CIRJE-F-272

Empirical Likelihood Estimation of Lévy Processes

Naoto Kunitomo
University of Tokyo

Takashi Owada
Bank of Japan

April 2004
Revised: March 2005

Empirical Likelihood Estimation of Lévy Processes *

Naoto Kunitomo[†]
and
Takashi Owada [‡]

May 2004 (1st Version)
February 2005 (Revised Version)

Abstract

We propose a new parameter estimation procedure for the Lévy processes and the class of infinitely divisible distribution. We shall show that the empirical likelihood method gives an easy way to estimate the key parameters of the infinitely divisible distributions including the class of stable distributions as a special case. The maximum empirical likelihood estimator by using the empirical characteristic functions gives the consistency, the asymptotic normality, and the asymptotic efficiency for the key parameters when the number of restrictions on the empirical characteristic functions is large. Test procedures can be also developed. Some extensions to the estimating equations problem with the infinitely divisible distributions are discussed.

Key Words

Stable Distribution, Lévy processes, Infinitely Divisible Distributions, Maximum Empirical Likelihood, Characteristic Function, Estimating Equation Methods.

AMS Subject Classification: Primary 60G51, 60G52; secondary 62F12.

*This paper is a revised version of Discussion Paper CIRJE-F-272 (2004), Graduate School of Economics, University of Tokyo. We thank Professor Y. Miyahara of Nagoya-City University, the associate editor and the referee of this journal for pointing out several problems contained in the original paper which have led us to revise it significantly.

†Graduate School of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, JAPAN.

‡Currently at Bank of Japan.

1. Introduction

There have been growing interests on the applications of the Lévy processes and the class of infinitely divisible distributions in several research fields including financial economics. One interesting class of infinitely divisible distributions is the class of stable distributions. Since they are important classes of probability distributions, there have been extensive studies by mathematicians in probability over several decades. See Feller (1971), Zolotarev (1986), and Sato (1999) for the details of related problems in the probability literatures. Several statistical applications of stable distributions have been applied for modeling the fat-tail phenomena sometimes observed in financial economics and other applied areas of statistics. See Mandelbrot (1963), Paulson et. al. (1975), and Nolan (2001) for the early studies of the subject in the statistics literatures. More recently, some applications of the more general Lévy processes and other class of infinitely divisible distributions have been reported in the analyses of financial data. See Bandorff-Nielsen et. al. (2001) and Carr et. al. (2002) for recent examples.

Several estimation methods for the key parameters of stable distributions have been proposed and developed over the past few decades. DuMouchel (1971) has investigated the parametric maximum likelihood estimation method and Nolan (1997) has extended a numerical algorithm of the likelihood evaluation. Since it is not possible to obtain any explicit form of the likelihood function for stable distributions except very special cases, Fama and Roll (1968, 1971) proposed a practical estimation method based on the percentiles of empirical distributions and later MuCulloch (1986) has improved their method. Also another estimation method based on the empirical characteristic function was originally proposed by Press (1972), and there have been several related studies by Paulson et. al. (1975), Koutrouvelis (1980), Kogon and Williams (1998), Feuerverger and McDunnough (1981a, 1981b). These estimation methods could be extended to the more general Lévy processes.

The main purpose of this paper is to develop a new parameter estimation procedure for the Lévy processes and some classes of infinitely divisible distributions based on the empirical likelihood approach. The empirical likelihood method was originally proposed by Owen (1988, 1990) for constructing nonparametric confidence intervals and later it has been extended to the estimating equations problem by Qin and Lawless (1994). In this paper first we shall show that we can apply the empirical likelihood approach to the estimation problem of unknown parameters for stable distributions and the resulting computational burden is not heavy. In particular, the maximum empirical likelihood (MEL) estimator for the parameters of stable distribution we are proposing has some desirable asymptotic properties; it has the consistency, the asymptotic normality, and the asymptotic efficiency when the number of restrictions on the empirical characteristic function is large under a set of general conditions. Also it is possible to develop the empirical likelihood ratio statistics for the parameters of stable distributions which have the desirable asymptotic property.

More importantly, it is rather straightforward to extend our estimation method for the unknown parameters of stable distributions to the estimation of Lévy processes and infinitely divisible distributions, and also to the estimating equations problem with stable disturbances and other infinitely divisible disturbances. We shall show that it is possible to estimate both the parameters of equations and the parameters of stable distributions (or some infinitely divisible distributions) for disturbances at the same time by our method. It seems that it is not easy to solve this estimation problem by the conventional methods proposed in the past studies and in this sense our estimation method has some advantage over other methods. For the estimating equations problem, Qin and Lawless (1994) have shown some asymptotic properties of the MEL estimator and Kitamura et. al. (2001) have extended their results to one direction. In this respect our study has some technical novelty because we are considering the case when the number of restrictions grows with the sample size. Hence this paper can be regarded as an extension of Qin and Lawless (1994) in another important direction. Our formulation of the empirical likelihood method with many restrictions is closely related to a recent unpublished work by Hjort et. al. (2004). Also the studies by Fan et. al. (2001), Shen et. al. (1999), and Zhang and Gibels (2003) on the hypotheses testing problems have discussed the related problems.

In Section 2, we formulate the empirical likelihood estimation method of the class of stable distributions in the standard situation and state our main results on the asymptotic properties of the MEL estimator and the related testing procedure. Then in Section 3, we discuss the estimation problem of the Lévy processes and several infinitely divisible distributions, and an extension to the estimating equations problem when the disturbance terms follow the stable distribution (or some infinitely divisible distributions). In Section 4, we report some simulation results and in Section 5 we give some concluding remarks. The proofs of main results are given in Mathematical Appendix.

2. Empirical Likelihood Estimation of Stable Distributions

In this section we first consider the situation when X_i $(i = 1, \ldots, n)$ are a sequence of independently and identically distributed random variables and they follow the class of stable distributions. Let the characteristic function of X_i be denoted by $\phi_\theta(t)$, and its real part and imaginary part be $\phi_\theta^R(t)$ and $\phi_\theta^I(t)$, respectively. We adopt the formulation of the characteristic function used by Chamber et. al. (1976) for the class of stable distribution and it is represented as

$$\phi_\theta(t) = \phi_\theta^R(t) + i\phi_\theta^I(t) \ , \tag{2.1}$$

where

$$\begin{aligned} \phi_\theta^R(t) &= e^{-|\gamma t|^\alpha} \cos[\delta t + \beta\gamma t(|\gamma t|^{\alpha-1} - 1)\tan\tfrac{\pi\alpha}{2}] \ , \\ \phi_\theta^I(t) &= e^{-|\gamma t|^\alpha} \sin[\delta t + \beta\gamma t(|\gamma t|^{\alpha-1} - 1)\tan\tfrac{\pi\alpha}{2}] \end{aligned}$$

and the parameter space is given by

$$\boldsymbol{\Theta} = \{0 < \alpha \le 2 \ , -1 \le \beta \le 1 \ , \gamma > 0 \ , \delta \in \mathbf{R}\} \ .$$

In the following analysis we denote the vector of unknown parameters $\boldsymbol{\theta} = (\alpha, \beta, \gamma, \delta)'$ and the stable distribution associated with $\boldsymbol{\theta}$ as $F_\theta(\cdot)$.

There are two non-standard problems in the estimation of the vector of unknown parameters $\boldsymbol{\theta}$. It has been well-known in probability theory that except some special cases (the normal distribution, the Cauchy distribution, and a Lévy distribution) we do not have a simple explicit form of the probability density function and distribution function. This makes some difficulty of the direct estimation of unknown parameters including the parametric maximum likelihood method. Also since the stable distributions do not necessarily have the first and/or second moments, some of the standard techniques in the statistical asymptotic theory cannot be directly applicable.

2.1 Empirical Likelihood Method

In order to estimate the unknown parameters of the stable distributions, we are proposing to use the empirical likelihood approach, which is similar to the one developed by Qin and Lawless (1994). Because the stable distributions do not necessarily have the first and second moments, we cannot utilize the moments of distributions. However, we can use the information from the empirical characteristic functions instead. We define the empirical likelihood function by

$$L_n(F_\theta) = \prod_{k=1}^{n} (F_\theta(X_k) - F_\theta(X_k-)) = \prod_{k=1}^{n} p_k \ , \tag{2.2}$$

where $F_\theta(\cdot)$ is the distribution function and p_k $(k = 1, \cdots, n)$ are the probability assigned to the data points of X_k. Without any further restrictions except $p_k \geq 0$ and $\sum_{k=1}^n p_k = 1$, the empirical likelihood function $L_n(F_\theta)$ can be maximized at $p_k = 1/n$ $(k = 1, \cdots, n)$. Let the empirical likelihood ratio function be

$$R_n(F_\theta) = \prod_{k=1}^{n} np_k \ . \tag{2.3}$$

Then we define the maximum empirical likelihood estimator $\hat{\boldsymbol{\theta}}_n$ for the vector of unknown coefficients by maximizing the function $R_n(F_\theta)$ under the restrictions

$$\begin{aligned}\mathcal{P}_n = \Big\{ \ p_k \geq 0 \ (k = 1, \ldots, n) \ , \sum_{k=1}^{n} p_k = 1 \ , \sum_{k=1}^{n} p_k \left(\cos(t_l X_k) - \phi_\theta^R(t_l)\right) = 0, \\ \sum_{k=1}^{n} p_k \left(\sin(t_l X_k) - \phi_\theta^I(t_l)\right) = 0 \ (l = 1, \ldots, m) \ \Big\} \ .\end{aligned}$$

In the above restrictions m is the number of restrictions on the empirical characteristic functions and we take $m \geq 2$ and two terms $\sum_{k=1}^n p_k \cos(t_l X_k)$ and $\sum_{k=1}^n p_k \sin(t_l X_k)$ are the real part and the imaginary part of the empirical characteristic function evaluated at m different points $t = t_l$ $(t_1 < t_2 < \cdots < t_m; l = 1, \cdots, m)$. The choice of m is important and it can be dependent on the sample size n, but we shall discuss this problem later.

Denote a $2m \times 1$ vector

$$\mathbf{g}(X_k, \boldsymbol{\theta}) = \left(\mathbf{g}^R(X_k, \boldsymbol{\theta})', \mathbf{g}^I(X_k, \boldsymbol{\theta})'\right)' \ , \tag{2.4}$$

where

$$\begin{aligned}\mathbf{g}^R(X_k, \boldsymbol{\theta}) &= \left(\cos(t_1 X_k) - \phi_\theta^R(t_1), \ldots, \cos(t_m X_k) - \phi_\theta^R(t_m)\right)' \ , \\ \mathbf{g}^I(X_k, \boldsymbol{\theta}) &= \left(\sin(t_1 X_k) - \phi_\theta^I(t_1), \ldots, \sin(t_m X_k) - \phi_\theta^I(t_m)\right)' \ ,\end{aligned}$$

and $\phi_\theta^R(t_k)$ and $\phi_\theta^I(t_k)$ are given by (2.1) evaluated at $t = t_k$ $(k = 1, \cdots, m)$. Then we have the conditions $\mathbf{E}_{\theta_0}[\mathbf{g}(X, \boldsymbol{\theta}_0)] = \mathbf{0}$, where $\mathbf{E}_{\theta_0}(\cdot)$ is the expectation operator with respect to $F_{\theta_0}(\cdot)$ and $\boldsymbol{\theta}_0$ is the vector of true parameter values.

We suppose that the convex hull $\mathcal{P}_n(\boldsymbol{\theta}) = \{\sum_{k=1}^n p_k \mathbf{g}(X_k, \boldsymbol{\theta}) \mid p_k \geq 0, \sum_{k=1}^n p_k = 1\}$ contains $\mathbf{0}$ and set the Lagrange form as

$$L_n(\boldsymbol{\theta}) = \sum_{k=1}^{n} \log(np_k) - \mu[\sum_{k=1}^{n} p_k - 1] - n\boldsymbol{\lambda}'[\sum_{k=1}^{n} p_k \ \mathbf{g}(X_k, \boldsymbol{\theta})] \ , \tag{2.5}$$

where μ is a scalar Lagrange multiplier and $\boldsymbol{\lambda} = (\lambda_{11}, \ldots, \lambda_{1m}, \lambda_{21}, \ldots, \lambda_{2m})'$ is the $2m \times 1$ vector of Lagrange multipliers. By differentiating $L_n(\boldsymbol{\theta})$ with respect to p_k, we have ${p_k}^{-1} = \mu + n\boldsymbol{\lambda}^{'}\mathbf{g}(X_k, \boldsymbol{\theta})$ $(k = 1, \cdots, n)$. Then we have $\hat{\mu} = n$, and $\hat{p}_k = (1/n)[1 + \boldsymbol{\lambda}^{'}\mathbf{g}(X_k, \boldsymbol{\theta})]^{-1}$ and $\boldsymbol{\lambda} = \boldsymbol{\lambda}(\boldsymbol{\theta})$ is the solution of $\mathbf{0} = \sum_{k=1}^{n} \hat{p}_k \mathbf{g}(X_k, \boldsymbol{\theta})$. When a $2m \times 2m$ matrix $(1/n) \sum_{k=1}^{n} \mathbf{g}(X_k, \boldsymbol{\theta})\mathbf{g}(X_k, \boldsymbol{\theta})^{'}$ is positive definite and $\hat{p}_k \geq 0$, the matrix

$$\frac{\partial^2}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}^{'}} \left(-\frac{1}{n} \sum_{k=1}^{n} \log[1 + \boldsymbol{\lambda}^{'}\mathbf{g}(X_k, \boldsymbol{\theta})] \right) = \frac{1}{n} \sum_{k=1}^{n} \frac{\mathbf{g}(X_k, \boldsymbol{\theta})\mathbf{g}(X_k, \boldsymbol{\theta})^{'}}{\left[1 + \boldsymbol{\lambda}^{'}\mathbf{g}(X_k, \boldsymbol{\theta})\right]^2}$$

is also positive definite and $\boldsymbol{\lambda} = \boldsymbol{\lambda}(\boldsymbol{\theta})$ is the unique solution of

$$\text{(2.6)} \qquad \text{argmin}_{\lambda} \left\{ -\frac{1}{n} \sum_{k=1}^{n} \log \left[1 + \boldsymbol{\lambda}^{'}\mathbf{g}(X_k, \boldsymbol{\theta}) \right] \right\} .$$

Then we define the maximum empirical likelihood (MEL) estimator for the vector of unknown parameters $\boldsymbol{\theta}$ by maximizing the log-likelihood function $l_n(\boldsymbol{\theta})$, which is given by

$$\text{(2.7)} \qquad l_n(\boldsymbol{\theta}) = \log \prod_{k=1}^{n} n\hat{p}_k = -\sum_{k=1}^{n} \log \left[1 + \boldsymbol{\lambda}(\boldsymbol{\theta})^{'}\mathbf{g}(X_k, \boldsymbol{\theta}) \right] .$$

The numerical maximization in the MEL estimation is usually done by the two-step optimization procedure. (See Owen (2001) for the details.)

2.2 Asymptotic Properties of MEL estimation

In this subsection we shall report some asymptotic properties of the MEL estimator of $\boldsymbol{\theta}$. For the problem of the general estimating equations Qin and Lawless (1994) have proven the consistency and the asymptotic normality of the MEL estimator under a set of conditions. When the number of restrictions m is fixed, we have an analogous result in our situation.

Theorem 2.1 : Let $X_1, \ldots, X_n$ are i.i.d. random variables with the stable distribution $F_\theta(\cdot)$ and the vector of true parameters $\boldsymbol{\theta}_0 = (\alpha_0, \beta_0, \gamma_0, \delta_0)^{'}$ is in $\text{Int}(\boldsymbol{\Theta}_1)$, where $\boldsymbol{\Theta}_1 = \{(\alpha, \beta, \gamma, \delta) | \epsilon \leq \alpha \leq 1 - \epsilon, 1 + \epsilon \leq \alpha \leq 2, \ -1 \leq \beta \leq 1, \ \epsilon \leq \gamma \leq M, \ -M \leq \delta \leq M\}$ with ϵ being a sufficiently small positive number and M being a sufficiently large positive number. Let the MEL estimator be $\hat{\boldsymbol{\theta}}_n = \text{argmax}_\theta \mathcal{R}_n(\boldsymbol{\theta})$, where

$$\text{(2.8)} \qquad \mathcal{R}_n(\boldsymbol{\theta}) = \left\{ \prod_{k=1}^{n} np_k | \sum_{k=1}^{n} p_k \mathbf{g}(X_k, \boldsymbol{\theta}) = \mathbf{0}, \ p_k \geq 0, \ \sum_{k=1}^{n} p_k = 1 \right\} ,$$

a $2m \times 1$ vector of restrictions $\boldsymbol{g}(\cdot, \cdot)$ is defined by (2.4) and the Lagrange multiplier vector $\hat{\boldsymbol{\lambda}}_n$ is the solution of

$$\text{(2.9)} \qquad \frac{1}{n} \sum_{k=1}^{n} \frac{\mathbf{g}(X_k, \hat{\boldsymbol{\theta}}_n)}{1 + \hat{\boldsymbol{\lambda}}_n^{'}\mathbf{g}(X_k, \hat{\boldsymbol{\theta}}_n)} = \mathbf{0} .$$

Then as $n \to +\infty$

$$\text{(2.10)} \qquad \sqrt{n} \begin{bmatrix} \hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \\ \hat{\boldsymbol{\lambda}}_n \end{bmatrix} \xrightarrow{d} N_{4+2m} \left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Omega}_m & \mathbf{O} \\ \mathbf{O} & \boldsymbol{\Gamma}_m \end{pmatrix} \right] ,$$

where we define a $2m \times 1$ vector $\mathbf{\Phi}_\theta = \left(\phi_\theta^R(t_1), \ldots, \phi_\theta^R(t_m), \phi_\theta^I(t_1), \ldots, \phi_\theta^I(t_m)\right)^{'}$, a $2m \times 4$ matrix $\mathbf{B}_m(\boldsymbol{\theta}) = (\frac{\partial \Phi_{\theta,i}}{\partial \theta_j})$, a $2m \times 2m$ matrix $\mathbf{A}_m(\boldsymbol{\theta}) = \mathbf{E}_\theta\left[\mathbf{g}(X_1, \boldsymbol{\theta})\mathbf{g}(X_1, \boldsymbol{\theta})^{'}\right]$, and

$$
\begin{aligned}
\mathbf{\Omega}_m &= [\mathbf{B}_m^{'}(\boldsymbol{\theta}_0)\mathbf{A}_m(\boldsymbol{\theta}_0)^{-1}\mathbf{B}_m(\boldsymbol{\theta}_0)]^{-1}, \\
\mathbf{\Gamma}_m &= \mathbf{A}_m(\boldsymbol{\theta}_0)^{-1}[\mathbf{A}_m(\boldsymbol{\theta}_0) - \mathbf{B}_m(\boldsymbol{\theta}_0)\mathbf{\Omega}_m\mathbf{B}_m(\boldsymbol{\theta}_0)^{'}]\mathbf{A}_m(\boldsymbol{\theta}_0)^{-1}.
\end{aligned}
$$

The (i,j)th elements of $\mathbf{A}_m(\boldsymbol{\theta})$ are given by

$$
\begin{aligned}
&\tfrac{1}{2}\left\{\phi_\theta^R(t_i + t_j) + \phi_\theta^R(t_i - t_j)\right\} - \phi_\theta^R(t_i)\phi_\theta^R(t_j) \quad (1 \le i, j \le m)\ , \\
&\tfrac{1}{2}\left\{\phi_\theta^I(t_i + t_j) - \phi_\theta^I(t_i - t_j)\right\} - \phi_\theta^R(t_i)\phi_\theta^I(t_j) \quad (1 \le i \le m,\ m+1 \le j \le 2m)\ , \\
&\tfrac{1}{2}\left\{\phi_\theta^I(t_i + t_j) + \phi_\theta^I(t_i - t_j)\right\} - \phi_\theta^I(t_i)\phi_\theta^R(t_j) \quad (m+1 \le i \le 2m,\ 1 \le j \le m)\ , \\
&-\tfrac{1}{2}\left\{\phi_\theta^R(t_i + t_j) - \phi_\theta^R(t_i - t_j)\right\} - \phi_\theta^I(t_i)\phi_\theta^I(t_j) \quad (m+1 \le i, j \le 2m)\ ,
\end{aligned}
$$

respectively.

This result is based on Qin and Lawless (1994) (their *Lemma 1* and *Theorem 1*) and its proof is to check their sufficient conditions in our situation. We need some regularity conditions on the functions $\boldsymbol{g}(\cdot, \cdot)$ with respect to $\boldsymbol{\theta}$ and use a neighborhood $\mathbf{NB}(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$ with some smoothness conditions. But it is rather straightforward to verify these conditions in our situation. For instance, in our case we can utilize the bounded condition that for $\forall \boldsymbol{\theta} \in \mathbf{NB}(\boldsymbol{\theta}_0)$ we have

$$
\|\mathbf{g}(x, \boldsymbol{\theta})\| = \left[\sum_{l=1}^{m}\left(\cos(t_l x) - \phi_\theta^R(t_l)\right)^2 + \sum_{l=1}^{m}\left(\sin(t_l x) - \phi_\theta^I(t_l)\right)^2\right]^{1/2} \le 2\sqrt{2m}\ .
$$

Also it is possible to show directly that $\partial \boldsymbol{g}(x, \boldsymbol{\theta})/\partial\theta_j$ and $\partial^2 \boldsymbol{g}(x, \boldsymbol{\theta})/\partial\theta_j\partial\theta_k$ are continuous in $\mathbf{NB}(\boldsymbol{\theta}_0)$. Since we take a compact set $\mathbf{\Theta}_1(\cdot)$, both $\partial \boldsymbol{g}(x, \boldsymbol{\theta})/\partial\theta_j$ and $\partial^2 \boldsymbol{g}(x, \boldsymbol{\theta})/\partial\theta_j\partial\theta_k$ are bounded in $\mathbf{NB}(\boldsymbol{\theta}_0)$ ($\mathbf{NB}(\boldsymbol{\theta}_0) \subset \mathbf{\Theta}_1$).

Let the density function of stable distribution be $f_\theta(x)$ with the vector of unknown parameters $\boldsymbol{\theta} = (\alpha, \beta, \gamma, \delta)^{'}$. By using the similar arguments as DuMouchel (1973), we can show that $f_\theta(x)$ has the following properties :

(i) : For $x \in \mathbf{R}$, $f_\theta(x)$ as a function of $\boldsymbol{\theta}$ is continuous in $\mathrm{Int}(\mathbf{\Theta}_1)$ and for any $\boldsymbol{\theta} \in \mathrm{Int}(\mathbf{\Theta}_1)$ it is twice continuously differentiable.

(ii) : Since for any $\boldsymbol{\theta} \in \mathrm{Int}(\mathbf{\Theta}_1)$,

$$
\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{'}}\int_{-\infty}^{\infty} f_\theta(x)dx = \int_{-\infty}^{\infty}\frac{\partial^2 f_\theta(x)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{'}}dx\ , \tag{2.11}
$$

then $\mathbf{E}_\theta\left[\frac{\partial \log f_\theta(X)}{\partial\boldsymbol{\theta}}\right] = \mathbf{0}$ and

$$
\mathbf{I}(\boldsymbol{\theta}) = \mathbf{E}_\theta\left[\left(\frac{\partial \log f_\theta(X)}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial \log f_\theta(X)}{\partial\boldsymbol{\theta}}\right)^{'}\right] = -\mathbf{E}_\theta\left[\frac{\partial^2 \log f_\theta(X)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{'}}\right]\ . \tag{2.12}
$$

(iii) : For any $\boldsymbol{\theta} \in \mathrm{Int}(\mathbf{\Theta}_1)$, the Fisher Information matrix $\mathbf{I}(\boldsymbol{\theta})$ is nonsingular.

Under these conditions we shall consider the asymptotic efficiency of the MEL estimator. By using the notation $\boldsymbol{\Omega}_m$ in *Theorem 2.1*, it is possible to show that for any non-zero vector $\mathbf{u} \in \mathbf{R}^4$ we have the inequality

$$\mathbf{u}'\mathbf{I}(\boldsymbol{\theta}_0)^{-1}\mathbf{u} \le \mathbf{u}'\boldsymbol{\Omega}_m\mathbf{u} \ . \tag{2.13}$$

This implies that the asymptotic covariance of the MEL estimator in *Theorem 2.1* is larger than the Cramér-Rao lower-bound in general and it is asymptotically inefficient when the number of restrictions m is fixed.

However, it is possible to consider the situation when m is dependent on the sample size n . In particular, we take the case when $m = m(n) = [n^{\frac{1}{3}-\eta}]$ (or $m(n) = [n^{\frac{1}{6}-\eta}]$) where $[c]$ is the largest integer not exceeding c and η is any positive number with $0 < \eta < 1/3$ (or $0 < \eta < 1/6$) . Also in order to impose m restrictions in the form of (2.4) we set $t_l = Kl/m$ $(l = 1, 2, \cdots, m)$ with some positive constant K. Then we have the consistency, the asymptotic normality, and the asymptotic efficiency of the MEL estimator as stated in the next theorem. The proof is lengthy and given in Mathematical Appendix.

Theorem 2.2 : We assume that $X_1, \ldots, X_n$ are i.i.d. random variables with the stable distribution $F_\theta(\cdot)$ and the true parameter vector $\boldsymbol{\theta}_0$ is in Int($\boldsymbol{\Theta}_2$), where $\boldsymbol{\Theta}_2 = \{(\alpha, \beta, \gamma, \delta) | \epsilon \le \alpha \le 2,\ -1 \le \beta \le 1,\ \epsilon \le \gamma \le M,\ -M \le \delta \le M\}$ with ϵ being a sufficiently small positive number and M being a sufficiently large positive number. The $2m \times 1$ restriction functions $\boldsymbol{g}(\cdot, \cdot)$ are defined by (2.4) at $t_l = Kl/m$ $(l = 1, \cdots, m)$ with some positive constant K and we take $m = m(n) = [n^{\frac{1}{3}-\eta}]$ with $0 < \eta < 1/3$. Also we define $\hat{\boldsymbol{\theta}}_n = \text{argmax}_\theta \mathcal{R}_n(\boldsymbol{\theta})$ and

$$\mathcal{R}_n(\boldsymbol{\theta}) = \left\{ \prod_{k=1}^n np_k : \sum_{k=1}^n p_k \mathbf{g}(X_k, \boldsymbol{\theta}) = \mathbf{0},\ p_k \ge 0,\ \sum_{k=1}^n p_k = 1 \right\} \ .$$

Then

$$\hat{\boldsymbol{\theta}}_n \xrightarrow{p} \boldsymbol{\theta}_0 \ . \tag{2.14}$$

When we restrict the parameter space such that the vector of true parameter values $\boldsymbol{\theta}_0$ is in Int($\boldsymbol{\Theta}_1$) and $\boldsymbol{\Theta}_1$ is the same as in *Theorem 2.1*, then

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} N_4[\mathbf{0}, \mathbf{J}_K(\boldsymbol{\theta}_0)] \ \ , \tag{2.15}$$

and

$$\lim_{K \to +\infty} \mathbf{J}_K(\boldsymbol{\theta}_0) = \mathbf{I}(\boldsymbol{\theta}_0)^{-1},$$

where $\mathbf{J}_K(\boldsymbol{\theta}_0) = \lim_{n\to\infty} \boldsymbol{\Omega}_m(\boldsymbol{\theta}_0)$ and $\boldsymbol{\Omega}_m = [\mathbf{B}_m'(\boldsymbol{\theta}_0)\mathbf{A}_m(\boldsymbol{\theta}_0)^{-1}\mathbf{B}_m(\boldsymbol{\theta}_0)]^{-1}$ evaluated at the points $t_l = Kl/m$ $(l = 1, \cdots, m)$ with $K > 0$.

There are two important special cases to be mentioned. First, when $\alpha = 1$ (i.e. the Cauchy distribution) and $\beta \neq 0$, we have the situation that for any finite t we have

$$\lim_{\alpha \to 1} \left|\frac{\partial^2 \phi_\theta(t)}{\partial \alpha^2}\right| \to +\infty \tag{2.16}$$

and the convergence rate of $\hat{\alpha}_n$ (the estimator of α) to 1 could be different from $\sqrt{n}$. When $\alpha = 2$ and $\beta \neq 0$, we have the situation that for any finite t

$$\lim_{\alpha \to 2} \frac{\partial \phi_\theta(t)}{\partial \beta} = \lim_{\alpha \to 2} \frac{\partial^2 \phi_\theta(t)}{\partial \beta^2} = 0 \ . \tag{2.17}$$

Then the vector of unknown parameters $\boldsymbol{\theta}$ is unidentified and the limiting information matrix is degenerate. In these boundary cases it is not still clear if we have the asymptotic normality and the asymptotic efficiency of the MEL estimator.

2.3 Empirical Likelihood Testing

It is possible to develop the empirical likelihood ratio statistics and testing procedures for the parameters of stable distribution which have the desirable asymptotic properties as stated in the next theorem. The brief proof is given in Mathematical Appendix.

Theorem 2.3 : In addition to the assumptions of *Theorem 2.2*, we take $m = m(n) = [n^{\frac{1}{6}-\eta}]$ with $0 < \eta < 1/6$ and we restrict the parameter space such that the vector of true parameter values $\boldsymbol{\theta}_0$ is in $\mathrm{Int}(\boldsymbol{\Theta}_1)$ as in *Theorem 2.1.*
(i) The empirical likelihood ratio statistic for testing the hypothesis $\mathbf{H}_0 \ : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ is given by $W_1 = 2[l_n(\hat{\boldsymbol{\theta}}_n) - l_n(\boldsymbol{\theta}_0)]$, where the log-likelihood function $l_n(\boldsymbol{\theta})$ is given by (2.7). Then

$$W_1 \xrightarrow{d} \chi^2(4) \tag{2.18}$$

as $n \longrightarrow +\infty$ when $\mathbf{H}_0$ is true.
(ii) To test the hypothesis of the whole restrictions $\mathbf{E}_{\theta_0}[\mathbf{g}(X, \boldsymbol{\theta}_0)] = \mathbf{0}$, the likelihood ratio statistic is given by $W_2 = -2l_n(\hat{\boldsymbol{\theta}}_n)$. Then

$$\frac{W_2 - 2m}{\sqrt{4m}} \xrightarrow{d} N(0,1) \tag{2.19}$$

as $n \to +\infty$ when the $2m$ restrictions imposed are true.

The first part of *Theorem 2.3* allow us to use the empirical likelihood ratio statistic for testing the standard hypothesis $\mathbf{H}_0$ as well as constructing confidence sets for parameters of $\boldsymbol{\theta}$. The second part may not be standard in the statistics literature, but it corresponds to the testing problem of the overidentifying restrictions in the econometric literatures since the classical study on the simultaneous equations models by Anderson and Rubin (1950). Since the degrees of freedom $L = 2m - 4$ in the second case becomes large as $n \to +\infty$, we have the normal distribution as the limit [1] .

Although we have the χ^2 distribution or the normal distribution as the limiting distribution as stated in *Theorem 2.3*, the asymptotic distributions of the likelihood ratio statistics are not known when the true parameters are on some boundaries of the parameter space $\boldsymbol{\Theta}_1$. The main difficulty of this problem is the same we have mentioned in the previous subsection for the estimation of parameters.

[1] The referee has pointed out the possible relation between our results and the general Wilks phenomena for the generalized empirical likelihood ratio statistics in more general situations. The latter problem has been recently discussed by Fan et. al. (2001), Shen et. al. (1999) and Zhang and Gibels (2003) in the statistics literature.

3. Estimation of Lévy Processes and Estimating Equations Problem

3.1 Estimation of Lévy Processes

We consider the estimation problem of unknown parameters in the class of Lévy processes. For any one-dimensional Lévy process Z_v at a positive finite time v (> 0), it can be represented as the sum of i.i.d. random variables X_{v_i}. For the notational convenience we take $v_i - v_{i-1} = 1$ $(v_i = i; i = 0, 1, \cdots, n)$ and write $Z_n = \sum_{i=1}^{n} X_i$. Then it has been well-known that the one-dimentional Lévy processes $\{Z_v\}$ and the infinitely divisible distributions for the random variables $\{X_i\}$ are completely determined by the characteristic function

$$\phi_\theta(t) = \exp\left\{ ibt - \frac{a}{2}t^2 + \int_{\mathbf{R}} [e^{itx} - 1 - itx\mathbf{I}(|x| < 1)]\nu_c(dx) \right\} , \tag{3.1}$$

where b and a (≥ 0) are real constants, $\mathbf{I}(\cdot)$ is the indicator function, $\nu_c(\cdot)$ is the Lévy measure satisfying $\nu_c(\{0\}) = 0$,

$$\int_{|x|>0} [|x|^2 \wedge 1]\nu_c(dx) < +\infty \tag{3.2}$$

and $\mathbf{c}$ is the vector of some parameters. (See Sato (1999) for the details of the Lévy processes and the infinitely divisible distributions.) Then the vector of unknown parameters of the infinitely divisible distributions is represented as $\boldsymbol{\theta} = (a, b, \mathbf{c}^{'})^{'}$.

For applications, we mention only three important cases of the infinitely divisible distributions used in the recent financial economics and mathematical finance. First, the class of stable distributions with the condition $0 < \alpha < 2$ can be characterized by the Lévy measure

$$\nu_c(dx) = \begin{cases} \frac{c_1}{|x|^{1+\alpha}}dx & \text{for } x < 0 \\ \frac{c_2}{|x|^{1+\alpha}}dx & \text{for } x > 0 \end{cases} , \tag{3.3}$$

where $\mathbf{c} = (c_1, c_2, \alpha)^{'}$. We should note that although the parameterizations of c_1 (> 0) and c_2 (> 0) are different from the ones appeared in Section 2 there is one-to-one correspondence between the vectors $(\alpha, \beta, \gamma, \delta)$ in Section 2 and (α, b, c_1, c_2) (see Sato (1999) for the details).

The second case is the CGMY process introduced by Carr et. al. (2002), which has been applied to describe the stochastic processes for financial prices. The Lévy measure for this process has been given by

$$\nu_c(dx) = C_0\{\mathbf{I}(x < 0)e^{-G|x|} + \mathbf{I}(x > 0)e^{-M|x|}\}|x|^{-(1+Y)}dx , \tag{3.4}$$

where the vector of parameters $\mathbf{c} = (C_0, G, M, Y)^{'}$ satisfies the conditions of $C_0 > 0, G \geq 0, M \geq 0$, and $Y < 2$. The characteristic function is given by

$$\begin{aligned} \phi_\theta(t) &= \exp\{i[b + C_0\Gamma(-Y)Y(M^{Y-1} - G^{Y-1})]t \\ &\quad + C_0\Gamma(-Y)((M - it)^Y - M^Y + (G + it)^Y - G^Y)]\} , \end{aligned} \tag{3.5}$$

where the vector of parameters is given by $\boldsymbol{\theta} = (b, C_0, G, M, Y)^{'}$ and $\Gamma(\cdot)$ is the Gamma function.
When $Y = 0$, then the CGMY process is reduced to the Variance Gamma process proposed by Madan and Seneta (1990). Miyahara (2002) has summarized the basic properties of the CGMY process and the Variance Gamma process in a systematic way.
Although the characteristic function given by (3.5) is continuous with respect to $\boldsymbol{\theta}$, we can find that for any finite t

$$|\frac{\partial \phi_\theta(t)}{\partial Y}| \to +\infty$$

as $Y \to 0$ or $Y \to 1$. Hence we should be careful to treat these cases as we have discussed for the class of stable distributions in Section 2.

Third example is the class of normal inverse Gaussian processes, which has been introduced and discussed by Bandorff-Nielsen (1998). The characteristic function for this class of distributions is given by

$$\phi_\theta(t) = \exp\{\delta[\sqrt{\alpha^2-\beta^2} - \sqrt{\alpha^2-(\beta+it)^2}] + i\ \mu t\}\ , \tag{3.6}$$

and the vector of parameters is given by $\boldsymbol{\theta} = (\mu, \alpha, \beta, \delta)^{'}$ in the present case.

In these infinitely divisible distributions it is not possible to obtain the simple form of the density function and the parametric maximum likelihood estimation method has computational problems except very special cases. In this respect, the maximum empirical likelihood (MEL) method can be directly applicable and we can establish the next result. The proof is similar to that of *Theorem 2.2* and it is omitted.

Theorem 3.1 : We assume that $X_1, \ldots, X_n$ are i.i.d. random variables with the characteristic function given by (3.1), which is continuous with respect to $\boldsymbol{\theta}$, and the L*é*vy measure ν_c is absolutely continuous with respect to the Lebesgue measure. The true parameter vector $\boldsymbol{\theta}_0$ is in Int$(\boldsymbol{\Theta}_3)$, and $\boldsymbol{\Theta}_3$ is a compact subset such that (3.1) is the characteristic function of the infinitely divisible distribution with non-degenerate continuous density $f_\theta(\cdot)$. We impose $2m \times 1$ restriction functions $\mathbf{g}(\cdot,\cdot)$ defined by (2.4) at $t_l = Kl/m$ $(l = 1, \cdots, m)$ with some positive constant K for the real part $\phi_\theta^R(t)$ and the imaginary part $\phi_\theta^I(t)$ of $\phi_\theta(t)$ and take $m = m(n) = [n^{\frac{1}{3}-\eta}]$ with $0 < \eta < 1/3$. Also we set $\hat{\boldsymbol{\theta}}_n = \text{argmax}_\theta \mathcal{R}_n(\boldsymbol{\theta})$ and

$$\mathcal{R}_n(\theta) = \left\{ \prod_{k=1}^n np_k | \sum_{k=1}^n p_k \mathbf{g}(X_k, \boldsymbol{\theta}) = \mathbf{0},\ p_k \geq 0,\ \sum_{k=1}^n p_k = 1 \right\}\ .$$

Then

$$\hat{\boldsymbol{\theta}}_n \xrightarrow{p} \boldsymbol{\theta}_0\ . \tag{3.7}$$

Furthermore, we restrict the parameter space such that the true parameter value $\boldsymbol{\theta}_0$ is in Int$(\boldsymbol{\Theta}_4)$ and $\boldsymbol{\Theta}_4$ is a compact subset such that $\mathbf{I}(\boldsymbol{\theta}_0)$ (the Fisher information matrix) is positive definite. Assume that $\phi_\theta(t)$ are continuously twice-differentiable with respect to $\boldsymbol{\theta}$ and their derivatives are bounded by the integrable functions.
Then

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} N[\mathbf{0}, \mathbf{J}_K(\boldsymbol{\theta}_0)]\ \ , \tag{3.8}$$

where $\lim_{K\to+\infty}\mathbf{J}_K(\boldsymbol{\theta}_0)=\mathbf{I}(\boldsymbol{\theta}_0)^{-1}$ and $\mathbf{J}_K(\boldsymbol{\theta}_0)$ is defined as in *Theorem 2.2*.

There can be simpler regularity conditions for the results in *Theorem 3.1*. Since the Lévy measure is not necessarily a finite measure in the general case, however, a careful analysis would be needed to impose further conditions on $\nu_c(\cdot)$.

3.2 Estimating Equations Problem

We consider a single structural equation in the econometric model (or the estimating equation model in statistics) represented by

$$y_{1j}=h_1(\mathbf{y}_{2j},\mathbf{z}_{1j},\boldsymbol{\theta}_1)+u_j\ (j=1,\cdots,n)\ , \tag{3.9}$$

where $h_1(\cdot,\cdot,\cdot)$ is a function, y_{1j} and $\mathbf{y}_{2j}$ are 1×1 and $G_1\times 1$ (vector of) endogenous variables, $\mathbf{z}_{1j}$ is a $K_1^*\times 1$ vector of included exogenous variables, $\boldsymbol{\theta}_1=(\theta_{1k})$ is an $r\times 1$ vector of unknown parameters, and $\{u_j\}$ are mutually independent disturbance terms with the infinitely divisible distribution $F_{\theta_2}(\cdot)$ and $\boldsymbol{\theta}_2=(\theta_{2k})$ is the vector of its unknown parameters.

We assume that (3.9) is the first equation in a system of $(1+G_1)$ structural equations which relate the vector of $1+G_1$ endogenous variables $\mathbf{y}_j^{'}=(y_{1j}^{'},\mathbf{y}_{2j}^{'})$ to the vector of K^* $(=K_1^*+K_2^*)$ instrumental (or exogenous) variables $\mathbf{z}_j$ $(j=1,\cdots,n)$, which includes the vector of explanatory variables $\mathbf{z}_{1j}$ appeared in the structural equation of interest as (3.9). The set of explanatory (or exogenous) variables $\mathbf{z}_j$ are often called the instrumental variables in the econometric literatures. The restrictions we impose on the real part and the imaginary part of the characteristic function at any m different points of t $(t_1<\cdots<t_m)$ are given by

$$\mathbf{E}\left[\boldsymbol{h}_2(\mathbf{z}_j)(e^{itu_j}-\phi_{\theta_2}(t))\right]=\mathbf{0}\ \ (j=1,\cdots,n)\ , \tag{3.10}$$

where $\boldsymbol{h}_2(\cdot)$ is a set of l functions of instrumental variables $\mathbf{z}_j$ $(l\leq K^*)$ and $\boldsymbol{\theta}_2=(a,b,\mathbf{c}^{'})^{'}$ is the vector of unknown parameters of the infinitely divisible distributions for the disturbance terms $\{u_j\}$. Because we do not specify the structural equations except (3.9) and we only have the limited information on the set of instrumental variables (or instruments), we are considering the limited information estimation method [2] .

As an important application of the general methodology we are developing, we shall consider the estimating equation problem when the disturbance terms follow the class of stable distributions. Let $\boldsymbol{\theta}=(\boldsymbol{\theta}_1^{'},\boldsymbol{\theta}_2^{'})^{'}$ and $\boldsymbol{\theta}_1$ be the vector of unknown parameters in the estimating equations when the disturbance terms $\{u_j\}$ in (3.9) follow the stable distribution $F_{\theta_2}(\cdot)$ with the vector of parameters $\boldsymbol{\theta}_2=(\alpha,\beta,\gamma,\delta)^{'}$. The maximum empirical likelihood (MEL) estimator for the vector of unknown parameters $\boldsymbol{\theta}$ can be

[2] See Chapter 12 of Anderson (2003), and Anderson and Rubin (1949) on the classical linear formulation of the related problems and see Kunitomo and Matsushita (2003) for the finite sample properties of the MEL estimator in the simple linear structural equation.

defined by maximizing the Lagrange form

$$
\begin{aligned}
L_n^*(\boldsymbol{\lambda},\boldsymbol{\theta}) \;=\;& \sum_{j=1}^{n}\log(n\,p_j)-\mu[\sum_{j=1}^{n}p_j-1]-n\sum_{j=1}^{n}p_j\boldsymbol{h}_2^{'}(\mathbf{z}_j)\\
&\times\left\{\sum_{k=1}^{m}\boldsymbol{\lambda}_{1k}\left[\cos\left(t_k(y_{1j}-h_1(\mathbf{y}_{2j},\mathbf{z}_{1j},\boldsymbol{\theta}_1))\right)-\phi_{\theta_2}^{R}(t_k)\right]\right.\\
&\left.+\sum_{k=1}^{m}\boldsymbol{\lambda}_{2k}\left[\sin\left(t_k(y_{1j}-h_1(\mathbf{y}_{2j},\mathbf{z}_{1j},\boldsymbol{\theta}_1))\right)-\phi_{\theta_2}^{I}(t_k)\right]\right\}\ ,
\end{aligned}
\tag{3.11}
$$

where μ is a scalar Lagrange multiplier, $\boldsymbol{\lambda}_{ik}$ $(i=1,2;k=1,\cdots,m)$ are $l\times 1$ vectors of Lagrange multipliers, $\phi_{\theta_2}^{R}(t)$ and $\phi_{\theta_2}^{I}(t)$ are the real part and the imaginary part of $\phi_{\theta_2}(t)$ as (2.4), respectively, and p_j $(j=1,\cdots,n)$ are the weighted probability functions to be chosen. The above maximization problem is the same as to maximize

$$
\begin{aligned}
L_n(\boldsymbol{\lambda},\boldsymbol{\theta}) \;=\;& -\sum_{j=1}^{n}\log\left(1+\boldsymbol{h}_2^{'}(\mathbf{z}_j)\left\{\sum_{k=1}^{m}\boldsymbol{\lambda}_{1k}\left[\cos(t_k(y_{1j}-h_1(\mathbf{y}_{2j},\mathbf{z}_{1j},\boldsymbol{\theta}_1)))-\phi_{\theta_2}^{R}(t_k)\right]\right.\right.\\
&\left.\left.+\sum_{k=1}^{m}\boldsymbol{\lambda}_{2k}\left[\sin(t_k(y_{1j}-h_1(\mathbf{y}_{2j},\mathbf{z}_{1j},\theta_1)))-\phi_{\theta_2}^{I}(t_k)\right]\right\}\right),
\end{aligned}
\tag{3.12}
$$

where we have used the relations $\hat{\mu}=n$ and

$$
\begin{aligned}
[n\hat{p}_j]^{-1} \;=\;& 1+\boldsymbol{h}_2^{'}(\mathbf{z}_j)\left\{\sum_{k=1}^{m}\boldsymbol{\lambda}_{1k}\left[\cos(t_k(y_{1j}-h_1(\mathbf{y}_{2j},\mathbf{z}_{1j},\boldsymbol{\theta}_1)))-\phi_{\theta_2}^{R}(t_k)\right]\right.\\
&\left.+\sum_{k=1}^{m}\boldsymbol{\lambda}_{2k}\left[\sin(t_k(y_{1j}-h_1(\mathbf{y}_{2j},\mathbf{z}_{1j},\boldsymbol{\theta}_1)))-\phi_{\theta_2}^{I}(t_k)\right]\right\}\ .
\end{aligned}
\tag{3.13}
$$

By differentiating (3.12) with respect to $2lm\times 1$ vector $\boldsymbol{\lambda}^{'}=(\boldsymbol{\lambda}_{11}^{'},\boldsymbol{\lambda}_{21}^{'},\cdots,\boldsymbol{\lambda}_{1m}^{'},\boldsymbol{\lambda}_{2m}^{'})$ and combining the resulting equation with (3.13), we have the MEL estimator for the vector of parameters $\boldsymbol{\theta}$. Because we have $r+4$ parameters and the number of restrictions is $2lm$, the degrees of overidentifying restrictions is given by

$$
L=2lm-r-4\ , \tag{3.14}
$$

where we assume that $L>0$.

In our formulation of the present problem the restrictions of (2.4) in Section 2 can be interpreted as the simplest case of (3.10) in this section when $r=0$, $l=1$, and $h_2(x)=x$. Also if we set $y_j=y_{1j}$ (we do not have any $\mathbf{y}_{2j}$), $\mathbf{x}_j=\mathbf{z}_{1j}$ $(=\mathbf{z}_j)$ $(j=1,\cdots,n)$, and the vector of $\mathbf{x}_j$ are exogenous, then we have the nonlinear regression model with the infinitely divisible disturbances or the stable disturbances.

More generally, the estimation problem of structural equations have been discussed under the standard moment conditions on disturbance terms and the generalized moment method by Hansen (1982) or the estimating equation method by Godambe (1960). The standard statistical estimation methods have been usually applied [3] . By applying the similar arguments as in Section 2, it may be possible to establish the asymptotic results as *Theorem 2.2*, *Theorem 2.3* and *Theorem 3.1* in the general estimating equations problem under a set of regularity conditions.

[3] See Hayashi (2000), for the details of standard results in the recent econometrics literatures.

4. Simulation Results

In order to examine the actual performance of our estimation procedure, we have done a set of Monte Carlo simulations. In the first experiment we have fixed $\gamma = 1$ and $\delta = 0$, and simulated 1,000 random numbers which follow the stable distribution by using the method of Chamber, Mallows and Stuck (1976). After some experiments, we imposed the constraints on the empirical characteristic functions at the points $t = 0.1,\ 1.1,\ 2.1,\ 3.1,\ 4.1$. By using the restrictions at only these five points, we can get relatively accurate estimation results when the true parameter values are $(\alpha, \beta) \in (0,\ 1.8) \times [-1,\ 1]$. For the case of $(\alpha, \beta) \in [1.8,\ 2) \times [-1,\ 1]$, however, we sometimes have slow convergences when we had imposed the restrictions only at near to the origin as $t = 0.1$.

From our experiments, when we have fat tails in the empirical study of returns sometimes encountered in financial economics and the true value α is near to 2, it may be enough to use the restrictions on the empirical characteristic functions at $t = 0.6,\ 1.1,\ 2.1,\ 3.1,\ 4.1$. When $\gamma_0 \neq 1, \delta_0 \neq 0$, it is computationally efficient to use the iterative procedure as

1. First we obtain a preliminary estimate by using an estimation method as McCulloch (1986) and obtain $\hat{\gamma}^{(0)}, \hat{\delta}^{(0)}$.
2. Apply the empirical likelihood method to the standardized data $(x_1 - \hat{\delta}^{(0)})/\hat{\gamma}^{(0)}, \ldots, (x_n - \hat{\delta}^{(0)})/\hat{\gamma}^{(0)}$, and set $\hat{\gamma}^{(1)}, \hat{\delta}^{(1)}$.
3. We set $\hat{\gamma} = \hat{\gamma}^{(0)}\hat{\gamma}^{(1)},\ \hat{\delta} = \hat{\delta}^{(0)} + \hat{\delta}^{(1)}\hat{\gamma}^{(0)}$ for the final estimates of the parameters γ and δ .

Although in our experiments we have set the sample size $n = 1,000$, we can estimate the key parameters satisfactorily even when $n \geq 100$ by imposing the restrictions at only 5 points.

We repeated our simulations 500 times in each case and calculated the average, the maximum, the minimum, and RMSE as reported in Table 1. Then we have compared the sample variance with the asymptotic variance for the parametric maximum likelihood estimator which were obtained numerically by DuMouchel (1971) and Nolan (2001). We define the efficiency of our estimator as the ratio of the asymptotic variance calculated from the inverse of the Fisher information and the sample variance of estimator in our simulations. Then we have summarized our numerical results on efficiency in Table 2 and we found that there are not many extreme cases where we have low efficiency and our estimation method give reasonable values in most cases.

When $\alpha = 1$ and the parameter values are near to the boundaries, we have found that some instability in estimation occurs and the estimation results often depend on the choice of the initial conditions. When $\alpha = 2$ and $\beta \neq 0$, there is an identification problem and some instability in numerical computations would occur without any further restrictions on the parameter space.

As the second simulation we have examined the actual performance of the empirical likelihood estimation for the regression model with the class of stable disturbance terms, which is defined by

$$Y_j = \theta_1 X_j + u_j \ (j = 1, \cdots, n), \tag{4.1}$$

Table 1: Simulation Results of α

We set $\gamma = 1.0$ and $\delta = 0.0$ in our simulations. The values of average, maximum, minimum, and RMSE are calculated from the estimates for each coefficients.

(α, β)	Average	Max	Min	RMSE
(1.95,0.0)	1.9492	2.0855	1.7983	0.0460
(1.80,0.0)	1.8021	1.9638	1.6050	0.0618
(1.65,0.0)	1.6502	1.8155	1.4717	0.0619
(1.50,0.0)	1.5023	1.6718	1.3238	0.0592
(1.30,0.0)	1.3045	1.4521	1.1382	0.0534
(1.25,0.0)	1.2539	1.4151	1.1017	0.0521
(1.00,0.0)	1.0018	1.1301	0.8880	0.0436
(0.80,0.0)	0.8004	0.9159	0.6929	0.0365
(1.50,0.5)	1.5037	1.6575	1.3516	0.0601
(1.10,0.5)	1.1052	1.2382	0.9829	0.0455
(1.00,0.5)	1.0024	1.1662	0.8849	0.0406
(0.60,0.5)	0.6009	0.6850	0.5146	0.0275
(0.50,0.5)	0.4996	0.5819	0.4340	0.0240

Table 2: Efficiency

We set $\gamma = 1.0$ and $\delta = 0.0$ in our simulations. The values in Table 2 are the efficiencies as the ratio of the asymptotic variance and the sample variance for each coefficients α , β , γ , δ in simulations.

(α, β)	α	β	γ	δ
(1.65,0.0)	0.794	0.973	0.986	0.899
(1.50,0.0)	0.851	1.052	0.932	0.985
(1.30,0.0)	0.904	0.966	0.926	0.914
(1.25,0.0)	0.908	0.924	0.906	0.876
(1.00,0.0)	0.949	0.878	0.877	0.904
(0.80,0.0)	0.981	0.813	0.919	0.865
(1.50,0.5)	0.820	0.779	0.929	0.897
(1.10,0.5)	0.945	0.824	0.860	0.890
(1.00,0.5)	0.966	0.783	0.877	0.890
(0.60,0.5)	1.015	0.590	1.004	0.863
(0.50,0.5)	0.987	0.516	1.054	0.815

where θ_1 is the unknown coefficient, Y_j is the dependent variable, X_j is the explanatory variable, and u_j is the disturbance term with the stable distribution. We have set the parameter values $\beta = 0, \gamma = 1, \delta = 0$ in the class of stable distributions and simulated $\{X_j\}$ such that they are a sequence of i.i,.d. random variables which follow the log-normal distribution $LN(0, 1)$. We have repeated our simulations 500 times for the sample size n $(= 3,000)$ with $\theta_1 = 1.0$, and calculated the average, the maximum, the minimum, and the RMSE in Table 3.

When $\alpha = 1.5$ we also have calculated the standard least squares estimator for the coefficient parameter θ_1 . The average and its RMSE were 1.0015 and 0.0561, respectively, while the maximum and the minimum were 1.4499 and 0.5873, respectively. It seems that the RMSE of the least squares estimator is more than twice of the MEL estimator when $1 < \alpha < 2$ in our simulations. In addition to this favorable result on our estimation method, the least squares estimation often fails when $0 < \alpha < 1$ in our limited experiments. On the other hand, we did not have any convergence problem in the MEL estimation as long as we have enough data size in the simulations. The MEL estimation procedure for the regression model with the stable disturbances has reasonable performances in all cases of our simulations.

Table 3: Simulation Results for Regression

We set $\beta = \delta = 0.0$ and set $\theta = (\theta_1, \alpha, \gamma)'$ in our simulations. The values of average, maximum, minimum, and RMSE of the MEL estimates are calculated from the estimates for each coefficients.

$(\alpha, \gamma, \theta_1)$=(0.6,1.0,1.0)	α	γ	θ_1
Average	0.6010	0.9969	1.0009
RMSE	0.0157	0.0412	0.0115
Max	0.6451	1.1429	1.0354
Min	0.5525	0.8509	0.9673
$(\alpha, \gamma, \theta_1)$=(1.5,1.0,1.0)	α	γ	θ_1
Average	1.4998	1.0013	1.0010
RMSE	0.0329	0.0222	0.0192
Max	1.5933	1.0708	1.0539
Min	1.4126	0.9386	0.9373

5. Conclusions

This paper first develops a new parameter estimation method of stable distributions based on the empirical likelihood approach. We have shown that we can apply the empirical likelihood approach to the estimation problem of stable distributions and the computational burden is not heavy in comparison with the parametric maximum likelihood estimation. The maximum empirical likelihood (MEL) estimator for the parameters of stable distributions has some desirable asymptotic properties; it has the consistency, the asymptotic normality, and the asymptotic efficiency when the number of restrictions is large. Also it is possible to develop the empirical likelihood ratio statistics for the parameters of stable distributions which have the desirable asymptotic property such as the asymptotic χ^2-distribution. Also we can construct a test procedure for the null-hypothesis of restrictions imposed and it is the same as the test of overidentifying conditions in the econometrics literature.

Second, it is rather straightforward to extend our estimation method for unknown parameters of the stable distributions to the estimation of the general Lévy processes and the infinitely divisible distributions. It is also possible to apply it to the estimating equations problem with stable and infinitely divisible disturbance terms. We have shown that it is possible to estimate both the parameters of equations and the parameters of the distributions for disturbances at the same time by our method. It seems that it is not easy to solve this estimation problem by the conventional methods proposed in the past studies and in this sense our estimation method developed has some advantages over other methods.

Finally, we should mention that our estimation method is so simple that the results can be extended to several directions. One obvious direction is to extend our method to the multivariate infinitely divisible distributions and it is straightforward to do it for the class of symmetric stable distributions. Although we have assumed that X_k $(k = 1, \cdots, n)$ are a sequence of i.i.d. random variables in this paper, there are many interesting applications when they are dependent. Because there have been growing interests on the applications of the back-ground driving Lévy processes (BDLP) financial applications (Bandorff-Nielsen et. al. (2001)), these problems are under our current investigations

6. Mathematical Appendix

In this appendix we give the proofs of *Theorem 2.2* and *Theorem 2.3* in Section 2. In our proof we have utilized some ideas of Kitamura et. al. (2001) and Hjort et. al. (2004) although their problem and formulation are quite different from ours. We first show the consistency of the MEL estimator, and then prove its asymptotic normality and asymptotic efficiency when m is large. Then a *key* lemma needed for the proof of *Theorem 2.2* will be given and we utilize the proof of *Theorem 2.2* to show *Theorem 2.3*. In our proofs we take a compact set $\boldsymbol{\Theta}$ for the parameter space and we assume that the vector of true parameters $\boldsymbol{\theta}_0$ is in $\mathrm{Int}(\boldsymbol{\Theta})$ and each elements of $\mathbf{A}_m(\boldsymbol{\theta})$ and $\mathbf{B}_m(\boldsymbol{\theta})$ are bounded.

Proof of Theorem 2.2 :
[i] **Consistency**: We take a sufficiently large K (> 0) and set $m = m(n) = [n^{\frac{1}{3}-\epsilon}]$, $0 < \epsilon < 1/3$, $t_l = K(l/m)$ $(l = 1, \ldots, m)$ in (2.4). Define a criterion function by

$$G_n(\boldsymbol{\theta}) = -\frac{1}{n}\sum_{k=1}^{n} \log[1 + K\boldsymbol{\lambda}^{'}(\boldsymbol{\theta})\mathbf{g}(X_k, \boldsymbol{\theta})] \,, \tag{6.1}$$

where the $2m$ restrictions are given by $\boldsymbol{g}(X_k, \boldsymbol{\theta}) = (\cos(t_1 X_k) - \phi_\theta^R(t_1), \ldots, \cos(t_m X_k) - \phi_\theta^R(t_m), \sin(t_1 X_k) - \phi_\theta^I(t_1), \ldots, \sin(t_m X_k) - \phi_\theta^I(t_m)\,)'$ and the Lagrange multipliers satisfy the equation as (2.9). Also define a function

$$\mathbf{u}(\boldsymbol{\theta}) = \frac{K\mathbf{E}_{\theta_0}[\mathbf{g}(X_1, \boldsymbol{\theta})]}{1 + K\|\mathbf{E}_{\theta_0}[\mathbf{g}(X_1, \boldsymbol{\theta})]\|} \ .$$

We also define the set of integers by $\mathbf{N} = \{n | m(n) \geq 2\}$. Then for any $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ and

$n \in \mathbf{N}$, we have

$$\text{(6.2)} \quad \mathbf{E}_{\theta_0}[K\mathbf{u}'(\boldsymbol{\theta})\mathbf{g}(X_1,\boldsymbol{\theta})] = K^2 \sum_{l=1}^{m} \frac{\left[\left(\phi^R_{\theta_0}(t_l) - \phi^R_{\theta}(t_l)\right)^2 + \left(\phi^I_{\theta_0}(t_l) - \phi^I_{\theta}(t_l)\right)^2\right]}{1 + K\|\mathbf{E}_{\theta_0}[\mathbf{g}(X_1,\boldsymbol{\theta})]\|} > 0 \,.$$

For any $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, we take a sufficiently small $\delta > 0$, then for $n \in \mathbf{N}$ we have $\mathbf{E}_{\theta_0}[\sup_{\theta^* \in NB(\theta_0,\delta)} - K\mathbf{u}'(\boldsymbol{\theta}^*)g(X_1,\boldsymbol{\theta}^*)] < 0$, where $\mathbf{NB}(\boldsymbol{\theta}_0,\delta)$ is an open ball with center $\boldsymbol{\theta}_0$ and radius δ. Also $\sup_{\theta\in\Theta} |n^{-\frac{1}{3}+\frac{\epsilon}{2}} K\mathbf{u}'(\boldsymbol{\theta})\mathbf{g}(X_1,\boldsymbol{\theta})| \leq n^{-\frac{1}{3}+\frac{\epsilon}{2}} 2K\sqrt{2m}$ and it goes to 0 as $n \to +\infty$. By using Taylor's Theorem, there exists a $t \in (0,1)$ such that

$$-\log\left[1 + n^{-\frac{1}{3}+\frac{\epsilon}{2}} K\mathbf{u}'(\boldsymbol{\theta})\mathbf{g}(X_1,\boldsymbol{\theta}\right] = -n^{-\frac{1}{3}+\frac{\epsilon}{2}} K\mathbf{u}'(\boldsymbol{\theta})\mathbf{g}(X_1,\boldsymbol{\theta}) + \frac{n^{-\frac{2}{3}+\epsilon}(K\mathbf{u}'(\boldsymbol{\theta})\mathbf{g}(X_1,\boldsymbol{\theta}))^2}{2[1 + n^{-\frac{1}{3}+\frac{\epsilon}{2}} tK\mathbf{u}'(\boldsymbol{\theta})\boldsymbol{g}(X_1,\boldsymbol{\theta})]^2}$$

and then

(6.3)

$$\begin{aligned} & n^{\frac{1}{3}-\frac{\epsilon}{2}} \mathbf{E}_{\theta_0}\left\{\sup_{\theta^* \in NB(\theta_0,\delta)} - \log\left[1 + n^{-\frac{1}{3}+\frac{\epsilon}{2}} K\mathbf{u}'(\boldsymbol{\theta}^*)\mathbf{g}(X_1,\boldsymbol{\theta}^*)\right]\right\} \\ \leq\ & \mathbf{E}_{\theta_0}\left[\sup_{\theta^* \in NB(\theta_0,\delta)} - K\mathbf{u}'(\boldsymbol{\theta}^*)\mathbf{g}(X_1,\boldsymbol{\theta}^*)\right] + \mathbf{E}_{\theta_0}\left\{\sup_{\theta^* \in NB(\theta_0,\delta)} \frac{n^{-\frac{1}{3}+\frac{\epsilon}{2}}(K\mathbf{u}'(\boldsymbol{\theta}^*)\mathbf{g}(X_1,\boldsymbol{\theta}^*))^2}{2[1 + tn^{-\frac{1}{3}+\frac{\epsilon}{2}} K\mathbf{u}'(\boldsymbol{\theta}^*)\mathbf{g}(X_1,\boldsymbol{\theta}^*)]^2}\right\}. \end{aligned}$$

Since $n^{-\frac{1}{3}+\frac{\epsilon}{2}}[K\mathbf{u}'(\boldsymbol{\theta})\boldsymbol{g}(X_1,\boldsymbol{\theta})]^2 \leq 8K^2 n^{-\frac{1}{3}+\frac{\epsilon}{2}} m$ and it goes to 0 as $n \to \infty$, the second term of (6.3) converges to 0. By using (6.2), we can show that for any $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ and sufficiently small $\delta > 0$, there exists $n(\boldsymbol{\theta},\delta)$ such that for $n \geq n(\boldsymbol{\theta},\delta)$ we have $n^{\frac{1}{3}-\frac{\epsilon}{2}} \mathbf{E}_{\theta_0}\{\sup_{\theta^* \in NB(\theta_0,\delta)} - \log[1 + n^{-\frac{1}{3}+\frac{\epsilon}{2}} K\mathbf{u}'(\boldsymbol{\theta}^*)\mathbf{g}(X_1,\boldsymbol{\theta}^*)]\} < 0$. Since the set $\boldsymbol{\Theta} \setminus \mathbf{NB}(\boldsymbol{\theta}_0,\delta)$ is compact, there exist $L \in \mathbf{B}$ and $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_L$ such that $\boldsymbol{\Theta} \setminus \mathbf{NB}(\boldsymbol{\theta}_0,\delta) \subset \bigcup_{\alpha=1}^{L} \mathbf{NB}(\boldsymbol{\theta}_\alpha,\delta)$. Then for any $n \geq n(\boldsymbol{\theta}_\alpha,\delta)$ we have

$$\begin{aligned} & \mathbf{P}\left(\frac{1}{n}\sum_{k=1}^{n} \sup_{\theta^* \in NB(\theta_\alpha,\delta)} - \log[1 + n^{-\frac{1}{3}+\frac{\epsilon}{2}} K\mathbf{u}'(\boldsymbol{\theta}^*)\mathbf{g}(X_k,\boldsymbol{\theta}^*)] \right. \\ & \qquad \left. > \frac{1}{3} n^{-\frac{1}{3}+\frac{\epsilon}{2}} \mathbf{E}_{\theta_0}[\sup_{\theta^* \in NB(\theta_\alpha,\delta)} - K\mathbf{u}'(\boldsymbol{\theta}^*)\mathbf{g}(X_1,\boldsymbol{\theta}^*)]\right) \\ \leq\ & \mathbf{P}\left(\left|\frac{1}{n}\sum_{k=1}^{n} \sup_{\theta^* \in NB(\theta_\alpha,\delta)} - \log[1 + n^{-\frac{1}{3}+\frac{\epsilon}{2}} K\mathbf{u}'(\boldsymbol{\theta}^*)\mathbf{g}(X_k,\boldsymbol{\theta}^*)] \right.\right. \\ & \qquad \left. - \mathbf{E}_{\theta_0}\{\sup_{\theta^* \in NB(\theta_\alpha,\delta)} - \log[1 + n^{-\frac{1}{3}+\frac{\epsilon}{2}} K\mathbf{u}'(\boldsymbol{\theta}^*)\mathbf{g}(X_1,\boldsymbol{\theta}^*)]\}\right| \\ & \qquad \left. > -\frac{1}{6} n^{-\frac{1}{3}+\frac{\epsilon}{2}} \mathbf{E}_{\theta_0}[\sup_{\theta^* \in NB(\theta_\alpha,\delta)} - K\mathbf{u}'(\boldsymbol{\theta}^*)\mathbf{g}(X_1,\boldsymbol{\theta}^*)]\right) \\ \leq\ & \frac{6^2}{n^{\frac{1}{3}+\epsilon}} \frac{\mathbf{Var}_{\theta_0}\{\sup_{\theta^* \in NB(\theta_\alpha,\delta)} - \log[1 + n^{-\frac{1}{3}+\frac{\epsilon}{2}} K\mathbf{u}'(\boldsymbol{\theta}^*)\mathbf{g}(X_1,\boldsymbol{\theta}^*)]\}}{(\mathbf{E}_{\theta_0}[\sup_{\theta^* \in NB(\theta_\alpha,\delta)} - K\mathbf{u}'(\boldsymbol{\theta}^*)\mathbf{g}(X_1,\boldsymbol{\theta}^*)])^2} \,. \end{aligned} \quad \text{(6.4)}$$

We notice that the denominator of the right-hand side (RHS) of (6.4) does not converges to zero because of (6.2). Then by using the fact that RHS of (6.4) goes to 0, there exists $\bar{n}(\boldsymbol{\theta}_\alpha, \delta) \in \mathbf{N}$ such that for any $n \geq \bar{n}(\boldsymbol{\theta}_\alpha, \delta)$ it is less than $\delta/2K$ $(\alpha = 1, \ldots, L)$. Hence for any $n \geq \max_{1\leq\beta\leq L} \bar{n}(\boldsymbol{\theta}_\beta, \delta)$ we have

$$\mathbf{P}\left(\sup_{\theta^*\in\Theta\setminus NB(\theta_0,\delta)} \frac{1}{n}\sum_{k=1}^{n} -\log[1 + n^{-\frac{1}{3}+\frac{\epsilon}{2}} K\mathbf{u}^{'}(\boldsymbol{\theta}^*)\mathbf{g}(X_k, \boldsymbol{\theta}^*)] \right.$$
$$\left. > \frac{1}{3} n^{-\frac{1}{3}+\frac{\epsilon}{2}} \max_{1\leq\beta\leq L} \mathbf{E}_{\theta_0}[\sup_{\theta^*\in NB(\theta_\beta,\delta)} -K\mathbf{u}^{'}(\boldsymbol{\theta}^*)\mathbf{g}(X_1, \boldsymbol{\theta}^*)] \right) < \frac{\delta}{2} \ .$$

We notice that $\boldsymbol{\lambda}(\hat{\boldsymbol{\theta}}_n)$ is the minimum of (2.6) and then we have the relation

$$\mathbf{P}\left(\sup_{\theta^*\in\Theta\setminus NB(\boldsymbol{\theta}_0,\delta)} G_n(\theta^*) > \frac{1}{3} n^{-\frac{1}{3}+\frac{\epsilon}{2}} \max_{1\leq\beta\leq L} \mathbf{E}_{\theta_0}[\sup_{\theta^*\in NB(\theta_\beta,\delta)} -K\mathbf{u}^{'}(\boldsymbol{\theta}^*)\mathbf{g}(X_1, \boldsymbol{\theta}^*)] \right) < \frac{\delta}{2} \ . \tag{6.5}$$

Now we investigate the stochastic order of the Lagrange multipliers and we write them at the true value as $\boldsymbol{\lambda}(\boldsymbol{\theta}_0) = \|\boldsymbol{\lambda}(\boldsymbol{\theta}_0)\|\boldsymbol{\xi}$, where $\boldsymbol{\xi}$ is the $2m \times 1$ unit vector. By using the fact that the Lagrange multipliers are the solution of

$$\frac{1}{n}\sum_{k=1}^{n} \frac{K\mathbf{g}(X_k, \boldsymbol{\theta}_0)}{1 + K\boldsymbol{\lambda}^{'}(\boldsymbol{\theta}_0)\mathbf{g}(X_k, \boldsymbol{\theta}_0)} = \mathbf{0} \ , \tag{6.6}$$

we have

$$\begin{aligned} 0 &= \frac{1}{m}\left\| \frac{1}{n}\sum_{k=1}^{n} \frac{K\mathbf{g}(X_k, \boldsymbol{\theta}_0)}{1 + K\boldsymbol{\lambda}^{'}(\boldsymbol{\theta}_0)\mathbf{g}(X_k, \boldsymbol{\theta}_0)} \right\| \\ &\geq \frac{1}{m}\left| \frac{1}{n}\boldsymbol{\xi}^{'}\left(K\sum_{k=1}^{n} \boldsymbol{g}(X_k, \boldsymbol{\theta}_0) - \|\boldsymbol{\lambda}(\boldsymbol{\theta}_0)\| \sum_{k=1}^{n} \frac{K^2\mathbf{g}(X_k, \boldsymbol{\theta}_0)\boldsymbol{\xi}^{'}\mathbf{g}(X_k, \boldsymbol{\theta}_0)}{1 + K\|\boldsymbol{\lambda}(\boldsymbol{\theta}_0)\|\boldsymbol{\xi}^{'}\mathbf{g}(X_k, \boldsymbol{\theta}_0)} \right) \right| \\ &\geq \frac{\|\boldsymbol{\lambda}(\boldsymbol{\theta}_0)\|}{1 + K\|\boldsymbol{\lambda}(\boldsymbol{\theta}_0)\| \max_{1\leq k\leq n} \|\mathbf{g}(X_k, \boldsymbol{\theta}_0)\|} \frac{1}{n}\sum_{k=1}^{n} \frac{K^2}{m}\boldsymbol{\xi}^{'}\mathbf{g}(X_k, \boldsymbol{\theta}_0)\mathbf{g}(X_k, \boldsymbol{\theta}_0)^{'}\boldsymbol{\xi} \\ &\quad - \left| \frac{1}{n}\sum_{k=1}^{n} \frac{K}{m}\boldsymbol{\xi}^{'}\mathbf{g}(X_k, \boldsymbol{\theta}_0) \right| \ . \end{aligned} \tag{6.7}$$

We use the inequality $\mathbf{E}_{\theta_0}[\left|(1/n)\sum_{k=1}^{n} \boldsymbol{\xi}^{'}\mathbf{g}(X_k, \boldsymbol{\theta}_0)\right|^2] \leq (1/n)\mathbf{E}_{\theta_0}[\|\mathbf{g}(X_1, \boldsymbol{\theta}_0)\|^2]$, which is less than m/n. Then we have the relation that $\left|(1/n)\sum_{k=1}^{n} \boldsymbol{\xi}^{'}\mathbf{g}(X_k, \boldsymbol{\theta}_0)\right| = O_p(\sqrt{m/n})$. We define a $2m \times 2m$ matrix $\mathbf{D}(n) = (D_{ij}(n))$ by

$$\mathbf{D}(n) = \frac{1}{n}\sum_{k=1}^{n} \frac{K^2}{m}\left\{ \mathbf{g}(X_k, \boldsymbol{\theta}_0)\mathbf{g}(X_k, \boldsymbol{\theta}_0)^{'} - \mathbf{E}_{\theta_0}[\mathbf{g}(X_k, \boldsymbol{\theta}_0)\mathbf{g}(X_k, \boldsymbol{\theta}_0)^{'}] \right\} \ . \tag{6.8}$$

Then by using the Markov inequality we have the conditions that for any $\epsilon > 0$ $\mathbf{P}(|D_{ij}(n)| \geq \epsilon) \leq c_1(1/\sqrt{n})^2(1/m)^2$ and hence $\mathbf{P}(\max_{1\leq i,j\leq n} |D_{ij}(n)| \geq \epsilon) \leq c_2(1/n)$, where c_i $(i = 1, 2)$ are some positive constants [4] . Thus they converge to zero in probability as $n \longrightarrow +\infty$. Also by using the similar arguments as (6.34) and *Lemma A.1* below (we can take a sequence of nonsingular matrices as $\tilde{\boldsymbol{W}}$), we can show that the characteristic roots of $2m \times 2m$ matrix $\boldsymbol{\Sigma}_m(\boldsymbol{\theta}_0) = (K^2/m)\mathbf{E}_{\theta_0}[\mathbf{g}(X_k, \boldsymbol{\theta}_0)\mathbf{g}(X_k, \boldsymbol{\theta}_0)^{'}]$ are

[4] We are adopting the arguments used in Section 5 of Hjort et. al. (2004).

positive and bounded both from the above and zero. Because $\|\boldsymbol{\lambda}(\boldsymbol{\theta}_0)\|/[1+K\|\boldsymbol{\lambda}(\boldsymbol{\theta}_0)\|\max_{1\le k\le n}\|\mathbf{g}(X_k,\boldsymbol{\theta}_0)\|] = O_p(\sqrt{m/n})$ and $\max_{1\le k\le n}\|\boldsymbol{g}(X_k,\theta_0)\| \le 2\sqrt{2m} = O(\sqrt{m})$, we find the *key* condition

$$\|\boldsymbol{\lambda}(\boldsymbol{\theta}_0)\| = O_p(\sqrt{\frac{m}{n}}) \tag{6.9}$$

and

$$G_n(\boldsymbol{\theta}_0) \ge -\frac{1}{n}\sum_{k=1}^{n} K\boldsymbol{\lambda}^{'}(\boldsymbol{\theta}_0)\mathbf{g}(X_k,\boldsymbol{\theta}_0) = -K\|\boldsymbol{\lambda}(\boldsymbol{\theta}_0)\|\frac{1}{n}\sum_{k=1}^{n}\boldsymbol{\xi}^{'}\mathbf{g}(X_k,\boldsymbol{\theta}_0)\ , \tag{6.10}$$

which is of the order $O_p(\sqrt{m/n})O_p(\sqrt{m/n}) = o_p(n^{-\frac{1}{3}})$. Then there exists $n' \in \mathbf{N}$ such that for any $n \ge n^{'}$ we have

$$\mathbf{P}\Big(n^{\frac{1}{3}-\frac{\epsilon}{2}}G_n(\boldsymbol{\theta}_0) < n^{-\frac{\epsilon}{2}}\max_{1\le\beta\le L}\mathbf{E}_{\theta_0}[\sup_{\theta^*\in NB(\theta_\beta,\delta)} -K\boldsymbol{u}^{'}(\boldsymbol{\theta}^*)\mathbf{g}(X_1,\boldsymbol{\theta}^*)]\Big) < \frac{\delta}{2}\ . \tag{6.11}$$

Furthermore, by noting the fact that $n^{\frac{1}{3}-\frac{\epsilon}{2}}G_n(\hat{\boldsymbol{\theta}}_n) = \sup_{\theta^*\in\Theta} n^{\frac{1}{3}-\frac{\epsilon}{2}}G_n(\boldsymbol{\theta}^*)$ (which is greater than $n^{\frac{1}{3}-\frac{\epsilon}{2}}G_n(\boldsymbol{\theta}_0)$) and using (6.11), we can evaluate

$$\mathbf{P}\Big(n^{\frac{1}{3}-\frac{\epsilon}{2}}G_n(\hat{\boldsymbol{\theta}}_n) < n^{-\frac{\epsilon}{2}}\max_{1\le\beta\le L}\mathbf{E}_{\theta_0}[\sup_{\theta^*\in NB(\theta_\beta,\delta)} -K\mathbf{u}^{'}(\boldsymbol{\theta}^*)\mathbf{g}(X_1,\boldsymbol{\theta}^*)]\Big) < \frac{\delta}{2}\ . \tag{6.12}$$

Hence by combining (6.5) and (6.12), we have the condition $\mathbf{P}\left(\hat{\boldsymbol{\theta}}_n \notin \mathbf{NB}(\boldsymbol{\theta}_0,\delta)\right) < \delta$, which implies that $\hat{\boldsymbol{\theta}}_n \xrightarrow{p} \boldsymbol{\theta}_0$ as $n \to +\infty$.

(ii)**Asymptotic Normality** : We consider the first order condition of the criterion function $\partial G_n(\hat{\boldsymbol{\theta}}_n)/\partial\boldsymbol{\theta} = \mathbf{0}$. Then by expanding $-\partial G_n(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}$ around at $\hat{\boldsymbol{\theta}}_n = \boldsymbol{\theta}_0$, we have

$$-\sqrt{n}\frac{\partial G_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} = \frac{\partial^2 G_n(\boldsymbol{\theta}_n^{\dagger})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{'}}\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\ , \tag{6.13}$$

where we have taken $\|\boldsymbol{\theta}_n^{\dagger} - \hat{\boldsymbol{\theta}}_n\| \le \|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|$. In order to show the asymptotic normality of the random vector $-\sqrt{n}\partial G_n(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}$, we write

$$\begin{aligned} -\sqrt{n}\tfrac{\partial G_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} &= \sqrt{n}\frac{1}{n}\sum_{k=1}^{n}\frac{K}{1+K\boldsymbol{\lambda}^{'}(\boldsymbol{\theta}_0)\mathbf{g}(X_k,\boldsymbol{\theta}_0)}\left(\frac{\partial\mathbf{g}(X_k,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}^{'}}\right)^{'}\boldsymbol{\lambda}(\boldsymbol{\theta}_0) \\ &= \sqrt{n}\frac{1}{n}\sum_{k=1}^{n}\frac{K}{1+K\boldsymbol{\lambda}^{'}(\boldsymbol{\theta}_0)\mathbf{g}(X_k,\boldsymbol{\theta}_0)}\left(-\frac{\partial\boldsymbol{\Phi}_{\theta_0}}{\partial\boldsymbol{\theta}^{'}}\right)^{'}\boldsymbol{\lambda}(\boldsymbol{\theta}_0)\ . \end{aligned} \tag{6.14}$$

By rewriting (6.6) with respect to the Lagrange multipliers, we have

$$\mathbf{0} = \frac{1}{n}\sum_{k=1}^{n}\frac{K}{m}\mathbf{g}(X_k,\boldsymbol{\theta}_0) - \left\{\frac{1}{n}\sum_{k=1}^{n}\frac{K^2}{m}\mathbf{g}(X_k,\boldsymbol{\theta}_0)\mathbf{g}(X_k,\boldsymbol{\theta}_0)^{'}\right\}\boldsymbol{\lambda}(\boldsymbol{\theta}_0) + r_{1n}\ , \tag{6.15}$$

where we set the remainder term

$$r_{1n} = \frac{1}{n}\sum_{k=1}^{n}\frac{K}{m}\frac{\mathbf{g}(X_k,\boldsymbol{\theta}_0)(K\boldsymbol{\lambda}^{'}(\boldsymbol{\theta}_0)\mathbf{g}(X_k,\boldsymbol{\theta}_0))^2}{1+K\boldsymbol{\lambda}^{'}(\boldsymbol{\theta}_0)\mathbf{g}(X_k,\boldsymbol{\theta}_0)}\ .$$

Because $\max_{1\le k\le n}|K\boldsymbol{\lambda}'(\boldsymbol{\theta}_0)\mathbf{g}(X_k,\boldsymbol{\theta}_0)| \le K\|\boldsymbol{\lambda}(\boldsymbol{\theta}_0)\| \max_{1\le k\le n}\|\mathbf{g}(X_k,\boldsymbol{\theta}_0)\|$ is of the order $O_p(\sqrt{m/n})O(\sqrt{m}) = o_p(1)$, we have $\max_{1\le k\le n} 1/[|1+K\boldsymbol{\lambda}'(\boldsymbol{\theta}_0)\mathbf{g}(X_k,\boldsymbol{\theta}_0)|] = O_p(1)$. Also there exists a positive constant c_3 such that $\|\mathbf{B}_m(\boldsymbol{\theta})\| \le c_3\sqrt{m}$. By using these evaluations, we find

$$\|r_{1n}\| \le \max_{1\le k\le n}\left(K^3/[|1+K\boldsymbol{\lambda}'(\boldsymbol{\theta}_0)\mathbf{g}(X_k,\boldsymbol{\theta}_0)|]\right)(1/m)\|\boldsymbol{\lambda}(\boldsymbol{\theta}_0)\|^2 \max_{1\le k\le n}\|\mathbf{g}(X_k,\boldsymbol{\theta}_0)\|^3 ,$$

which is of the order $O_p(m^{-1})O_p(m/n)O(m^{3/2}) = O_p(m^{3/2}/n)$. Then we can approximate the random vector $-\sqrt{n}\partial G_n(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}$ as

(6.16)
$$\begin{aligned} & -\sqrt{n}\frac{\partial G_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} \\ = \quad & \sqrt{n}\left\{\frac{1}{n}\sum_{k=1}^n \frac{K}{1+K\boldsymbol{\lambda}'(\boldsymbol{\theta}_0)\mathbf{g}(X_k,\boldsymbol{\theta}_0)}\left(-\frac{\partial\boldsymbol{\Phi}_{\theta_0}}{\partial\boldsymbol{\theta}'}\right)'\right\}\mathbf{S}_{n,\theta_0}^{-1}\left\{\frac{1}{n}\sum_{k=1}^n\frac{K}{m}\mathbf{g}(X_k,\boldsymbol{\theta}_0)\right\} + r_{2n} , \end{aligned}$$

where

$$\mathbf{S}_{n,\theta} = \frac{1}{n}\sum_{k=1}^n \frac{K^2}{m}\mathbf{g}(X_k,\boldsymbol{\theta})\mathbf{g}(X_k,\boldsymbol{\theta})'$$

and r_{2n} is the remainder term. By applying the similar arguments to r_{1n}, we find that $r_{2n} = o_p(1)$ and then we can further approximate (6.16) as

$$-\sqrt{n}\frac{\partial G_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} = \frac{1}{\sqrt{n}}\sum_{k=1}^n K\left(-\frac{\partial\boldsymbol{\Phi}_{\theta_0}}{\partial\boldsymbol{\theta}'}\right)'\mathbf{S}_{n,\theta_0}^{-1}\frac{K}{m}\mathbf{g}(X_k,\boldsymbol{\theta}_0) + o_p(1) . \tag{6.17}$$

In order to show the asymptotic normality of (6.17), for any vector $\boldsymbol{\zeta}\in\mathbf{R}^4$ we define a set of random variables Y_{mk} $(k=1,\ldots,n)$ by

$$Y_{mk} = \boldsymbol{\zeta}'K\left(-\frac{\partial\boldsymbol{\Phi}_{\theta_0}}{\partial\boldsymbol{\theta}'}\right)'\{\mathbf{E}_{\theta_0}(\boldsymbol{S}_{n,\theta_0})\}^{-1}\frac{K}{m}\mathbf{g}(X_k,\boldsymbol{\theta}_0) .$$

Then we find that for any $\eta > 0$

$$\begin{aligned} & \frac{\sum_{k=1}^n \mathbf{E}_{\theta_0}[Y_{mk}^2 : |Y_{mk}| > \eta\,(\sum_{k=1}^n \mathbf{Var}_{\theta_0}(Y_{mk}))^{1/2}]}{\sum_{k=1}^n \mathbf{Var}_{\theta_0}(Y_{mk})} \\ = \quad & \frac{\mathbf{E}_{\theta_0}\left[(Y_{m1}/\sqrt{m})^2 mI\left(|Y_{m1}/\sqrt{m}| > \eta\,((n/m)\mathbf{Var}_{\theta_0}(Y_{m1}))^{1/2}\right)\right]}{\mathbf{Var}_{\theta_0}(Y_{m1})} \to 0 \end{aligned} \tag{6.18}$$

and

$$\mathbf{Var}_{\theta_0}(Y_{m1}) = \frac{K^2}{m}\boldsymbol{\zeta}'\left(-\frac{\partial\boldsymbol{\Phi}_{\theta_0}}{\partial\boldsymbol{\theta}'}\right)'\{\mathbf{E}_{\theta_0}(\mathbf{S}_{n,\theta_0})\}^{-1}\left(-\frac{\partial\boldsymbol{\Phi}_{\theta_0}}{\partial\boldsymbol{\theta}'}\right)\boldsymbol{\zeta} \tag{6.19}$$

In the above derivation we have used several evaluation of stochastic orders, which are similar to the previous ones. By applying the Lindeberg-type condition for the central limit theorem, we can prove that for any vector $\zeta\in\mathbf{R}^4$

$$\frac{1}{\sqrt{n}}\sum_{k=1}^n \boldsymbol{\zeta}'K\left(-\frac{\partial\boldsymbol{\Phi}_{\theta_0}}{\partial\boldsymbol{\theta}'}\right)'\mathbf{S}_{n,\theta_0}^{-1}\frac{K}{m}\mathbf{g}(X_k,\boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0},\boldsymbol{\zeta}'\mathbf{J}_K(\boldsymbol{\theta}_0)\boldsymbol{\zeta}) , \tag{6.20}$$

where the variance-covariance matrix is given by
$\mathbf{J}_K(\boldsymbol{\theta}_0) = \lim_{n\to\infty} \mathbf{B}_m(\boldsymbol{\theta}_0)^{'}\mathbf{A}_m(\boldsymbol{\theta}_0)^{-1}\mathbf{B}_m(\boldsymbol{\theta}_0)$. (See *Lemma A.1* below.)
Because the vector $\boldsymbol{\zeta}$ is arbitrary, we use the Cramér-Wold device to obtain that $-\sqrt{n}\partial G_n(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}$ converges to the normal distribution with the variance-covariance matrix $\mathbf{J}_K(\boldsymbol{\theta}_0)$ as $n \longrightarrow +\infty$.

Next for the second derivatives we shall show

$$\frac{\partial^2 G_n(\boldsymbol{\theta}_n^{\dagger})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{'}} = -\mathbf{J}_K(\boldsymbol{\theta}_0) + o_p(1) \ . \tag{6.21}$$

In order to do it, we first take a compact set around the true vector $\boldsymbol{\theta}_0$ as $\mathbf{NB}(\boldsymbol{\theta}_0) = \{\boldsymbol{\theta}|\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| \leq n^{-1/3}\}$ and the Lagrange multiplier vector $\lambda(\boldsymbol{\theta})$ satisfying

$$\frac{1}{n}\sum_{k=1}^{n}\frac{K\mathbf{g}(X_k,\boldsymbol{\theta})}{1+K\boldsymbol{\lambda}(\boldsymbol{\theta})^{'}\mathbf{g}(X_k,\boldsymbol{\theta})} = \mathbf{0} \ . \tag{6.22}$$

Then we need to consider each terms of the second derivatives and evaluate their stochastic orders, which are given by

$$\begin{aligned}
\frac{\partial^2 G_n(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{'}} &= \frac{K^2}{n}\sum_{k=1}^{n}\frac{\left(-\partial\boldsymbol{\Phi}_{\theta}/\partial\boldsymbol{\theta}^{'}\right)^{'}\boldsymbol{\lambda}(\boldsymbol{\theta})\boldsymbol{\lambda}^{'}(\boldsymbol{\theta})\left(-\partial\boldsymbol{\Phi}_{\theta}/\partial\boldsymbol{\theta}^{'}\right)}{[1+K\boldsymbol{\lambda}^{'}(\boldsymbol{\theta})\mathbf{g}(X_k,\boldsymbol{\theta})]^2} \\
&+ \frac{K^2}{n}\sum_{k=1}^{n}\frac{\left(-\partial\boldsymbol{\Phi}_{\theta}/\partial\boldsymbol{\theta}^{'}\right)^{'}\boldsymbol{\lambda}(\boldsymbol{\theta})\mathbf{g}(X_k,\boldsymbol{\theta})^{'}\left(\partial\boldsymbol{\lambda}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}^{'}\right)}{[1+K\boldsymbol{\lambda}^{'}(\boldsymbol{\theta})\mathbf{g}(X_k,\boldsymbol{\theta})]^2} \\
&- \frac{K}{n}\sum_{k=1}^{n}\frac{\left(-\partial\boldsymbol{\Phi}_{\theta}/\partial\boldsymbol{\theta}^{'}\right)^{'}\left(\partial\boldsymbol{\lambda}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}^{'}\right)}{1+K\boldsymbol{\lambda}^{'}(\boldsymbol{\theta})\mathbf{g}(X_k,\boldsymbol{\theta})} \\
&- \frac{K}{n}\sum_{k=1}^{n}\sum_{i=1}^{2}\sum_{l=1}^{m}\frac{\lambda_{il}(\boldsymbol{\theta})(\partial^2 g_{(i-1)m+l}(X_k,\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{'})}{1+K\boldsymbol{\lambda}^{'}(\boldsymbol{\theta})\mathbf{g}(X_k,\boldsymbol{\theta})} \ ,
\end{aligned} \tag{6.23}$$

where $\mathbf{g}(X_k,\boldsymbol{\theta}) = (g_l(X_k,\boldsymbol{\theta}))$ $(k=1,\cdots,n)$.
There are some complications in our evaluations partly because we have the 3rd term involving the derivatives of the Lagrange multipliers. By differentiating (6.22) with respect to $\boldsymbol{\theta}$, we have the relation
(6.24)

$$\mathbf{T}_n(\theta)\frac{\partial\boldsymbol{\lambda}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^{'}} = \frac{1}{n}\sum_{k=1}^{n}\frac{K}{m}\frac{\left(-\partial\boldsymbol{\Phi}_{\theta}/\partial\boldsymbol{\theta}^{'}\right)}{1+K\boldsymbol{\lambda}^{'}(\boldsymbol{\theta})\mathbf{g}(X_k,\boldsymbol{\theta})} - \frac{1}{n}\sum_{k=1}^{n}\frac{K^2}{m}\frac{\boldsymbol{g}(X_k,\theta)\boldsymbol{\lambda}^{'}(\boldsymbol{\theta})\left(-\partial\boldsymbol{\Phi}_{\theta}/\partial\boldsymbol{\theta}^{'}\right)}{(1+K\boldsymbol{\lambda}^{'}(\boldsymbol{\theta})\boldsymbol{g}(X_k,\boldsymbol{\theta}))^2} \ ,$$

where we define a $2m \times 2m$ random matrix $\mathbf{T}_n(\boldsymbol{\theta})$ by

$$\mathbf{T}_{n,\theta} = \frac{1}{n}\sum_{k=1}^{n}\frac{K^2}{m}\frac{\boldsymbol{g}(X_k,\boldsymbol{\theta})\boldsymbol{g}(X_k,\boldsymbol{\theta})^{'}}{[1+K\boldsymbol{\lambda}^{'}(\boldsymbol{\theta})\boldsymbol{g}(X_k,\boldsymbol{\theta})]^2} \ .$$

Then it is straightforward to show that the second term of LHS of (6.24) is of the smaller order than the first term because it is dominated by

$$\sup_{\theta\in NB(\theta_0)} \left\| \frac{K^2}{n} \sum_{k=1}^{n} \frac{\left(-\partial \boldsymbol{\Phi}_\theta / \partial \boldsymbol{\theta}'\right)' \boldsymbol{\lambda}(\boldsymbol{\theta}) \boldsymbol{g}(X_k, \boldsymbol{\theta})}{[1 + K\boldsymbol{\lambda}'(\boldsymbol{\theta})\boldsymbol{g}(X_k, \boldsymbol{\theta})]^2} \right\|$$

$$\leq \sup_{\theta\in NB(\theta_0)} \left\{ K^2 \max_{1\leq k\leq n} \frac{1}{[1+K\boldsymbol{\lambda}'(\boldsymbol{\theta})\boldsymbol{g}(X_k,\boldsymbol{\theta})]^2} \left\| \frac{\partial \boldsymbol{\Phi}_\theta}{\partial \boldsymbol{\theta}'} \right\| \|\boldsymbol{\lambda}(\boldsymbol{\theta})\| \max_{1\leq k\leq n} \|\boldsymbol{g}(X_k, \boldsymbol{\theta})\| \right\} ,$$

which can be evaluated as $o_p(1)$. Then we find that the second term of (6.23) is of the order $o_p(1)$. Similarly, we find that the first term and the fourth term of (6.23) are dominated by

$$\sup_{\theta\in NB(\theta_0)} \left\{ K^2 \max_{1\leq k\leq n} \frac{1}{[1 + K\boldsymbol{\lambda}'(\boldsymbol{\theta})\boldsymbol{g}(X_k, \boldsymbol{\theta})]^2} \|\boldsymbol{\lambda}(\boldsymbol{\theta})\|^2 \left\| \frac{\partial \boldsymbol{\Phi}_\theta}{\partial \boldsymbol{\theta}'} \right\|^2 \right\}$$

and

$$\sup_{\theta\in NB(\theta_0)} \left\{ K \max_{1\leq k\leq n} \frac{1}{|1 + K\boldsymbol{\lambda}'(\boldsymbol{\theta})\boldsymbol{g}(X_k, \boldsymbol{\theta})|} \left\| \sum_{i=1}^{2} \sum_{l=1}^{m} \lambda_{il}(\boldsymbol{\theta}) \frac{\partial^2 g_{(i-1)m+l}(X_k, \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right\| \right\} ,$$

respectively. Then by evaluating the stochastic orders of these terms, we can find that they are of the order $o_p(1)$.

As the dominant term, we need to evaluate the 3rd term of (6.23). After tedious but straightforward calculations on the 3rd term of (6.23), it is possible to show that

$$\sup_{\theta\in NB(\theta_0)} \left\| \left\{ \frac{K}{n} \sum_{k=1}^{n} \frac{\left(-\partial \boldsymbol{\Phi}_\theta / \partial \boldsymbol{\theta}'\right)'}{1 + K\boldsymbol{\lambda}'(\boldsymbol{\theta})\boldsymbol{g}(X_k, \boldsymbol{\theta})} \right\} T_{n,\theta}^{-1} \left\{ \frac{1}{n} \sum_{k=1}^{n} \frac{K^2}{m} \frac{\boldsymbol{g}(X_k, \boldsymbol{\theta})\boldsymbol{\lambda}'(\boldsymbol{\theta}) \left(-\partial \boldsymbol{\Phi}_\theta / \partial \boldsymbol{\theta}'\right)}{(1 + K\boldsymbol{\lambda}'(\boldsymbol{\theta})\boldsymbol{g}(X_k, \boldsymbol{\theta}))^2} \right\} \right\| = o_p(1)$$

and then we find that

$$\sup_{\theta\in NB(\theta_0)} \left\| \frac{K}{n} \sum_{k=1}^{n} \frac{\left(-\partial \boldsymbol{\Phi}_\theta / \partial \boldsymbol{\theta}'\right)' \left(\partial \boldsymbol{\lambda}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}'\right)}{1 + K\boldsymbol{\lambda}'(\boldsymbol{\theta})\boldsymbol{g}(X_k, \boldsymbol{\theta})} - J_K(\boldsymbol{\theta}_0) \right\| = o_p(1) . \tag{6.25}$$

Therefore, we have established the asymptotic normality

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, \mathbf{J}_K(\boldsymbol{\theta}_0)^{-1}) . \tag{6.26}$$

By taking sufficiently large K and use **Lemma A.1** below, we have $\lim_{K\to\infty} \mathbf{J}_{K,\theta_0}^{-1} = \mathbf{I}(\theta_0)^{-1}$.

(**Q.E.D.**)

Now we present the *key* Lemma for the asymptotic distributions and asymptotic efficiency of the MEL estimation.

Lemma A.1 : Under the assumptions of *Theorem 2.2*, we have

$$\lim_{K\to\infty} \mathbf{J}_K(\boldsymbol{\theta}_0) = \lim_{K\to\infty} \lim_{n\to\infty} \left[\mathbf{B}_m(\boldsymbol{\theta}_0)' \mathbf{A}_m(\boldsymbol{\theta}_0)^{-1} \mathbf{B}_m(\boldsymbol{\theta}_0) \right] = \mathbf{I}(\boldsymbol{\theta}_0) \tag{6.27}$$

as $m = m(n) \to +\infty$ $(n \to +\infty)$, where $\mathbf{A}_m(\boldsymbol{\theta}_0)$ and $\mathbf{B}_m(\boldsymbol{\theta}_0)$ are defined in *Theorem 2.1* at $t_l = Kl/m$ $(l = 1, \cdots, m)$.

Proof of Lemma A.1:
For the random variable X followed by the stable distribution $F_\theta(\cdot)$, we define $2m \times 1$ complex vectors $\tilde{\mathbf{g}}(X, \boldsymbol{\theta})$ and $\widetilde{\boldsymbol{\Phi}}_\theta$ by
$\tilde{\mathbf{g}}(X, \boldsymbol{\theta}) = \left[e^{it_1X} - \phi_\theta(t_1), \cdots, e^{it_mX} - \phi_\theta(t_m), e^{-it_1X} - \phi_\theta(-t_1), \cdots, e^{-it_mX} - \phi_\theta(-t_m)\right]'$
and $\widetilde{\Phi}_\theta = [\phi_\theta(t_1), \cdots, \phi_\theta(t_m), \phi_\theta(-t_1), \cdots, \phi_\theta(-t_m)]'$.
Then we find that

$$\mathbf{B}_m(\boldsymbol{\theta}_0)'\mathbf{A}_m(\boldsymbol{\theta}_0)^{-1}\mathbf{B}_m(\boldsymbol{\theta}_0) = \left(\frac{\partial \widetilde{\boldsymbol{\Phi}}_{\theta_0}}{\partial \boldsymbol{\theta}'}\right)' \{\mathbf{E}_{\theta_0}[\tilde{\mathbf{g}}(X, \boldsymbol{\theta}_0)\tilde{\mathbf{g}}(X, \boldsymbol{\theta}_0)']\}^{-1} \frac{\partial \widetilde{\boldsymbol{\Phi}}_{\theta_0}}{\partial \boldsymbol{\theta}'} .$$

Furthremore, we set a 4×1 vector $\boldsymbol{w}_\theta(t)$ by

$$\boldsymbol{w}_\theta(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\partial \log f_\theta(x)}{\partial \boldsymbol{\theta}} \, e^{-itx} dx \; , \tag{6.28}$$

where $f_\theta(x)$ is the density function with the parameter $\boldsymbol{\theta}$ and $\widetilde{\boldsymbol{W}}_\theta$ is a $2m \times 4$ matrix $\widetilde{\boldsymbol{W}}_\theta = (\boldsymbol{w}_\theta(t_1), \cdots, \boldsymbol{w}_\theta(t_m), \boldsymbol{w}_\theta(-t_1), \cdots, \boldsymbol{w}_\theta(-t_m))'$.
Then we have the convergence as
(6.29)

$$\frac{K}{m} \widetilde{\boldsymbol{W}}_{\theta_0}' \frac{\partial \widetilde{\boldsymbol{\Phi}}_{\theta_0}}{\partial \boldsymbol{\theta}'} = \frac{K}{m} \sum_{l=-m, l \neq 0}^{l=m} \boldsymbol{w}_{\theta_0}(t_l) \left(\frac{\partial \phi_{\theta_0}(t_l)}{\partial \boldsymbol{\theta}}\right)' \longrightarrow \int_{-K}^{K} \boldsymbol{w}_{\theta_0}(t) \left(\frac{\partial \phi_{\theta_0}(t)}{\partial \boldsymbol{\theta}}\right)' dt \; ,$$

as $n \to \infty$. Similarly, we have the convergence

$$\begin{aligned} & \tfrac{K^2}{m^2} \widetilde{\boldsymbol{W}}_{\theta_0}' \mathbf{E}_{\theta_0}[\tilde{\mathbf{g}}(X, \boldsymbol{\theta}_0)\tilde{\mathbf{g}}(X, \boldsymbol{\theta}_0)'] \widetilde{\boldsymbol{W}}_{\theta_0} \\ \longrightarrow \; & \int_{-K}^{K} \int_{-K}^{K} \{\phi_{\theta_0}(s+t) - \phi_{\theta_0}(s)\phi_{\theta_0}(t)\} \boldsymbol{w}_{\theta_0}(s) \boldsymbol{w}_{\theta_0}'(t) ds dt \; , \end{aligned} \tag{6.30}$$

as $n \to \infty$. Hence
(6.31)

$$\begin{aligned} & \lim_{n \to \infty} \left(\widetilde{\boldsymbol{W}}_{\theta_0}' \frac{\partial \widetilde{\boldsymbol{\Phi}}_{\theta_0}}{\partial \boldsymbol{\theta}'}\right)' \{\widetilde{\boldsymbol{W}}_{\theta_0}' \mathbf{E}_{\theta_0}[\tilde{\boldsymbol{g}}(X, \boldsymbol{\theta}_0)\tilde{\boldsymbol{g}}(X, \boldsymbol{\theta}_0)'] \widetilde{\boldsymbol{W}}_{\theta_0}\}^{-1} \widetilde{\boldsymbol{W}}_{\theta_0}' \frac{\partial \widetilde{\boldsymbol{\Phi}}_{\theta_0}}{\partial \boldsymbol{\theta}'} \\ = \; & \left\{\int_{-K}^{K} \boldsymbol{w}_{\theta_0}(t) \left(\frac{\partial \phi_{\theta_0}(t)}{\partial \boldsymbol{\theta}}\right)' dt\right\}' \left\{\int_{-K}^{K}\int_{-K}^{K} \{\phi_{\theta_0}(s+t) - \phi_{\theta_0}(s)\phi_{\theta_0}(t)\} \boldsymbol{w}_{\theta_0}(s) \boldsymbol{w}_{\theta_0}(t) ds dt\right\}^{-1} \\ & \times \left\{\int_{-K}^{K} \boldsymbol{w}_{\theta_0}(t) \left(\frac{\partial \phi_{\theta_0}(t)}{\partial \boldsymbol{\theta}}\right)' dt\right\} . \end{aligned}$$

Let denote the RHS of (6.31) as $\boldsymbol{\Xi}_K(\boldsymbol{\theta}_0)$. Then for any (non-degenerate) $2m \times 4$ matrix $\mathbf{v}$ we have

$$\left[\left(\mathbf{v}' \frac{\partial \widetilde{\boldsymbol{\Phi}}_{\theta_0}}{\partial \boldsymbol{\theta}'}\right)' \{\mathbf{v}' \mathbf{E}_{\theta_0}[\tilde{\mathbf{g}}(X, \boldsymbol{\theta}_0)\tilde{\mathbf{g}}(X, \boldsymbol{\theta}_0)'] \mathbf{v}\}^{-1} \left(\mathbf{v}' \frac{\partial \widetilde{\boldsymbol{\Phi}}_{\theta_0}}{\partial \boldsymbol{\theta}'}\right)\right]^{-1} \tag{6.32}$$

can be minimized at

$$\mathbf{v} = \left\{\mathbf{E}_{\theta_0}[\tilde{\mathbf{g}}(X, \boldsymbol{\theta}_0)\tilde{\mathbf{g}}(X, \boldsymbol{\theta}_0)']\right\}^{-1} \frac{\partial \widetilde{\boldsymbol{\Phi}}_{\theta_0}}{\partial \boldsymbol{\theta}'}$$

and the minimum value is

$$\left[\left(\frac{\partial\widetilde{\boldsymbol{\Phi}}_{\theta_0}}{\partial\boldsymbol{\theta}^{'}}\right)^{'}\{\mathbf{E}_{\theta_0}[\tilde{\mathbf{g}}(X,\boldsymbol{\theta}_0)\tilde{\mathbf{g}}(X,\boldsymbol{\theta}_0)^{'}]\}^{-1}\frac{\partial\widetilde{\boldsymbol{\Phi}}_{\theta_0}}{\partial\boldsymbol{\theta}^{'}}\right]^{-1}.$$

It has been well-known that the asymptotic efficiency bound is given by $\mathbf{I}(\boldsymbol{\theta}_0)^{-1}$, provided that it is nonsingular. Thus for any 4×1 non-zero vector $\mathbf{u}$,
(6.33)

$$\begin{aligned}
\mathbf{u}^{'}\mathbf{I}(\boldsymbol{\theta}_0)^{-1}\mathbf{u} &\leq \mathbf{u}^{'}\left[\left(\frac{\partial\widetilde{\boldsymbol{\Phi}}_{\theta_0}}{\partial\boldsymbol{\theta}^{'}}\right)^{'}\{\mathbf{E}_{\theta_0}[\tilde{\mathbf{g}}(X,\boldsymbol{\theta}_0)\tilde{\mathbf{g}}(X,\boldsymbol{\theta}_0)^{'}]\}^{-1}\frac{\partial\widetilde{\boldsymbol{\Phi}}_{\theta_0}}{\partial\boldsymbol{\theta}^{'}}\right]^{-1}\mathbf{u} \\
&\leq \mathbf{u}^{'}\left[\left(\widetilde{\boldsymbol{W}}_{\theta_0}^{'}\frac{\partial\widetilde{\boldsymbol{\Phi}}_{\theta_0}}{\partial\boldsymbol{\theta}^{'}}\right)^{'}\{\widetilde{\boldsymbol{W}}_{\theta_0}^{'}\mathbf{E}_{\theta_0}[\tilde{\mathbf{g}}(X,\boldsymbol{\theta}_0)\tilde{\mathbf{g}}(X,\boldsymbol{\theta}_0)^{'}]\boldsymbol{W}_{\theta_0}\}^{-1}\left(\widetilde{\boldsymbol{W}}_{\theta_0}^{'}\frac{\partial\widetilde{\boldsymbol{\Phi}}_{\theta_0}}{\partial\boldsymbol{\theta}^{'}}\right)\right]^{-1}\mathbf{u}\,,
\end{aligned}$$

and we have

$$\lim_{K\to\infty}\boldsymbol{\Xi}_K(\boldsymbol{\theta}_0)=\mathbf{I}(\boldsymbol{\theta}_0)$$

by using the same arguments developed by Feuerverger and McDunnough (1981a) for the information matrix. Hence we have obtained the result.
(**Q.E.D.**)

Proof of Theorem 2.3:
[i] : The first part of the proof of *Theorem 2.3* is similar to the proof of the testing hypothesis problems given by Owen (1990), and Quin and Lawles (1994) except the fact that the number of restrictions $m(n)$ increases as $n\longrightarrow\infty$. The precise evaluations of stochastic orders in our derivations are quite tedious, but most of them are straightforward as in the proof of *Theorem 2.2*.

Let $Y_k(\boldsymbol{\theta})=K\boldsymbol{\lambda}(\boldsymbol{\theta})^{'}\boldsymbol{g}(X_k,\boldsymbol{\theta})$ $(k=1,\cdots,n)$. Then the criterion function $G_n(\hat{\boldsymbol{\theta}}_n)$ of (6.1) in the MEL estimation can be rewritten as

$$G_n(\hat{\boldsymbol{\theta}}_n)=(-\frac{1}{n})\sum_{k=1}^{n}\left[Y_k(\hat{\boldsymbol{\theta}}_n)-\frac{1}{2}Y_k(\hat{\boldsymbol{\theta}}_n)^2+\frac{1}{3}Y_k(\boldsymbol{\theta}^*)^3\right]\,, \tag{6.34}$$

where we have $\|\boldsymbol{\theta}^*-\hat{\boldsymbol{\theta}}_n\|\leq\|\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0\|$. Then we find that

$$\begin{aligned}
\sum_{k=1}^{n}|Y_k(\boldsymbol{\theta}^*)|^3 &\leq K^2\left[\max_{1\leq k\leq n}|Y_k(\boldsymbol{\theta}^*)|\right]\sum_{k=1}^{n}|\boldsymbol{\lambda}(\boldsymbol{\theta}^*)^{'}\mathbf{g}(X_k,\boldsymbol{\theta}^*)|^2 \\
&\leq Km^{3/2}\|\boldsymbol{\lambda}(\boldsymbol{\theta}^*)\|\left[\sqrt{n}\boldsymbol{\lambda}(\boldsymbol{\theta}^*)\right]^{'}\left[\frac{K^2}{n}\sum_{k=1}^{n}\frac{1}{m}\mathbf{g}(X_k,\boldsymbol{\theta}^*)\mathbf{g}(X_k,\boldsymbol{\theta}^*)\right]\left[\sqrt{n}\boldsymbol{\lambda}(\boldsymbol{\theta}^*)\right]
\end{aligned}$$

and the stochastic order of th last term is $[m^{3/2}\times O_p(\sqrt{m/n})][\sqrt{n}\times O_p(\sqrt{m/n})]^2=o_p(1)$. Here we have used the fact that $m(n)^6/n\longrightarrow 0$ by the assumption of *Theorem 2.3*. By expanding $\sum_{k=1}^{n}Y_k(\hat{\boldsymbol{\theta}}_n)$ around the true parameter values $\boldsymbol{\theta}_0$, it can be approximated by

(6.35)

$$
\begin{aligned}
&\sum_{k=1}^{n} Y_k(\hat{\boldsymbol{\theta}}_n) \\
= \;& K\sqrt{n}\boldsymbol{\lambda}(\boldsymbol{\theta}_0)^{'}[\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\mathbf{g}(X_k,\boldsymbol{\theta}_0)] + K\sqrt{n}(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)^{'}\frac{\partial\boldsymbol{\lambda}(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}[\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\mathbf{g}(X_k,\boldsymbol{\theta}_0)] \\
+ \;& K\sqrt{n}\boldsymbol{\lambda}(\boldsymbol{\theta}_0)^{'}[-\frac{\partial\boldsymbol{\Phi}_{\theta_0}}{\partial\boldsymbol{\theta}^{'}}](\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)\sqrt{n} + K\sqrt{n}(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)^{'}\frac{\partial\boldsymbol{\lambda}(\boldsymbol{\theta}_0)^{'}}{\partial\boldsymbol{\theta}}[-\frac{\partial\boldsymbol{\Phi}_{\theta_0}}{\partial\boldsymbol{\theta}^{'}}](\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)\sqrt{n} + o_p(1)\;.
\end{aligned}
$$

From (6.24), (6.34) and *Theorem 2.2*, the sum of dominant factors in the last two terms of (6.35) are cancelled out and it is of the order of $o_p(1)$. Hence (6.35) can be further approximated as

(6.36)

$$
\begin{aligned}
&\sum_{k=1}^{n} Y_k(\hat{\boldsymbol{\theta}}_n) \\
= \;& K\sqrt{n}\boldsymbol{\lambda}(\boldsymbol{\theta}_0)^{'}\left[\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\mathbf{g}(X_k,\boldsymbol{\theta}_0)\right] \\
- \;& K\sqrt{n}\boldsymbol{\lambda}(\boldsymbol{\theta}_0)^{'}(\frac{\partial\boldsymbol{\Phi}_{\theta_0}}{\partial\boldsymbol{\theta}}^{'})\left[(\frac{\partial\boldsymbol{\Phi}_{\theta_0}}{\partial\boldsymbol{\theta}^{'}})^{'}\mathbf{S}_{n,\theta_0}^{-1}(\frac{\partial\boldsymbol{\Phi}_{\theta_0}}{\partial\boldsymbol{\theta}^{'}})\right]^{-1}(\frac{\partial\boldsymbol{\Phi}_{\theta_0}}{\partial\boldsymbol{\theta}^{'}})^{'}K\sqrt{n}\boldsymbol{\lambda}(\boldsymbol{\theta}_0) + o_p(1)\;.
\end{aligned}
$$

Also by expanding $\sum_{k=1}^{n} Y_k(\hat{\boldsymbol{\theta}}_n)^2$ around the true parameter values $\boldsymbol{\theta}_0$ and using the similar arguments, it can be further approximated by

(6.37)

$$
\begin{aligned}
\sum_{k=1}^{n} Y_k(\hat{\boldsymbol{\theta}}_n)^2 \;=\;& K^2\textstyle\sum_{k=1}^{n}[\boldsymbol{\lambda}(\boldsymbol{\theta}_0)^{'}\mathbf{g}(X_k,\boldsymbol{\theta}_0)]^2 \\
-\;& K\sqrt{n}\boldsymbol{\lambda}(\boldsymbol{\theta}_0)^{'}(\frac{\partial\boldsymbol{\Phi}_{\theta_0}}{\partial\boldsymbol{\theta}^{'}})\left[(\frac{\partial\boldsymbol{\Phi}_{\theta_0}}{\partial\boldsymbol{\theta}})^{'}\mathbf{S}_{n,\theta_0}^{-1}(\frac{\partial\boldsymbol{\Phi}_{\theta_0}}{\partial\boldsymbol{\theta}^{'}})\right]^{-1}(\frac{\partial\boldsymbol{\Phi}_{\theta_0}}{\partial\boldsymbol{\theta}^{'}})^{'}K\sqrt{n}\boldsymbol{\lambda}(\boldsymbol{\theta}_0) + o_p(1)\;.
\end{aligned}
$$

Then the second terms of (6.36) and (6.37) are asymptotically equivalent and it is straightforward to show that

$$
(6.38)\quad 2n\left[G_n(\hat{\boldsymbol{\theta}}_n)\right] = 2n\left[G_n(\boldsymbol{\theta}_0)\right] + (\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)^{'}(\frac{\partial\boldsymbol{\Phi}_{\theta_0}}{\partial\boldsymbol{\theta}^{'}})^{'}\mathbf{S}_{n,\theta_0}^{-1}(\frac{\partial\boldsymbol{\Phi}_{\theta_0}}{\partial\boldsymbol{\theta}^{'}})(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0) + o_p(1)\;,
$$

where

$$
2n\left[G_n(\boldsymbol{\theta}_0)\right] = \left[\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\frac{K}{\sqrt{m}}\mathbf{g}(X_k,\boldsymbol{\theta}_0)\right]^{'}\mathbf{S}_{n,\theta_0}^{-1}\left[\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\frac{K}{\sqrt{m}}\mathbf{g}(X_k,\boldsymbol{\theta}_0)\right] + o_p(1)\;.
$$

Hence by using the asymptotic normality of the MEL estimator in *Theorem 3.2*, we find that (4.39) converges to $\chi^2(q)$ distribution ($q=4$) as $n \longrightarrow +\infty$.
[*ii*] : In order to show the second part of *Theorem 2.3*, we define a $2m \times 2m$ matrix $\boldsymbol{\Sigma}_m(\boldsymbol{\theta}_0) = (K^2/m)\mathbf{E}_{\theta_0}[\mathbf{g}(X_1,\boldsymbol{\theta}_0)\mathbf{g}(X_1,\boldsymbol{\theta}_0)^{'}]$ and $2m \times 1$ random vectors $\mathbf{X}_n = (X_{nj})$ by

$$
(6.39)\qquad \mathbf{X}_m = \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\frac{K}{\sqrt{m}}\mathbf{g}(X_k,\boldsymbol{\theta}_0)\;.
$$

By using the silimar arguments as in the proof of *Theorem 2.2* and the above result of [*i*], it is possible to show that we can approximate the test statistic as $W_2 = \mathbf{X}_n' \mathbf{\Sigma}_m(\boldsymbol{\theta}_0)^{-1}\mathbf{X}_n + O_p(1)$. Then we rewrite

(6.40)

$$
\begin{aligned}
\frac{\mathbf{X}_n'\mathbf{\Sigma}_m(\boldsymbol{\theta}_0)^{-1}\mathbf{X}_n - 2m}{\sqrt{4m}} &= \frac{1}{\sqrt{4m}}\frac{1}{n}\sum_{i=i'=1}^{n}\sum_{j,k=1}^{2m}\left[\frac{K^2}{m} g_j(X_i,\theta_0)g_k(X_{i'},\theta_0)\sigma^{jk}(m) - \sigma^{jk}(m)\sigma_{jk}(m)\right] \\
&+ \frac{1}{\sqrt{4m}}\frac{1}{n}\sum_{i\neq i'=1}^{n}\sum_{j,k=1}^{2m}\left[\frac{K^2}{m} g_j(X_i,\boldsymbol{\theta}_0)g_k(X_{i'},\boldsymbol{\theta}_0)\sigma^{jk}(m)\right],
\end{aligned}
$$

where $\mathbf{g}(X_i, \boldsymbol{\theta}_0) = (g_j(X_i, \boldsymbol{\theta}_0))$ and $\mathbf{\Sigma}_m(\boldsymbol{\theta})^{-1} = (\sigma^{jk}(m))$ $(i = 1, \cdots, n; j, k = 1, \cdots, 2m)$. Since the expected values in the right-hand side of (6.40) are zeros and the variance of the first term is less than $(1/n)(1/4m)(2m)^2 = m/n$, it goes to zero as $n \to +\infty$. Because each elements of the second term in (6.40) are bounded, we can apply the Lindeberg central limit theorem to obtain

$$
\frac{\mathbf{X}_n'\mathbf{\Sigma}_m(\boldsymbol{\theta}_0)^{-1}\mathbf{X}_n - 2m}{\sqrt{4m}} \xrightarrow{d} N(0,1) \tag{6.41}
$$

as $n \to +\infty$. Hence we have the desired result.

(**Q.E.D**)

References

[1] Anderson, T.W. (2003) *An Introduction to Multivariate Statistical Analysis*, 3rd Edition, John-Wiley.

[2] Anderson, T.W. and H. Rubin (1949), "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, Vol. 20, 46-63.

[3] Bandorff-Nielsen, O. (1998), "Processes of Normal Inverse Gaussian Type," *Finance and Stochastics*, Vol. 2, 41-68.

[4] Bandorff-Nielsen, O., Mikosch, T., and Resnick, S. (2001), *Lévy Processes: Theory and Applications*, Birkhauser, Boston.

[5] Carr, P., Geman, H., Madane, D., and Yor, M. (2002), "The Fine Structure of Asset Returns : An Empirical Investigation," *Journal of Business*, Vol. 75, 305-332.

[6] Chambers, J.M., Mallows, C.L. and Stuck, B.W. (1976), "A Method for Simulating Stable Random Variables," *Journal of the American Statistical Association*, Vol.71, 340-344.

[7] DuMouchel, W.H. (1971), "Stable Distributions in Statistical Inference," Unpublished Ph.D. dissertation, Yale University.

[8] DuMouchel, W.H. (1973), "On the Asymptotic Normality of the Maximum Likelihood Estimate When Sampling From a Stable Distribution, " *The Annals of Statistics*, Vol. 1, 948-957.

[9] Fama, E.F. (1965), "The Behavior of Stock Market Prices," *Journal of Business*, Vol.38, 34-105.

[10] Fama, E.F. and Roll, R. (1968), "Some Properties of Symmetric Stable Distributions," *Journal of the American Statistical Association*, Vol. 63, 817-836.

[11] Fama, E.F. and Roll, R. (1971), "Parameter Estimates for Symmetric Stable Distributions," *Journal of the American Statistical Association*, Vol.66, 331-338.

[12] Fan, Jianqing, Chunming Zhang, and Zian Zhang (2001), "Generalized Likelihood Ratio Statistics and Wilks Phenomenon," *The Annals of Statistics*, Vol. 29, 153-193.

[13] Feller, W. (1971), *An Introduction to Probability Theory and Its Applications*, 2nd edition. John Wiley, New York.

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

[15] Feuerverger, A. and McDunnough, P. (1981b), "On some Fourier Methods for Inference, " *Journal of the American Statistical Association*, Vol.76, 379-387.

[16] Godambe, V.P. (1960), "An Optimum Property of Regular Maximum Likelihood Equation," *Annals of Mathematical Statistics*, Vol. 31, 1208-1211.

[17] Hansen, L. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, Vol. 50, 1029-1054.

[18] Hayashi, F. (2000), *Econometrics*, Princeton University Press.

[19] Hjort, N. L., I.W. McKeague, and I.V. Keilegom (2004), "Extending the Scope of Empirical Likelihood," Unpublished Manuscript.

[20] Kitamura, Y., Tripathi, G. and Ahn, H. (2001), "Empirical Likelihood-Based Inference in Conditional Moment Restriction Models," Discussion Paper CIRJE-F-124, Graduate School of Economics, University of Tokyo (http://www.e.u-tokyo.ac.jp/cirje/research).

[21] Kogon, S.M. and Williams, D.B. (1998), "Characteristic Function Based Estimation of Stable Distribution Parameters," *A Practical Guide to Heavy Tailed Data*, Birkhauser, Boston, MA, 311-335.

[22] Koutrouvelis I.A. (1980), "Regression-Type Estimation of the Parameters of Stable Laws," *Journal of the American Statistical Association*, Vol. 75, 918-928.

[23] Kunitomo,N. and Matsushita, Y. (2003), "Finite Sample Distributions of the Empirical Likelihood Estimator and the GMM Estimator," Discussion Paper CIRJE-F-200. Faculty of Economics, University of Tokyo (http://www.e.u-tokyo.ac.jp/cirje/research).

[24] Lukacs, E. (1970), *Characteristic Functions*, 2nd Edition. Griffin, London.

[25] Madan, D. and Seneta, E. (1990), "The Variance Gamma (V.G.) Model for Share Market Returns," *Journal of Business*, Vol. 63, 511-524.

[26] Mandelbrot, B. (1963), "The Variation of Certain Speculative Prices," *Journal of Business*, Vol. 36, 394-419.

[27] Miyahara, Y. (2002), "Estimation of Lévy Processes," Discussion Papers in Economics No.318, Department of Economics, Nagoya City University.

[28] MuCulloch, J.H. (1986), "Simple Consistent Estimators of Stable Distribution Parameters," *Communications in Statistics : Simulation and Computation*, Vol. 15, 1109-1136.

[29] Nolan, J.P. (1997), "Numerical Calculation of Stable Densities and Distribution Functions," *Communications in Statistics : Stochastic Models*, Vol. 13, 759-774.

[30] Nolan, J.P. (2001), "Maximum Likelihood Estimation and Diagnostics for Stable Distributions," in *Levy Processes: Theory and Applications*, Birkhauser, Boston, 379-400.

[31] Owen, A.B. (1988), "Empirical Likelihood Ratio Confidence Intervals for a Single Functional," *Biometrika*, Vol. 75, 237-249.

[32] Owen, A.B. (1990), "Empirical Likelihood Ratio Confidence Regions," *The Annals of Statistics*, Vol. 18, 90-120.

[33] Owen, A.B. (2001), *Empirical Likelihood*, Chapman and Hall.

[34] Paulson, A.S., Holcomb, E.W. and Leitch, R.A. (1975), "The Estimation of the Parameters of the Stable Laws," *Biometrika*, Vol. 62, 163-170.

[35] Press, S.J. (1972), "Estimation in Univariate and Multivariate Stable Distributions," *Journal of the American Statistical Association*, Vol. 67, 842-846.

[36] Qin, J. and Lawless, J. (1994), "Empirical Likelihood and General Estimating Equations," *The Annals of Statistics*, Vol. 22, 300-325.

[37] Samorodnitsky G. and Taqqu M.S. (1994), *Stable Non-Gaussian Random Processes*, Chapman and Hall, New York.

[38] Sato, K. (1999), *Lévy Processes and Infinitely Divisible Distributions*, Cambridge University Press.

[39] Shen, Xiaotong, Jian Shi, and Wing, Huang Wong (1999), "Random Sieve Likelihood and General Regression Models," *Journal of the American Statistical Association*, Vol. 94, 835-846.

[40] Zhang, J. and I. Gibels (2003), "Sieve Empirical Likelihood and Extension of the Generalized Least Squares," *Scandinavian Journal of Statistics*, Vol. 30, 1-24.

[41] Zolotarev, V.M. (1986), *One-Dimensional Stable Distributions*, American Mathematical Society, Providence, RI.