Society, Ser. B*, 43, 173–176.

Li, G. (2003), "Nonparametric Likelihood Ratio Goodness-of-Fit Tests for Survival Data," *Journal of Multivariate Analysis*, 86, 166–182.

Madan, D. B., and Seneta, E. (1987), "Chebyshev Polynomial Approximations and Characteristic Function Estimation," *Journal of the Royal Statistical Society, Ser. B*, 49, 163–169.

Madan, D. B., Carr, P. P., and Chang, E. C. (1998), "The Variance Gamma Process and Option Pricing," *European Finance Review*, 2, 79–105.

Marcus, M. B. (1981), "Weak Convergence of the Empirical Characteristic Function," *The Annals of Probability*, 9, 194–201.

Meintanis, S. G. (2005), "Consistent Tests for Symmetric Stability With Finite Mean Based on the Empirical Characteristic Function," *Journal of Statistical Planning and Inference*, 128, 373–380.

Merton, R. (1976), "Option Pricing When the Underlying Stock Returns Are Discontinuous," *Journal of Financial Economics*, 3, 125–144.

Owen, A. (1988), "Empirical Likelihood Ratio Confidence Intervals for a Single Functional," *Biometrika*, 75, 237–249.

Qin, J., and Lawless, J. (1994), "Empirical Likelihood and General Estimating Equations," *The Annals of Statistics*, 22, 300–326.

Sato, K.-I. (1999), *Lévy Processes and Infinite Divisibility*, Cambridge: Cambridge University Press.

Singleton, K. J. (2001), "Estimation of Affine Asset Pricing Models Using the Empirical Characteristic Function," *Journal of Econometrics*, 102, 111–141.

Wells, M. T. (1992), "The Weak Approximation of the Empirical Characteristic Function Process When Parameters Are Estimated," *Stochastic Processes and Their Applications*, 40, 83–102.

Yu, J. (2004), "Empirical Characteristic Function Estimation and Its Applications," *Econometric Reviews*, 23, 93–123.

Yukich, J. E. (1985), "Weak Convergence of the Empirical Characteristic Function," *Proceedings of the American Mathematical Society*, 95, 470–473.