Empirical Likelihood Estimation of Lévy Processes [*]

Naoto Kunitomo[†]

and

Takashi Owada [‡]

December 2006

Summary

We propose an estimation procedure for the parameters of Lévy processes and the infinitely divisible distributions. The empirical likelihood method gives an easy way to estimate the parameters of the infinitely divisible distributions including the stable distributions. The maximum empirical likelihood estimator by using the empirical characteristic functions gives the consistency, the asymptotic normality, and the asymptotic efficiency when the number of restrictions is large. Test procedures and some extensions to the regression and estimating equations problems with the infinitely divisible disturbances are developed. A simple empirical example on the analysis of stock index returns in Japan is discussed.

Key Words

Lévy process, Infinitely Divisible Distribution, Stable Distribution, Empirical Likelihood, Empirical Characteristic Function, Regression and Estimating Equation.

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 over several decades. See Feller (1971), Zolotarev (1986), and Sato (1999) for the details of related problems in the probability literature. 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 earlier studies of the subject in the statistics literature. More recently, some applications of the more general Lévy processes and other classes of infinitely divisible distributions have been used 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 distributions and later MuCulloch (1986) has improved their method. Also another 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 classical 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 em-

pirical likelihood approach to the estimation problem of unknown parameters for stable distributions and the resulting computational burden is not very heavy. In particular, the maximum empirical likelihood (MEL) estimator for the parameters of stable distribution 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 regularity 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 more general estimation problem of Lévy processes and infinitely divisible distributions. Also we can apply our method to the regression and the general estimating equation problems 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 an important direction.

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 an empirical analysis of the stock index returns in Japan by the use of CGMY process. We give some concluding remarks in Section 6 and the proofs of main results are given in Section 7.

2. Empirical Likelihood Estimation of Stable Distribution

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 by Chamber et. al. (1976) for the class of stable distribution and it is represented as

$$\phi_\theta(t) = \phi_\theta^R(t) + i\phi_\theta^I(t) , \qquad (2.1)$$

where

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

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

and the parameter space is given by

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

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 H_θ .

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 in 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 standard techniques in the 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). Although the stable distributions do not necessarily have the first and second moments, we can utilize the information from the empirical characteristic function. We define the empirical likelihood function by

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

4

where p_k $(k = 1, \cdots, n)$ are the probability assigned to the data points of X_k. Without any further restriction except $p_k \geq 0$ and $\sum_{k=1}^{n} p_k = 1$, the empirical likelihood function $L_n(H_\theta)$ can be maximized at $p_k = 1/n$ $(k = 1, \cdots, n)$. Hence the empirical likelihood ratio function is given by $R_n(H_\theta) = \prod_{k=1}^{n} np_k$. Then we define the maximum empirical likelihood estimator $\hat{\boldsymbol{\theta}}_n$ for the vector of unknown coefficients by maximizing $R_n(H_\theta)$ under the restrictions :

$$
\mathcal{P}_n = \left\{ p_k \geq 0, \sum_{k=1}^{n} p_k = 1, \sum_{k=1}^{n} p_k \left(\cos(t_l X_k) - \phi_\theta^R(t_l) \right) = 0, \right.
$$
$$
\left. \sum_{k=1}^{n} p_k \left(\sin(t_l X_k) - \phi_\theta^I(t_l) \right) = 0 \ (k = 1, \cdots, n; l = 1, \ldots, m) \right\} . \tag{2.2}
$$

In the above restrictions m is the number of restrictions on the characteristic function (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.

Define $2m \times 1$ vectors

$$
\mathbf{g}(X_k, \boldsymbol{\theta}) = \left(\mathbf{g}^R(X_k, \boldsymbol{\theta})', \mathbf{g}^I(X_k, \boldsymbol{\theta})' \right)' , \tag{2.3}
$$

where $\mathbf{g}^R(X_k, \boldsymbol{\theta}) = \left(\cos(t_1 X_k) - \phi_\theta^R(t_1), \ldots, \cos(t_m X_k) - \phi_\theta^R(t_m) \right)'$, $\mathbf{g}^I(X_k, \boldsymbol{\theta}) = \left(\sin(t_1 X_k) - \phi_\theta^I(t_1), \ldots, \sin(t_m X_k) - \phi_\theta^I(t_m) \right)'$, $\phi_\theta^R(t_k)$ and $\phi_\theta^I(t_k)$ are given by (2.1) evaluated at $t = t_k$ $(k = 1, \cdots, m)$. Then we have the orthogonality condition $\mathbf{E}_{\theta_0}[\mathbf{g}(X, \boldsymbol{\theta}_0)] = \mathbf{0}$, where $\mathbf{E}_{\theta_0}[\,\cdot\,]$ is the expectation operator and $\boldsymbol{\theta}_0$ is the vector of true parameter values.

We assume 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 use the Lagrange form as

$$
L_n(\boldsymbol{\theta}) = \sum_{k=1}^{n} \log(np_k) - \mu \left(\sum_{k=1}^{n} p_k - 1 \right) - n\boldsymbol{\lambda}' \sum_{k=1}^{n} p_k \, \mathbf{g}(X_k, \boldsymbol{\theta}) , \tag{2.4}
$$

where μ and $\boldsymbol{\lambda} = (\lambda_{11}, \ldots, \lambda_{1m}, \lambda_{21}, \ldots, \lambda_{2m})'$ are 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 $\hat{\mu} = n$, $\hat{p}_k = (1/n)[1 + \boldsymbol{\lambda}' \mathbf{g}(X_k, \boldsymbol{\theta})]^{-1}$, 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 $\hat{p}_k > 0$ $(k = 1, \cdots, n)$, 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

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

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

$$l_n(\boldsymbol{\theta}) = \log \prod_{k=1}^{n} n\hat{p}_k = -\sum_{k=1}^{n} \log \left[1 + \boldsymbol{\lambda}' \mathbf{g}(X_k, \boldsymbol{\theta}) \right] . \tag{2.5}$$

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

2.2 Asymptotic Properties of MEL estimation

We shall investigate the asymptotic properties of the MEL estimator for $\boldsymbol{\theta}$. 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 regularity conditions. When the number of restrictions m is fixed, we have an analogous result in our situation, which is the starting point of subsequent developments.

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

$$R_n(\boldsymbol{\theta}) = \left\{ \prod_{k=1}^{n} np_k \mid \sum_{k=1}^{n} p_k \mathbf{g}(X_k, \boldsymbol{\theta}) = \mathbf{0}, p_k \geq 0, \sum_{k=1}^{n} p_k = 1 \right\}, \tag{2.6}$$

and a $2m \times 1$ vector of restrictions $\mathbf{g}(\cdot, \cdot)$ is defined by (2.3). Let the Lagrange multiplier $\hat{\boldsymbol{\lambda}}_n$ be the solution of

$$\frac{1}{n} \sum_{k=1}^{n} \frac{\mathbf{g}(X_k, \hat{\boldsymbol{\theta}}_n)}{1 + \boldsymbol{\lambda}' \mathbf{g}(X_k, \hat{\boldsymbol{\theta}}_n)} = \mathbf{0} . \tag{2.7}$$

6

Then as $n \to +\infty$

$$\sqrt{n}\left[\begin{array}{c} \hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \\ \hat{\boldsymbol{\lambda}}_n \end{array}\right] \xrightarrow{d} N_{4+2m}\left[\left(\begin{array}{c} \mathbf{0} \\ \mathbf{0} \end{array}\right), \left(\begin{array}{cc} \boldsymbol{\Omega}_m & \mathbf{O} \\ \mathbf{O} & \boldsymbol{\Gamma}_m \end{array}\right)\right], \qquad (2.8)$$

where

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

and we define a $2m \times 1$ vector $\boldsymbol{\Phi}_\theta = \left(\phi_\theta^R(t_1), \ldots, \phi_\theta^R(t_m), \phi_\theta^I(t_1), \ldots, \phi_\theta^I(t_m)\right)'$, a $2m \times 4$ matrix $\mathbf{B}_m(\boldsymbol{\theta}) = \left(\frac{\partial \boldsymbol{\Phi}_\theta}{\partial \boldsymbol{\theta}}\right)$, 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 the (i,j)-th elements of $\mathbf{A}_m(\boldsymbol{\theta})$ are given by

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

respectively.

The above statement 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 $\mathbf{g}(x, \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ and use a neighborhood $\mathbf{N}(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$ with some smoothness conditions. But it is rather straightforward to verify those conditions in our situation. For instance, we can utilize the condition that for $\forall \boldsymbol{\theta} \in \mathbf{N}(\boldsymbol{\theta}_0)$ (a compact set) 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 $\partial \mathbf{g}(x, \boldsymbol{\theta})/\partial \theta_j$ and $\partial^2 \mathbf{g}(x, \boldsymbol{\theta})/\partial \theta_j \partial \theta_k$ are continuous in $\mathbf{N}(\boldsymbol{\theta}_0)$, and both $\partial \mathbf{g}(x, \boldsymbol{\theta})/\partial \theta_j$ and $\partial^2 \mathbf{g}(x, \boldsymbol{\theta})/\partial \theta_j \partial \theta_k$ $(i, j = 1, \cdots, 4)$ are bounded in $\mathbf{N}(\boldsymbol{\theta}_0)$ $(\mathbf{N}(\boldsymbol{\theta}_0) \subset \boldsymbol{\Theta}_1)$.

Next we consider the density function of stable distribution $f_\theta(x)$ 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 $\mathrm{Int}(\boldsymbol{\Theta}_1)$ and for any

$\boldsymbol{\theta} \in \mathrm{Int}(\boldsymbol{\Theta}_1)$ it is twice continuously differentiable.

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

$$\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \int_{-\infty}^{\infty} f_\theta(x) dx = \int_{-\infty}^{\infty} \frac{\partial^2 f_\theta(x)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} dx \ , \tag{2.9}$$

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

$$\mathbf{I}(\boldsymbol{\theta}) = \mathbf{E}_\theta \left[\left(\frac{\partial \log f_\theta(X)}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial \log f_\theta(X)}{\partial \boldsymbol{\theta}} \right)' \right] = -\mathbf{E}_\theta \left[\frac{\partial^2 \log f_\theta(X)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right] \ . \tag{2.10}$$

(iii) For any $\boldsymbol{\theta} \in \mathrm{Int}(\boldsymbol{\Theta}_1)$, the Fisher Information matrix $\mathbf{I}(\boldsymbol{\theta})$ is non-singular. Hence for any non-zero vector $\mathbf{u} \in \mathbf{R}^4$ we have the inequality

$$\mathbf{u}'\mathbf{I}(\boldsymbol{\theta}_0)^{-1}\mathbf{u} \leq \mathbf{u}'\boldsymbol{\Omega}_m \mathbf{u} \ , \tag{2.11}$$

which implies that $\boldsymbol{\Omega}_m$ is positive definite.

The last inequality (2.11) implies that the asymptotic variance 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. For the estimation problem, we take some η $(0 < \eta < 1)$ and $m = m_n = [n^\eta]$, where $[c]$ is the largest integer not exceeding c. In order to impose m_n restrictions in the form of (2.2), we set $t_l = Kl/m_n$ $(l = 1, 2, \cdots, m_n)$ with some large constant K (> 0). 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 Section 7.

Theorem 2.2 : We assume that $X_1, \ldots, X_n$ are i.i.d. random variables with the stable distribution H_θ. We set $m = m_n = [n^{1/2-\epsilon}]$ $(0 < \epsilon < 1/2)$ and define $\hat{\boldsymbol{\theta}}_n = \mathrm{argmax}_\theta R_n(\boldsymbol{\theta})$, where $R_n(\boldsymbol{\theta})$ is given by (2.6).

(i) Assume that the true parameter vector $\boldsymbol{\theta}_0$ is in $\mathrm{Int}(\boldsymbol{\Theta}_2)$ and $\boldsymbol{\Theta}_2 = \{(\alpha, \beta, \gamma, \delta) : \epsilon \leq \alpha \leq 2, \ -1 \leq \beta \leq 1, \ \epsilon \leq \gamma \leq M, \ -M \leq \delta \leq M\}$ with ϵ (a sufficiently small positive number) and M (a sufficiently large positive number). Then as $n \longrightarrow \infty$

$$\hat{\boldsymbol{\theta}}_n \overset{p}{\longrightarrow} \boldsymbol{\theta}_0 \ . \tag{2.13}$$

(ii) We restrict the parameter space such that the vector of true parameter values $\boldsymbol{\theta}_0$ is in $\mathrm{Int}(\boldsymbol{\Theta}_1)$ and $\boldsymbol{\Theta}_1$ is the same as in *Theorem 2.1*. Then as $n \longrightarrow \infty$

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \overset{d}{\longrightarrow} N_4[0, \mathbf{J}_K(\boldsymbol{\theta}_0)] \ , \tag{2.14}$$

8

and

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

where $\mathbf{J}_K(\boldsymbol{\theta}_0) = \lim_{m \to \infty} \boldsymbol{\Omega}_m(\boldsymbol{\theta}_0)$ 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 $t_l = Kl/m_n$ $(l = 1, \cdots, m_n)$ and $\mathbf{I}(\boldsymbol{\theta}_0)$ is positive definite.

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,

$$\lim_{\alpha \to 1} \left| \frac{\partial^2 \phi_\theta(t)}{\partial \alpha^2} \right| \to +\infty \tag{2.15}$$

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

$$\lim_{\alpha \to 2} \frac{\partial \phi_\theta(t)}{\partial \beta} = \lim_{\alpha \to 2} \frac{\partial^2 \phi_\theta(t)}{\partial \beta^2} = 0 . \tag{2.16}$$

Then the vector of unknown parameters $\boldsymbol{\theta}$ is unidentified and the limiting information matrix is degenerate. (See Matsui and Takemura (2004), for instance.) It is not still clear if we have the asymptotic normality and the asymptotic efficiency of the MEL estimator in these boundary cases.

2.3 Empirical Likelihood Testing

It is also 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 proof is given in Section 7.

Theorem 2.3 : Suppose the assumptions in (ii) of *Theorem 2.2* hold and we set $m = m_n = [n^{1/2-\epsilon}]$ $(0 < \epsilon < 1/2)$.
(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 (2.5). Then

$$W_1 \xrightarrow{d} \chi^2(4) \tag{2.17}$$

as $n \longrightarrow +\infty$ when $\mathbf{H}_0$ is true.
(ii) For the hypothesis of the 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

$$\frac{W_2 - 2m}{\sqrt{4m}} \xrightarrow{d} N(0, 1) \tag{2.18}$$

9

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

The first part of *Theorem 2.3* allows 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 literature. 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.

3. Estimation of Lévy Processes, Regression and Estimating Equation Problems

3.1 Estimation of Lévy Processes

It is straightforward to apply our approach explained in Section 2 to several problems which have some difficulty in the standard procedures. For instance, 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} with $0 = v_0 \leq v_1 \leq \cdots \leq v_n \leq v$. For the notational convenience we take $v_i - v_{i-1} = 1$ $(i = 1, \cdots, n = v; v_0 = 0, \cdots, v_n = n)$ and write $Z_n = \sum_{i=1}^{n} X_i$. Then it has been well-known that the one-dimensional Lévy processes $\{Z_v\}$ and the infinitely divisible distributions for the random variables $\{X_i\}$ are completely determined by the characteristic function

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

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

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

and we denote $\mathbf{c}$ 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 possible 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 the parameterization of c_1 (> 0) and c_2 (> 0) is different from the one appeared in Section 2, however, there is one-to-one correspondence between the vectors $(\alpha, \beta, \gamma, \delta)$ in Section 2 and (α, b, c_1, c_2) (see Chapter 2 of 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 restrictions $C_0 > 0, G \geq 0, M \geq 0$, 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 see that for any finite t

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

as $Y \to 0$ or $Y \to 1$. Hence we should be careful to treat the boundary 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

11

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 a sequence of i.i.d. random variables with the characteristic function given by (3.1), which is continuous with respect to $\boldsymbol{\theta}$, and the Lévy measure ν_c is absolutely continuous with respect to the Lebesgue measure. The true parameter vector $\boldsymbol{\theta}_0$ is in $\mathrm{Int}(\boldsymbol{\Theta}_3)$, and $\boldsymbol{\Theta}_3$ is a compact subset such that (3.1) is the characteristic function of the infinitely divisible distribution with a non-degenerate density $f_\theta(\cdot)$ everywhere positive in $\mathbf{R}$. We impose $2m \times 1$ restriction functions as $\mathbf{g}(X_k, \boldsymbol{\theta})$ defined as (2.3) at $t_l = Kl/m_n$ $(l = 1, \cdots, m_n)$ with some positive constant K for the real part $\phi_\theta^R(t)$ and the imaginary part $\phi_\theta^I(t)$ of $\phi_\theta(t)$. We set $m = m_n = [n^{1/2-\epsilon}]$ $(0 < \epsilon < 1/2)$.
(i) Define $\hat{\boldsymbol{\theta}}_n = \mathrm{argmax}_\theta R_n(\boldsymbol{\theta})$ and $R_n(\theta)$ is given by (2.6). Then

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

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

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

where $\lim_{K \to +\infty} \mathbf{J}_K(\theta_0) = \mathbf{I}(\boldsymbol{\theta}_0)^{-1}$ and $\mathbf{J}_K(\theta_0)$ are defined by the corresponding quantities as in *Theorem 2.2*.

There could 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 $\nu_c(\cdot)$.

12

3.2 Regression and Estimating Equation Problems

We turn to the estimation problems of the nonlinear regression and 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 measurable 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 H_{θ_2} and $\boldsymbol{\theta}_2 = (\theta_{2k})$ is the vector of unknown parameters.

When there does not exists any endogenous variable (i.e. $\mathbf{y}_{2j}$) in the right-hand side of (3.9), we have the nonlinear regression model. When (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 $\mathbf{z}_j \ (j = 1, \cdots, n)$. The set of instrumental variables includes the vector of explanatory variables $\mathbf{z}_{1j}$ appeared in the structural equation of our interest as (3.9). The restrictions we impose on the real part and the imaginary part of the characteristic function are given such that for any m different points $t_1 < \cdots < t_m$,

$$\mathbf{E}\left[h_2(\mathbf{z}_j)(e^{itu_j} - \phi_{\theta_2}(t))\right] = \mathbf{0} \ (j = 1, \cdots, n) , \tag{3.10}$$

where $h_2(\cdot)$ 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 actually considering the limited information estimation method in econometrics. (See Anderson (2003) and Anderson and Rubin (1949) on the classical linear formulation of the related problems and see Anderson et. al. (2005) for the finite sample properties of the MEL estimator in the simple linear structural equation.)

As an important application of our general procedure, we shall consider the nonlinear regression and the estimating equation problems when the disturbance terms follow the class of infinitely divisible distributions. Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')'$ and $\boldsymbol{\theta}_2 = (\alpha, \beta, \gamma, \delta)'$ be the vectors of unknown parameters in the estimating equations

with the infinitely divisible disturbances. The maximum empirical likelihood (MEL) estimator for the vector of unknown parameters can be defined by maximizing the Lagrange form

$$
\begin{aligned}
L_n^*(\boldsymbol{\lambda}, \boldsymbol{\theta}) \;=\; & \sum_{j=1}^{n} \log(n\, p_j) - \mu \left(\sum_{j=1}^{n} p_j - 1 \right) - n \sum_{j=1}^{n} p_j h_2^{'}(\mathbf{z}_j) \\
& \times \left\{ \sum_{k=1}^{m} \boldsymbol{\lambda}_{1k} [\cos(t_k(y_{1j} - h_1(\mathbf{y}_{2j}, \mathbf{z}_{1j}, \boldsymbol{\theta}_1))) - \phi_{\theta_2}^R(t_k)] \right. \\
& \left. + \sum_{k=1}^{m} \boldsymbol{\lambda}_{2k} [\sin(t_k(y_{1j} - h_1(\mathbf{y}_{2j}, \mathbf{z}_{1j}, \boldsymbol{\theta}_1))) - \phi_{\theta_2}^I(t_k)] \right\} \; ,
\end{aligned}
\tag{3.11}
$$

where μ is a scalar Lagrange multiplier, and $\boldsymbol{\lambda}_{1k}$ and $\boldsymbol{\lambda}_{2k}$ $(k = 1, \cdots, m)$ 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 $\phi_{\theta_2}(t)$, 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

$$
\begin{aligned}
L_n(\boldsymbol{\lambda}, \boldsymbol{\theta}) \;=\; & -\sum_{j=1}^{n} \log \left(1 + h_2^{'}(\mathbf{z}_j) \left\{ \sum_{k=1}^{m} \boldsymbol{\lambda}_{1k} [\cos(t_k(y_{1j} - h_1(\mathbf{y}_{2j}, \mathbf{z}_{1j}, \boldsymbol{\theta}_1))) - \phi_{\theta_2}^R(t_k)] \right. \right. \\
& \left. \left. + \sum_{k=1}^{m} \boldsymbol{\lambda}_{2k} [\sin(t_k(y_{1j} - h_1(\mathbf{y}_{2j}, \mathbf{z}_{1j}, \boldsymbol{\theta}_1))) - \phi_{\theta_2}^I(t_k)] \right\} \right) ,
\end{aligned}
\tag{3.12}
$$

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

$$
\begin{aligned}
[n\hat{p}_j]^{-1} \;=\; & 1 + h_2^{'}(\mathbf{z}_j) \left\{ \sum_{k=1}^{m} \boldsymbol{\lambda}_{1k} [\cos(t_k(y_{1j} - h_1(\mathbf{y}_{2j}, \mathbf{z}_{1j}, \boldsymbol{\theta}_1))) - \phi_{\theta_2}^R(t_k)] \right. \\
& \left. + \sum_{k=1}^{m} \boldsymbol{\lambda}_{2k} [\sin(t_k(y_{1j} - h_1(\mathbf{y}_{2j}, \mathbf{z}_{1j}, \boldsymbol{\theta}_1))) - \phi_{\theta_2}^I(t_k)] \right\} \; .
\end{aligned}
\tag{3.13}
$$

By differentiating (3.12) with respect to $\boldsymbol{\lambda}^{'} = (\boldsymbol{\lambda}_{11}^{'}, \boldsymbol{\lambda}_{21}^{'}, \cdots, \boldsymbol{\lambda}_{1m}^{'}, \boldsymbol{\lambda}_{2m}^{'})$ 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 (2.2) in Section 2 can be interpreted as the simplest case of (3.10) in this section when $r = 0$, $l = 1$,

14

and $h_2(x) = 1$. Also if we set $y_j = y_{1j}$, $\mathbf{x}_j = \mathbf{z}_{1j}$ $(j = 1, \cdots, n)$, 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 has been discussed under the standard moment conditions on disturbance terms and the generalized method of moments by Hansen (1982) or the estimating equation method by Godambe (1960). The semi-parametric statistical estimation methods have been ususally applied. (See Hayashi (2000), for the details of standard results in the recent econometrics literature.) 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.

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 of the stable distribution by using the method of Chamber, Mallows and Stuck (1976). After some experiments, we have 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 $(\alpha, \beta) \in (0, 1.8) \times [-1, 1]$. When $(\alpha, \beta) \in [1.8, 2) \times [-1, 1]$, however, we have had sometimes slow convergence when we had imposed the restrictions 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 $\hat{\gamma} = \hat{\gamma}^{(0)}\hat{\gamma}^{(1)}$, $\hat{\delta} = \hat{\delta}^{(0)} + \hat{\delta}^{(1)}\hat{\gamma}^{(0)}$ as the final estimates of the parameters γ and δ .

Although in our experiments we have set the sample size $n = 1000$, we can estimate the key parameters satisfactorily as far as the cases when $n \geq 100$ by imposing the restrictions at only 5 points.

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

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 was obtained numerically by DuMouchel (1971) and Nolan (2001). We can 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. we have summarized our numerical results on efficiency in Table 2 and we have found that there are not many extreme cases when we have low efficiency and our estimation method gives reasonable values in most cases. When $\alpha = 2$ and $\beta \neq 0$, there is an identification problem and

we have confirmed that some instability in numerical computations would occur without any restrictions on the parameter space.

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

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

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

where θ_1 is the unknown (scalar) 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 (true) 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 the true parameter value $\theta_1 = 1.0$, 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,

Table 3: Simulation Results for Regression

We set $\beta = \delta = 0.0$ and set $\boldsymbol{\theta} = (\theta_1, \alpha, \gamma)'$ 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

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 RMSE 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 performance in all cases of our simulation.

5. An empirical example of the stock index returns in Japan

As an empirical example, we discuss an analysis of the stock index returns R_i ($i = 1, \cdots, n$) in Japan. We have applied the stable distribution, the CGMY process and also the AR models with the stable disturbances and the CGMY disturbances. In our analysis we have used the daily data of TOPIX, which is the major stock index in Tokyo from March 20, 1990 to August 16, 2005. In each case we took $m = 7$ as the restrictions on the empirical characteristic function.

First, we estimated the stable distribution with four parameters and found a significant departure of the distributions of log-returns from the Gaussian distribution. The estimated parameters and the standard errors are $\hat{\alpha} = 1.6747$ (0.0292),

18

$\hat{\gamma} = 0.7395$ (0.0121), $\hat{\beta} = 0.0210$ (0.0880), $\hat{\delta} = -0.0227$ (0.0223). This agrees with a wide spread observation on the distributions of daily returns on the stock markets. Second, we have estimated the CGMY process which includes some stable distributions as special cases. The estimated parameters and the standard errors are $\hat{C} = 0.5633$ (0.1610), $\hat{G} = 0.6794$ (0.1841), $\hat{M} = 0.7107$ (0.1843), $\hat{Y} = 0.8696$ (0.2270). All estimated parameters of the CGMY process are statistically significant and the formulation of stable distributions for the stock index returns as well as the Gaussian distribution may not be appropriate because the estimates of G and M are quite different from 0.

Since there are non-negligible autocorrelations in the TOPIX returns data, we have fitted the second order AR model and the disturbances follow the CGMY process with the restriction $G = M$. (We have examined several different AR models and the AR(2) model was chosen by minimizing the resulting empirical AIC.) The estimated parameters with the restriction and the standard errors are

$$R_i = -.0231 R_{i-1} - .2219 R_{i-2} + u_i \tag{5.1}$$
$$(0.0165) \quad (0.0165)$$

and $\hat{C} = 0.4490$ (0.1401), $\hat{G} = 0.5486$ (0.1820), $\hat{Y} = 1.1242$ (0.2158). The resulting estimates are fully consistent with our observations.

However, we have found that when we use finer intervals for stock returns, the estimate of α becomes smaller in its magniutude, for instance. This contradicts with the class of stable distributions for stock returns. Although there have been some empirical findings that the stable distribution may be appropriate for the stock returns data, it seems that the CGMY model may be better in some sense. Our empirical investigation suggests that the CGMY process is a good candidate for applications. Our empirical analysis here, however, is still at the preliminary stage and a fuller investigation will be needed on the related problems.

6. 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 and the asymptotic normal 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. We can use the testing procedures based on the empirical likelihood ratio statistics in these situations.

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 directly possible to apply our approach to the nonlinear regression and the estimating equation problems with 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 advantage over other methods.

Finally, we should mention that our estimation method is so simple that the results can be extended to some 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.

7. Proof of Theorems

In this mathematical appendix we give the proofs of *Theorem 2.2* and *Theorem 2.3* in Section 2. We first show the consistency of the MEL estimator, and then prove its asymptotic normality and asymptotic efficiency when m is large. We set $m = m_n = [n^{1/2-\epsilon}]$ $(0 < \epsilon < 1/2)$. 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 Int$(\boldsymbol{\Theta})$ and each elements of $\mathbf{A}_m(\boldsymbol{\theta})$ and $\mathbf{B}_m(\boldsymbol{\theta})$ are bounded.

We first prepare two lemmas (*Lemma 1* and *Lemma 2*), which are needed for the

proof of *Theorems.*

Lemma 1 : For a sufficiently large K (> 0), as $n \to \infty$

$$\lambda_{\min} [\boldsymbol{\Sigma}_n(\boldsymbol{\theta}_0)] > 0 , \tag{7.1}$$

where $\lambda_{\min}[\, \cdot \,]$ is the minimum characteristic root of the $2m \times 2m$ matrix

$$\boldsymbol{\Sigma}_n(\boldsymbol{\theta}_0) = \frac{K^2}{m} \mathbf{E}_{\theta_0} \left[\mathbf{g}(X_1, \boldsymbol{\theta}_0) \mathbf{g}(X_1, \boldsymbol{\theta}_0)' \right] . \tag{7.2}$$

Proof of Lemma 1: Consider the L^2-class of functions $c(x)$ such that $0 < \int_{-\infty}^{\infty} c(x)^2 dx < \infty$ and define the Fourier tansform by

$$c^*(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} c(x) e^{-itx} dx .$$

Then we take $c^*(t_l)$ for $t_l = Kl/m$ $(l = -m, \cdots, -1, 1, \cdots, m)$ such that $0 < \frac{1}{m} \sum_{l=-m}^{m} |c^*(t_l)|^2 < \infty$. We also define a $2m \times 1$ complex vector by $\tilde{\mathbf{C}}_m = [c^*(t_1), \cdots, c^*(t_m), c^*(-t_1), \cdots, c^*(-t_m)]'$ and $\tilde{\mathbf{g}}(X, \boldsymbol{\theta}) = [e^{it_1 X} - \phi_\theta(t_1), \cdots, e^{it_m X} - \phi_\theta(t_m), e^{-it_1 X} - \phi_\theta(-t_1), \cdots, e^{-it_m X} - \phi_\theta(-t_m)]'$. Then

$$
\begin{aligned}
& \frac{K^2}{m^2} \tilde{\mathbf{C}}_m' \mathbf{E}_{\theta_0} \left[\tilde{\mathbf{g}}(X, \boldsymbol{\theta}_0) \tilde{\mathbf{g}}(X, \boldsymbol{\theta}_0)' \right] \tilde{\mathbf{C}}_m \\
= \; & \frac{K^2}{m^2} \mathbf{E}_{\theta_0} \left[\sum_{l,l'=-m, l\neq 0, l' \neq 0}^{m} c^*(t_l) c^*(t_{l'}) (e^{it_l X} - \phi_{\theta_0}(t_l))(e^{it_{l'} X} - \phi_{\theta_0}(t_{l'})) \right] \\
= \; & \frac{K^2}{m^2} \sum_{l,l'=-m, l\neq 0, l' \neq 0}^{m} [c^*(t_l) c^*(t_{l'}) (\phi_{\theta_0}(t_l + t_{l'}) - \phi_{\theta_0}(t_l) \phi_{\theta_0}(t_{l'}))] \\
\longrightarrow \; & \int_{-K}^{K} \int_{-K}^{K} \{\phi_{\theta_0}(s + t) - \phi_{\theta_0}(s) \phi_{\theta_0}(t)\} c^*(s) c^*(t) ds dt
\end{aligned}
\tag{7.3}
$$

as $m = m_n \longrightarrow \infty$ $(n \longrightarrow \infty)$. For sufficiently large K,

$$
\begin{aligned}
A \; = \; & \int_{-K}^{K} \int_{-K}^{K} \phi_{\theta_0}(s + t) c^*(s) c^*(t) ds dt \\
\sim \; & \int_t \int_x \left[\frac{1}{2\pi} \int_s e^{-i(s+t)x} \phi_{\theta_0}(s + t) ds \right] \times e^{itx} c(x) c^*(t) dx dt \\
= \; & \int_x c(x)^2 f(x) dx ,
\end{aligned}
$$

21

and

$$B = \left[\int_{-K}^{K}\int_{-K}^{K}\phi_{\theta_0}(s)c^*(s)ds\right]^2 \sim \left\{\int_x\left[\frac{1}{2\pi}\int_s\phi_{\theta_0}(s)e^{-isx}ds\right]c(x)dx\right\}^2$$

$$= \left[\int_x c(x)f(x)dx\right]^2 .$$

We use the relation $(\int_t a(t)b(t)dt)^2 = \int_s a(s)^2 ds \int_t b(t)^2 dt - (1/2)\int_s\int_t[a(s)b(t) - a(t)b(s)]^2 dsdt$ and set $a(s) = c(s)\sqrt{f(s)}$ and $b(t) = \sqrt{f(t)}$. Then for sufficiently large K,

$$\begin{aligned}
A - B &\sim \frac{1}{2}\int_x\int_y[c(x) - c(y)]^2[\sqrt{f(x)f(y)}]^2 dxdy \\
&= \frac{1}{2}\int_{x,y}[c(x) - c(y)]^2 f(x)f(y)dxdy > 0
\end{aligned}$$

(7.4)

because $c(x)$ $(\in L^2)$ cannot be constant.

Finally, we use the fact that we can set the complex vectors $\tilde{\mathbf{C}}_m$ and $\tilde{\mathbf{g}}(X,\boldsymbol{\theta}_0)$ such that the corrsponding real vectors are given by

$$\mathbf{g}(X,\boldsymbol{\theta}_0) = \left[\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2i} & -\frac{1}{2i} \end{pmatrix}\otimes \mathbf{I}_m\right]\tilde{\mathbf{g}}(X,\boldsymbol{\theta}_0), \ \mathbf{C}_m = \left[\begin{pmatrix} \frac{1}{2} & \frac{1}{2i} \\ \frac{1}{2} & -\frac{1}{2i} \end{pmatrix}\otimes \mathbf{I}_m\right]^{-1}\tilde{\mathbf{C}}_m .$$

Q.E.D.

Lemma 2 : Under the assumptions of *Theorem 2.2*, we have

$$\lim_{K\to\infty}\mathbf{J}_K(\boldsymbol{\theta}_0) = \lim_{K\to\infty}\lim_{n\to\infty}\left[\mathbf{B}_m(\boldsymbol{\theta}_0)'\mathbf{A}_m(\boldsymbol{\theta}_0)^{-1}\mathbf{B}_m(\boldsymbol{\theta}_0)\right]^{-1} = \mathbf{I}(\boldsymbol{\theta}_0)^{-1} \qquad (7.5)$$

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 2: We define a $2m \times 1$ complex vector $\tilde{\boldsymbol{\Phi}}_\theta$ by $\tilde{\boldsymbol{\Phi}}_\theta = [\phi_\theta(t_1),\cdots,\phi_\theta(t_m),\phi_\theta(-t_1),\cdots,\phi_\theta(-t_m)]'$. Then we find that

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

Let $\mathbf{w}_\theta(t)$ be

$$\mathbf{w}_\theta(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\partial\log f_\theta(x)}{\partial\boldsymbol{\theta}}e^{-itx}dx , \qquad (7.6)$$

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

22

convergence as

$$\frac{K}{m}\tilde{\boldsymbol{W}}'_{\theta_0}\frac{\partial\tilde{\boldsymbol{\Phi}}_{\theta_0}}{\partial\boldsymbol{\theta}'} = \frac{K}{m}\sum_{l=-m,l\neq 0}^{m}\mathbf{w}_{\theta_0}(t_l)\left(\frac{\partial\phi_{\theta_0}(t_l)}{\partial\boldsymbol{\theta}}\right)' \longrightarrow \int_{-K}^{K}\mathbf{w}_{\theta_0}(t)\left(\frac{\partial\phi_{\theta_0}(t)}{\partial\boldsymbol{\theta}}\right)' dt \quad (7.7)$$

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

$$\frac{K^2}{m^2}\tilde{\boldsymbol{W}}'_{\theta_0}\mathbf{E}_{\theta_0}[\tilde{\mathbf{g}}(X,\boldsymbol{\theta}_0)\tilde{\mathbf{g}}(X,\boldsymbol{\theta}_0)']\tilde{\boldsymbol{W}}_{\theta_0}$$
$$\tag{7.8}$$
$$\longrightarrow \int_{-K}^{K}\int_{-K}^{K}\{\phi_{\theta_0}(s+t)-\phi_{\theta_0}(s)\phi_{\theta_0}(t)\}\mathbf{w}_{\theta_0}(s)\mathbf{w}'_{\theta_0}(t)dsdt$$

as $n\to\infty$. Hence

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

$$= \left\{\int_{-K}^{K}\mathbf{w}_{\theta_0}(t)\left(\frac{\partial\phi_{\theta_0}(t)}{\partial\boldsymbol{\theta}}\right)'dt\right\}'\left\{\int_{-K}^{K}\int_{-K}^{K}\{\phi_{\theta_0}(s+t)-\phi_{\theta_0}(s)\phi_{\theta_0}(t)\}\mathbf{w}_{\theta_0}(s)\mathbf{w}'_{\theta_0}(t)dsdt\right\}^{-1}$$

$$\times\left\{\int_{-K}^{K}\mathbf{w}_{\theta_0}(t)\left(\frac{\partial\phi_{\theta_0}(t)}{\partial\boldsymbol{\theta}}\right)'dt\right\}.$$
$$\tag{7.9}$$

We denote the RHS of (7.9) as $\boldsymbol{\Xi}_K(\boldsymbol{\theta}_0)$. Then for any (non-degenerate) $2m\times 4$ matrix $\mathbf{v}$ we have

$$\left[\left(\mathbf{v}'\frac{\partial\tilde{\boldsymbol{\Phi}}_{\theta_0}}{\partial\boldsymbol{\theta}'}\right)'\{\mathbf{v}'\mathbf{E}_{\theta_0}[\tilde{\mathbf{g}}(X,\boldsymbol{\theta}_0)\tilde{\mathbf{g}}(X,\boldsymbol{\theta}_0)']\mathbf{v}\}^{-1}\left(\mathbf{v}'\frac{\partial\tilde{\boldsymbol{\Phi}}_{\theta_0}}{\partial\boldsymbol{\theta}'}\right)\right]^{-1} \tag{7.10}$$

can be minimized at

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

and the minimum value is

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

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

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

$$\leq \mathbf{u}'\left[\left(\tilde{\boldsymbol{W}}'_{\theta_0}\frac{\partial\tilde{\boldsymbol{\Phi}}_{\theta_0}}{\partial\boldsymbol{\theta}'}\right)'\{\tilde{\boldsymbol{W}}'_{\theta_0}\mathbf{E}_{\theta_0}[\tilde{\mathbf{g}}(X,\boldsymbol{\theta}_0)\tilde{\mathbf{g}}(X,\boldsymbol{\theta}_0)']\tilde{\boldsymbol{W}}_{\theta_0}\}^{-1}\left(\tilde{\boldsymbol{W}}'_{\theta_0}\frac{\partial\tilde{\boldsymbol{\Phi}}_{\theta_0}}{\partial\boldsymbol{\theta}'}\right)\right]^{-1}\mathbf{u},$$
$$\tag{7.11}$$

23

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 and then we obtain the desired result. **Q.E.D.**

Proof of Theorem 2.2 :

$[i]$ **Consistency**: We take a sufficiently large K (> 0) and set $m = m_n = [n^{1/2-\epsilon}]$ $(0 < \epsilon < 1/2)$. We use the fact that

$$\frac{K}{m}\|\mathbf{E}_{\theta_0}[\mathbf{g}(X,\boldsymbol{\theta})]\|^2 = \int_0^K \left[(\phi_{\theta_0}^R(t) - \phi_\theta^R(t))^2 + (\phi_{\theta_0}^I(t) - \phi_\theta^I(t))^2\right] dt + o(1) , \quad (7.12)$$

where the first term of the right hand side is denoted by $\boldsymbol{\Gamma}_2(K,\boldsymbol{\theta}_0,\boldsymbol{\theta})$. Define a criterion function by

$$G_n(\boldsymbol{\theta}) = -\frac{1}{n}\sum_{k=1}^n \log\left[1 + \boldsymbol{\lambda}'(\boldsymbol{\theta})\mathbf{g}(X_k,\boldsymbol{\theta})\right] \tag{7.13}$$

and a function

$$\mathbf{u}(\boldsymbol{\theta}) = \frac{\sqrt{\frac{K}{m}}\mathbf{E}_{\theta_0}[\mathbf{g}(X,\boldsymbol{\theta})]}{1 + \|\sqrt{\frac{K}{m}}\mathbf{E}_{\theta_0}[\mathbf{g}(X,\boldsymbol{\theta})]\|} . \tag{7.14}$$

For any $\delta > 0$ we take a neighborhood $\mathbf{N}(\boldsymbol{\theta}_0,\delta)$ and then

$$\sup_{\boldsymbol{\theta}^* \in \boldsymbol{\Theta} \setminus \boldsymbol{N}(\boldsymbol{\theta}_0,\delta)} \left|\sqrt{\frac{K}{m}}\mathbf{E}_{\theta_0}\left[-\mathbf{u}'(\boldsymbol{\theta}^*)\mathbf{g}(X,\boldsymbol{\theta}^*)\right] - \left[\frac{-\boldsymbol{\Gamma}_2(K,\boldsymbol{\theta}_0,\boldsymbol{\theta}^*)}{1 + \sqrt{\boldsymbol{\Gamma}_2(K,\boldsymbol{\theta}_0,\boldsymbol{\theta}^*)}}\right]\right| = o(1) . \tag{7.15}$$

We set $\beta = 1/2$ and then by using Taylor's Theorem, there exists a $t \in (0,1)$ such that

$$\sup_{\boldsymbol{\theta}^* \in \boldsymbol{\Theta}} \left|n^\beta\sqrt{\frac{K}{m}}\frac{1}{n}\sum_{k=1}^n\{-\log\left[1 + n^{-\beta}\mathbf{u}'(\boldsymbol{\theta}^*)\mathbf{g}(X_k,\boldsymbol{\theta}^*)\right]\} - \sqrt{\frac{K}{m}}\frac{1}{n}\sum_{k=1}^n\left[-\mathbf{u}'(\boldsymbol{\theta}^*)\mathbf{g}(X_k,\boldsymbol{\theta}^*)\right]\right|$$

$$\leq \sup_{\boldsymbol{\theta}^* \in \boldsymbol{\Theta}} \sqrt{\frac{K}{m}}\frac{1}{n}\left|\sum_{k=1}^n \frac{n^{-\beta}(\mathbf{u}'(\boldsymbol{\theta}^*)\mathbf{g}(X_k,\boldsymbol{\theta}^*))^2}{2\left[1 + tn^{-\beta}\mathbf{u}'(\boldsymbol{\theta}^*)\mathbf{g}(X_k,\boldsymbol{\theta}^*)\right]^2}\right| = o(1)$$

because we have $n^{-\beta}|\mathbf{u}'(\boldsymbol{\theta}^*)\mathbf{g}(X_k,\boldsymbol{\theta}^*)| \leq 2\sqrt{2}n^{-\beta}\sqrt{m} \to 0$ and $m^{-1/2}[n^{-\beta}\mathbf{u}'(\boldsymbol{\theta}^*)\mathbf{g}(X_k,\boldsymbol{\theta}^*)]^2 \to 0$ as $n \longrightarrow \infty$. For any $\epsilon > 0$,

$$\mathbf{P}\left(\sup_{\boldsymbol{\theta}^* \in \boldsymbol{\Theta}} \left|\sqrt{\frac{K}{m}}\frac{1}{n}\sum_{k=1}^n\left[-\mathbf{u}'(\boldsymbol{\theta}^*)\mathbf{g}(X_k,\boldsymbol{\theta}^*)\right] - \sqrt{\frac{K}{m}}\mathbf{E}_{\theta_0}\left[-\mathbf{u}'(\boldsymbol{\theta}^*)\mathbf{g}(X,\boldsymbol{\theta}^*)\right]\right| > \epsilon\right)$$

$$\leq \frac{1}{\epsilon^2}\mathbf{E}_{\theta_0}\left[\sup_{\boldsymbol{\theta}^* \in \boldsymbol{\Theta}} \frac{K}{m}\frac{1}{n^2}\left|\mathbf{u}'(\boldsymbol{\theta}^*)\sum_{k=1}^n(\mathbf{g}(X_k,\boldsymbol{\theta}^*) - \mathbf{E}_{\theta_0}(\mathbf{g}(X,\boldsymbol{\theta}^*)))\right|^2\right] = o(1) .$$

24

Then, by using (7.12) and (7.15), for any $\delta > 0$ there exists n_0 and $H(\delta) > 0$ such that for all $n \geq n_0$,

$$\mathbf{P}\left(\sup_{\boldsymbol{\theta}^* \in \boldsymbol{\Theta} \setminus \boldsymbol{N}(\boldsymbol{\theta}_0, \delta)} n^\beta \sqrt{\frac{K}{m}} \frac{1}{n} \sum_{k=1}^n \{-\log\left[1 + n^{-\beta} \mathbf{u}'(\boldsymbol{\theta}^*) \mathbf{g}(X_k, \boldsymbol{\theta}^*)\right]\} > -H(\delta)\right) < \frac{\delta}{2} .$$

$$(7.16)$$

Since

$$\sup_{\boldsymbol{\theta}^* \in \boldsymbol{\Theta} \setminus \boldsymbol{N}(\boldsymbol{\theta}_0, \delta)} \frac{1}{n} \sum_{k=1}^n \{-\log\left[1 + n^{-\beta} \mathbf{u}'(\boldsymbol{\theta}^*) \mathbf{g}(X_k, \boldsymbol{\theta}^*)\right]\} \geq \sup_{\boldsymbol{\theta}^* \in \boldsymbol{\Theta} \setminus \boldsymbol{N}(\boldsymbol{\theta}_0, \delta)} G_n(\boldsymbol{\theta}^*) ,$$

we have

$$\mathbf{P}\left(n^\beta \sqrt{\frac{K}{m}} \sup_{\boldsymbol{\theta}^* \in \boldsymbol{\Theta} \setminus \boldsymbol{N}(\boldsymbol{\theta}_0, \delta)} G_n(\boldsymbol{\theta}^*) > -H(\delta)\right) < \frac{\delta}{2} . \qquad (7.17)$$

Now we investigate the stochastic order of the Lagrange multipliers at the true value, which is the solution of

$$\frac{1}{n} \sum_{k=1}^n \frac{\mathbf{g}(X_k, \boldsymbol{\theta}_0)}{1 + \boldsymbol{\lambda}'(\boldsymbol{\theta}_0) \mathbf{g}(X_k, \boldsymbol{\theta}_0)} = \mathbf{0} \qquad (7.18)$$

and we write $\boldsymbol{\lambda}(\boldsymbol{\theta}_0) = \|\boldsymbol{\lambda}(\boldsymbol{\theta}_0)\| \boldsymbol{\xi}$ and $\boldsymbol{\xi}$ is the $2m \times 1$ unit vector. By multiplying $\boldsymbol{\xi}'$ from the left hand side to (7.18),

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

and we have the inequality

$$\|\boldsymbol{\lambda}(\boldsymbol{\theta}_0)\| \frac{1}{1 + \|\boldsymbol{\lambda}(\boldsymbol{\theta}_0)\| \max_{1 \leq k \leq n} \|\mathbf{g}(X, \boldsymbol{\theta}_0)\|} \frac{K^2}{mn} \sum_{k=1}^n \boldsymbol{\xi}' \mathbf{g}(X_k, \boldsymbol{\theta}_0) \mathbf{g}(X_k, \boldsymbol{\theta}_0)' \boldsymbol{\xi}$$

$$\leq \left| \frac{K^2}{mn} \sum_{k=1}^n \boldsymbol{\xi}' \mathbf{g}(X_k, \boldsymbol{\theta}_0) \right| .$$

Since

$$\mathbf{E}_{\theta_0}\left[\left| \frac{K^2}{mn} \sum_{k=1}^n \boldsymbol{\xi}' \mathbf{g}(X_k, \boldsymbol{\theta}_0) \right|^2\right] \leq \left(\frac{K^2}{mn}\right)^2 \mathbf{E}_{\theta_0}\left[\left\| \sum_{k=1}^n \mathbf{g}(X_k, \boldsymbol{\theta}_0) \right\|^2\right] = O\left(\frac{1}{mn}\right) ,$$

25

and by the use of *Lemma 1*,

$$\|\boldsymbol{\lambda}(\boldsymbol{\theta}_0)\| = O_p\left(\sqrt{\frac{1}{mn}}\right) \tag{7.19}$$

and

$$G_n(\boldsymbol{\theta}_0) \geq -\frac{1}{n}\sum_{k=1}^{n}\boldsymbol{\lambda}'(\boldsymbol{\theta}_0)\mathbf{g}(X_k, \boldsymbol{\theta}_0) = -\|\boldsymbol{\lambda}(\boldsymbol{\theta}_0)\|\frac{1}{n}\sum_{k=1}^{n}\boldsymbol{\xi}'\mathbf{g}(X_k, \boldsymbol{\theta}_0) , \tag{7.20}$$

which is of the order $O_p(m^{-1/2}n^{-1/2})O_p(\sqrt{m/n}) = O_p(n^{-1})$. Hence for nay $\delta > 0$ there exists $H'(\delta) > 0$ such that for all $n \geq n_1$,

$$\mathbf{P}\left(\sqrt{\frac{K}{m}}n^\beta G_n(\boldsymbol{\theta}_0) < -H'(\delta)\right) < \frac{\delta}{2} . \tag{7.21}$$

Thus by combining (7.17) and (7.21), we have $\mathbf{P}\left(\hat{\boldsymbol{\theta}}_n \notin \mathbf{N}(\boldsymbol{\theta}_0, \delta)\right) < \delta$ and it 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 $\boldsymbol{\theta}_0 = \hat{\boldsymbol{\theta}}_n$, we have

$$-\sqrt{n}\frac{\partial G_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} = \frac{\partial^2 G_n(\boldsymbol{\theta}_n^\dagger)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) , \tag{7.22}$$

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

$$\begin{aligned}-\sqrt{n}\frac{\partial G_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} &= \sqrt{n}\frac{1}{n}\sum_{k=1}^{n}\frac{1}{1+\boldsymbol{\lambda}'(\boldsymbol{\theta}_0)\mathbf{g}(X_k, \boldsymbol{\theta}_0)}\left(\frac{\partial\mathbf{g}(X_k, \boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}\right)'\boldsymbol{\lambda}(\boldsymbol{\theta}_0) \\ &= \sqrt{n}\frac{1}{n}\sum_{k=1}^{n}\frac{1}{1+\boldsymbol{\lambda}'(\boldsymbol{\theta}_0)\mathbf{g}(X_k, \boldsymbol{\theta}_0)}\left(-\frac{\partial\boldsymbol{\Phi}_{\theta_0}}{\partial\boldsymbol{\theta}'}\right)'\boldsymbol{\lambda}(\boldsymbol{\theta}_0) .\end{aligned} \tag{7.23}$$

By rewriting (7.18) with respect to the Lagrange multipliers, we have

$$\mathbf{0} = \frac{1}{n}\sum_{k=1}^{n}\frac{K^2}{m}\mathbf{g}(X_k, \boldsymbol{\theta}_0) - \left\{\frac{1}{n}\sum_{k=1}^{n}\frac{K^2}{m}\mathbf{g}(X_k, \boldsymbol{\theta}_0)\mathbf{g}(X_k, \boldsymbol{\theta}_0)'\right\}\boldsymbol{\lambda}(\boldsymbol{\theta}_0) + r_{1n} , \tag{7.24}$$

where we set the remainder term

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

26

By using *Lemma 1* and (7.19), we find that $\|r_{1n}\|$ is of the order $O_p(n^{-1}m^{-1/2})$. Since we have the condition $m_n/n \longrightarrow 0$, we can approximate the random vector $-\sqrt{n}\partial G_n(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}$ as

$$-\sqrt{n}\frac{\partial G_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} = \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\left(-\frac{K}{\sqrt{m}}\frac{\partial\Phi_{\theta_0}}{\partial\boldsymbol{\theta}'}\right)'\mathbf{S}_n(\boldsymbol{\theta}_0)^{-1}\frac{K}{\sqrt{m}}\mathbf{g}(X_k,\boldsymbol{\theta}_0) + o_p(1) , \quad (7.25)$$

where

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

In order to show the asymptotic normality of (7.25), for any constant vector $\boldsymbol{\zeta}\in\mathbf{R}^4$ we define a set of random variables X_{nk} $(k=1,\ldots,n)$ by

$$X_{nk} = \boldsymbol{\zeta}'\left(-\frac{K}{\sqrt{m}}\frac{\partial\Phi'_{\theta_0}}{\partial\boldsymbol{\theta}'}\right)[\boldsymbol{\Sigma}_n(\boldsymbol{\theta}_0)]^{-1}\frac{K}{\sqrt{m}}\mathbf{g}(X_k,\boldsymbol{\theta}_0) .$$

Then by using the Lindberge central limit theorem,

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n}X_{nk} \xrightarrow{d} N(0,\boldsymbol{\zeta}'\mathbf{I}_K(\boldsymbol{\theta}_0)\boldsymbol{\zeta}) , \qquad (7.26)$$

where $\mathbf{I}_K(\boldsymbol{\theta}_0)$ $(=\mathbf{J}_K(\boldsymbol{\theta}_0)^{-1}) = \lim_{n\to\infty}[\mathbf{B}_m(\boldsymbol{\theta}_0)'\mathbf{A}_m(\boldsymbol{\theta}_0)^{-1}\mathbf{B}_m(\boldsymbol{\theta}_0)]$ (See *Lemma 2*). For an arbitrary vector $\boldsymbol{\zeta}$, we consider the stochastic order of

$$r_{2n} = \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\left(-\frac{K}{\sqrt{m}}\frac{\partial\Phi_{\theta_0}}{\partial\boldsymbol{\theta}'}\right)'(\mathbf{S}_n(\boldsymbol{\theta}_0)^{-1} - \boldsymbol{\Sigma}_n(\boldsymbol{\theta}_0)^{-1})\frac{K}{\sqrt{m}}\mathbf{g}(X_k,\boldsymbol{\theta}_0) . \qquad (7.27)$$

Then we need to evaluate $\|\mathbf{S}_n(\boldsymbol{\theta}_0)^{-1}\|\|\mathbf{S}_n(\boldsymbol{\theta}_0) - \boldsymbol{\Sigma}_n(\boldsymbol{\theta}_0)\|\|\boldsymbol{\Sigma}_n(\boldsymbol{\theta}_0)^{-1}\|$. For this purpose we take $\delta = \epsilon/2$ (> 0) and then $\|\mathbf{S}_n(\boldsymbol{\theta}_0) - \boldsymbol{\Sigma}_n(\boldsymbol{\theta}_0)\| = o_p(n^{-1/2+\delta})$ and there exists a η (> 0) such that $\|\boldsymbol{\Sigma}_n(\boldsymbol{\theta}_0)\|^{-1} \leq \eta\sqrt{m}$ by *Lemma 1*. Hence we have $\|\mathbf{S}_n(\boldsymbol{\theta}_0)^{-1}\| = \|\{\boldsymbol{\Sigma}_n(\boldsymbol{\theta}_0) + [\mathbf{S}_n(\boldsymbol{\theta}_0) - \boldsymbol{\Sigma}_n(\boldsymbol{\theta}_0)]\}^{-1}\| = O_p(m^{1/2})$ and $\|r_{2n}\| \leq O_p(mn^{-(1/2-\delta)})$. Then r_{2n} is $o_p(1)$ since we have set $m_n = n^{1/2-\epsilon}$ $(0 < \epsilon < 1/2)$. By gathering these evaluations, $-\sqrt{n}\partial G_n(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}$ converges to the normal distribution with the variance-covariance matrix $\mathbf{I}_K(\boldsymbol{\theta}_0)$ $(=\mathbf{J}_K(\boldsymbol{\theta}_0)^{-1})$ as $n \longrightarrow +\infty$.

Next for the second derivatives we shall show

$$\frac{\partial^2 G_n(\boldsymbol{\theta}_n^\dagger)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} = -\mathbf{I}_K(\boldsymbol{\theta}_0) + o_p(1) . \qquad (7.28)$$

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

$$
\begin{aligned}
\frac{\partial^2 G_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} &= \frac{1}{n} \sum_{k=1}^{n} \frac{\left(-\partial \boldsymbol{\Phi}_\theta / \partial \boldsymbol{\theta}'\right)' \boldsymbol{\lambda}(\boldsymbol{\theta}) \boldsymbol{\lambda}'(\boldsymbol{\theta}) \left(-\partial \boldsymbol{\Phi}_\theta / \partial \boldsymbol{\theta}'\right)}{[1 + \boldsymbol{\lambda}'(\boldsymbol{\theta}) \mathbf{g}(X_k, \boldsymbol{\theta})]^2} \\
&+ \frac{1}{n} \sum_{k=1}^{n} \frac{\left(-\partial \boldsymbol{\Phi}_\theta / \partial \boldsymbol{\theta}'\right)' \boldsymbol{\lambda}(\boldsymbol{\theta}) \mathbf{g}(X_k, \boldsymbol{\theta})' \left(\partial \boldsymbol{\lambda}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}'\right)}{[1 + \boldsymbol{\lambda}'(\boldsymbol{\theta}) \mathbf{g}(X_k, \boldsymbol{\theta})]^2} \\
&- \frac{1}{n} \sum_{k=1}^{n} \frac{\left(-\partial \boldsymbol{\Phi}_\theta / \partial \boldsymbol{\theta}'\right)' \left(\partial \boldsymbol{\lambda}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}'\right)}{1 + \boldsymbol{\lambda}'(\boldsymbol{\theta}) \mathbf{g}(X_k, \boldsymbol{\theta})} - \frac{1}{n} \sum_{k=1}^{n} \sum_{l=1}^{2m} \frac{\lambda_l(\boldsymbol{\theta}) (\partial^2 g_l(X_k, \boldsymbol{\theta}) / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}')}{1 + \boldsymbol{\lambda}'(\boldsymbol{\theta}) \mathbf{g}(X_k, \boldsymbol{\theta})}.
\end{aligned}
\tag{7.29}
$$

There are some complications partly because we have 2nd and 3rd terms with the derivatives of the Lagrange multipliers.

By differentiating $(1/n) \sum_{k=1}^{n} \mathbf{g}(X_k, \boldsymbol{\theta}) / [1 + \boldsymbol{\lambda}'(\boldsymbol{\theta}) \mathbf{g}(X_k, \boldsymbol{\theta})] = \mathbf{0}$ with respect to $\boldsymbol{\theta}$, we have the relation

$$
\begin{aligned}
&\left[\frac{1}{n} \sum_{k=1}^{n} \frac{K^2}{m} \frac{\mathbf{g}(X_k, \boldsymbol{\theta}) \mathbf{g}(X_k, \boldsymbol{\theta})'}{[1 + \boldsymbol{\lambda}'(\boldsymbol{\theta}) \mathbf{g}(X_k, \boldsymbol{\theta})]^2} \right] \frac{\partial \boldsymbol{\lambda}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \\
&= \frac{1}{n} \sum_{k=1}^{n} \frac{K}{m} \frac{\left(-\partial \boldsymbol{\Phi}_\theta / \partial \boldsymbol{\theta}'\right)}{1 + \boldsymbol{\lambda}'(\boldsymbol{\theta}) \mathbf{g}(X_k, \boldsymbol{\theta})} - \frac{1}{n} \sum_{k=1}^{n} \frac{K^2}{m} \frac{\mathbf{g}(X_k, \boldsymbol{\theta}) \boldsymbol{\lambda}'(\boldsymbol{\theta}) \left(-\partial \boldsymbol{\Phi}_\theta / \partial \boldsymbol{\theta}'\right)}{(1 + \boldsymbol{\lambda}'(\theta) \mathbf{g}(X_k, \boldsymbol{\theta}))^2}.
\end{aligned}
\tag{7.30}
$$

For the last term of (7.30), we find

$$
\begin{aligned}
&\left\| \frac{1}{n} \sum_{k=1}^{n} \frac{(-\partial \boldsymbol{\Phi}_\theta / \partial \boldsymbol{\theta}')' \boldsymbol{\lambda}(\boldsymbol{\theta}) \mathbf{g}(X_k, \boldsymbol{\theta})'}{\left[1 + \boldsymbol{\lambda}'(\boldsymbol{\theta}) \mathbf{g}(X_k, \boldsymbol{\theta})\right]^2} \right\| \\
&\leq \max_{1 \leq k \leq n} \frac{1}{[1 + \boldsymbol{\lambda}'(\boldsymbol{\theta}) \mathbf{g}(X_k, \boldsymbol{\theta})]^2} \left\| \frac{\partial \boldsymbol{\Phi}_\theta}{\partial \boldsymbol{\theta}'} \right\| \|\boldsymbol{\lambda}(\boldsymbol{\theta})\| \max_{1 \leq k \leq n} \|\mathbf{g}(X_k, \boldsymbol{\theta})\|,
\end{aligned}
$$

is of the order $o_p(1)$. Hence the second term of (7.29) is of the order $o_p(1)$. Similarly, for the first term and the fourth term of (7.29) are dominated by

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

and

$$
\max_{1 \leq k \leq n} \frac{1}{|1 + \boldsymbol{\lambda}'(\boldsymbol{\theta}) \mathbf{g}(X_k, \boldsymbol{\theta})|} \left\| \sum_{l=1}^{2m} \lambda_l(\boldsymbol{\theta}) \frac{\partial^2 g_l(X_k, \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right\|,
$$

respectively. By evaluating the stochastic orders of these terms as before, we find that they are of the order $o_p(1)$.

As the dominant term, we need to evaluate the third term of (7.29). After straightforward (but tedious) evaluations on the 3rd term of (7.29), it is possible to show that

$$
\left\| \left[\frac{1}{n} \sum_{k=1}^{n} \frac{\left(-\partial \boldsymbol{\Phi}_\theta / \partial \boldsymbol{\theta}'\right)'}{1 + \boldsymbol{\lambda}'(\boldsymbol{\theta}) \mathbf{g}(X_k, \boldsymbol{\theta})} \right] \times \left[\frac{1}{n} \sum_{k=1}^{n} \frac{K^2}{m} \frac{\mathbf{g}(X_k, \boldsymbol{\theta}) \mathbf{g}(X_k, \boldsymbol{\theta})'}{[1 + \boldsymbol{\lambda}'(\boldsymbol{\theta}) \mathbf{g}(X_k, \boldsymbol{\theta})]^2} \right]^{-1} \right.
$$
$$
\left. \times \left[\frac{1}{n} \sum_{k=1}^{n} \frac{K^2}{m} \frac{\mathbf{g}(X_k, \boldsymbol{\theta}) \boldsymbol{\lambda}'(\boldsymbol{\theta}) \left(-\partial \boldsymbol{\Phi}_\theta / \partial \boldsymbol{\theta}'\right)}{[1 + \boldsymbol{\lambda}'(\boldsymbol{\theta}) \mathbf{g}(X_k, \boldsymbol{\theta})]^2} \right] \right\| = o_p(1) \ .
$$

Then for the third term of (7.29), we find

$$
\left\| \frac{1}{n} \sum_{k=1}^{n} \frac{\left(-\partial \boldsymbol{\Phi}_\theta / \partial \boldsymbol{\theta}'\right)' \left(\partial \boldsymbol{\lambda}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}'\right)}{1 + \boldsymbol{\lambda}'(\boldsymbol{\theta}) \mathbf{g}(X_k, \boldsymbol{\theta})} - \mathbf{I}_K(\boldsymbol{\theta}_0) \right\| = o_p(1) \tag{7.31}
$$

and $\mathbf{I}_K(\boldsymbol{\theta}_0) = \mathbf{J}_K(\boldsymbol{\theta}_0)^{-1}$. By using (7.26) and (7.28), we have established the asymptotic normality

$$
\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, \mathbf{J}_K(\boldsymbol{\theta}_0)) \ . \tag{7.32}
$$

Finally, by taking sufficiently large $K > 0$ and use *Lemma 2*, we have that $\lim_{K \to \infty} \mathbf{J}_K(\boldsymbol{\theta}_0) = \mathbf{I}(\boldsymbol{\theta}_0)^{-1}$, which is positive definite. **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 problem given by Owen (1990), and Qin and Lawless (1994) except the fact that the number of restrictions $m = m_n$ increases as $n \to \infty$. The precise evaluations of stochastic orders in our derivations are straightforward as in the proof of *Theorem 2.2*.

Let $Y_k(\boldsymbol{\theta}) = \boldsymbol{\lambda}(\boldsymbol{\theta})' \mathbf{g}(X_k, \boldsymbol{\theta})$ $(k = 1, \cdots, n)$. Then the criterion function $G_n(\hat{\boldsymbol{\theta}}_n)$ at the MEL estimation can be re-written as

$$
G_n(\hat{\boldsymbol{\theta}}_n) = -\frac{1}{n} \sum_{k=1}^{n} \left[Y_k(\hat{\boldsymbol{\theta}}_n) - \frac{1}{2} Y_k(\hat{\boldsymbol{\theta}}_n)^2 + \frac{1}{3} Y_k(\boldsymbol{\theta}^*)^3 \right] \ , \tag{7.33}
$$

where we have $\|\boldsymbol{\theta}^* - \hat{\boldsymbol{\theta}}_n\| \le \|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|$. We first approximate $Y_k(\hat{\boldsymbol{\theta}}_n)$ by the sum of four terms as

$$
\boldsymbol{\lambda}(\boldsymbol{\theta}_0)' \mathbf{g}(X_k, \boldsymbol{\theta}_0) + (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' \left(\partial \boldsymbol{\lambda} / \partial \boldsymbol{\theta}'\right)' \mathbf{g}(X_k, \boldsymbol{\theta}_0)
$$
$$
+ \ \boldsymbol{\lambda}(\boldsymbol{\theta}_0)' \left(-\partial \boldsymbol{\Phi}_{\theta_0} / \partial \boldsymbol{\theta}'\right) (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' \left(\partial \boldsymbol{\lambda} / \partial \boldsymbol{\theta}'\right)' \left(-\partial \boldsymbol{\Phi}_{\theta_0} / \partial \boldsymbol{\theta}'\right) (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)
$$

29

because the remaining terms are of smaller orders. Also we notice the fact that

$$\sum_{k=1}^{n} |Y_k(\boldsymbol{\theta}^*)|^3 \leq \sum_{k=1}^{n} \|\boldsymbol{\lambda}(\boldsymbol{\theta}^*)'\|^3 \|\mathbf{g}(X_k, \boldsymbol{\theta}^*)\|^3 ,$$

which is of the order $O_p(n^{-1/2})$. Then by using (7.22), (7.25),(7.28), (7.30) and (7.31), $\sum_{k=1}^{n} Y_k(\hat{\boldsymbol{\theta}}_n)$ can be approximated by

$$\sum_{k=1}^{n} Y_k(\hat{\boldsymbol{\theta}}_n) = \sqrt{n}\boldsymbol{\lambda}(\boldsymbol{\theta}_0)' \left[\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \mathbf{g}(X_k, \boldsymbol{\theta}_0) \right]$$

$$+ \left[\sqrt{n}\boldsymbol{\lambda}(\boldsymbol{\theta}_0)' \frac{\partial \boldsymbol{\Phi}_{\theta_0}}{\partial \boldsymbol{\theta}'} \right] \left[\left(\frac{K}{\sqrt{m}} \frac{\partial \boldsymbol{\Phi}_{\theta_0}}{\partial \boldsymbol{\theta}} \right)' \boldsymbol{\Sigma}_n(\boldsymbol{\theta}_0)^{-1} \left(\frac{K}{\sqrt{m}} \frac{\partial \boldsymbol{\Phi}_{\theta_0}}{\partial \boldsymbol{\theta}'} \right) \right]^{-1} \left(\frac{\partial \boldsymbol{\Phi}_{\theta_0}}{\partial \boldsymbol{\theta}'} \right)' \left[\sqrt{n}\boldsymbol{\lambda}(\boldsymbol{\theta}_0) \right]$$

$$+ \quad o_p(1).$$

$$(7.34)$$

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

$$\sum_{k=1}^{n} Y_k(\hat{\boldsymbol{\theta}}_n)^2 = \sum_{k=1}^{n} \left[\boldsymbol{\lambda}(\boldsymbol{\theta}_0)' \mathbf{g}(X_k, \boldsymbol{\theta}_0) \right]^2$$

$$+ \left[\sqrt{n}\boldsymbol{\lambda}(\boldsymbol{\theta}_0)' \frac{\partial \boldsymbol{\Phi}_{\theta_0}}{\partial \boldsymbol{\theta}'} \right] \left[(\frac{K}{\sqrt{m}} \frac{\partial \boldsymbol{\Phi}_{\theta_0}}{\partial \boldsymbol{\theta}})' \boldsymbol{\Sigma}_m(\boldsymbol{\theta}_0)^{-1} (\frac{K}{\sqrt{m}} \frac{\partial \boldsymbol{\Phi}_{\theta_0}}{\partial \boldsymbol{\theta}'}) \right]^{-1} (\frac{\partial \boldsymbol{\Phi}_{\theta_0}}{\partial \boldsymbol{\theta}'})' \left[\sqrt{n}\boldsymbol{\lambda}(\boldsymbol{\theta}_0) \right]$$

$$+ \quad o_p(1).$$

$$(7.35)$$

Since the second terms of (7.34) and (7.35) are asymptotically equivalent, it is straightforward to show that

$$2n \left[G_n(\hat{\boldsymbol{\theta}}_n) \right] = 2n \left[G_n(\boldsymbol{\theta}_0) \right] + (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' (\frac{K}{\sqrt{m}} \frac{\partial \boldsymbol{\Phi}_{\theta_0}}{\partial \boldsymbol{\theta}'}) \boldsymbol{\Sigma}_n(\boldsymbol{\theta}_0)^{-1} (\frac{K}{\sqrt{m}} \frac{\partial \boldsymbol{\Phi}_{\theta_0}}{\partial \boldsymbol{\theta}'})' (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + o_p(1) ,$$

$$(7.36)$$

where $2n \left[G_n(\boldsymbol{\theta}_0) \right] = \mathbf{X}_n' \boldsymbol{\Sigma}_n(\boldsymbol{\theta}_0)^{-1} \mathbf{X}_n + o_p(1)$ and $\mathbf{X}_n = (1/\sqrt{n}) \sum_{k=1}^{n} (K/\sqrt{m}) \mathbf{g}(X_k, \boldsymbol{\theta}_0)$. Then by using the asymptotic normality of the MEL estimator in *Theorem 2.2*, we find that (2.17) converges to $\chi^2(q)$ distribution $(q = 4)$ as $n \to +\infty$.

$[ii]$: By using the same arguments in the proof of *Theorem 2.2*, we approximate

30

the test statistic as $W_2 = \mathbf{X}'_n \mathbf{\Sigma}_n(\boldsymbol{\theta}_0)^{-1} \mathbf{X}_n + o_p(1)$. Then we rewrite

$$\frac{\mathbf{X}'_n \mathbf{\Sigma}_n(\boldsymbol{\theta}_0)^{-1} \mathbf{X}_n - 2m}{\sqrt{4m}}$$

$$= \frac{1}{\sqrt{4m}} \frac{1}{n} \sum_{i=1}^{n} \sum_{j,k=1}^{2m} \left[\frac{1}{m} g_j(X_i, \boldsymbol{\theta}_0) g_k(X_i, \boldsymbol{\theta}_0) \sigma^{jk}(n) - \sigma^{jk}(n) \sigma_{jk}(n) \right] \qquad (7.37)$$

$$+ \frac{1}{\sqrt{4m}} \frac{1}{n} \sum_{i \neq i'=1}^{n} \sum_{j,k=1}^{2m} \left[\frac{K^2}{m} g_j(X_i, \boldsymbol{\theta}_0) g_k(X_{i'}, \boldsymbol{\theta}_0) \sigma^{jk}(n) \right] ,$$

where we denote $\mathbf{g}(X_i, \boldsymbol{\theta}_0) = (g_j(X_i, \boldsymbol{\theta}_0))$ and $\mathbf{\Sigma}_n(\boldsymbol{\theta})^{-1} = (\sigma^{jk}(n))$ $(i = 1, \cdots, n; j, k = 1, \cdots, 2m)$.

The expected values of the right-hand side of (7.37) are zeros and the variance of the first term is less than $(1/n)^2 (1/4m)(2m)^2 = m/n^2$ and it goes to zero as $n \to +\infty$. Because each elements of the second term in (7.37) are bounded, we can apply the central limit theorem to obtain

$$\frac{\mathbf{X}'_n \mathbf{\Sigma}_n(\boldsymbol{\theta}_0)^{-1} \mathbf{X}_n - 2m}{\sqrt{4m}} \xrightarrow{d} N(0, 1) \qquad (7.38)$$

as $n \to +\infty$. By applying the similar arguments as in the proof of *Theorem 2.2*, we have the desired result. **Q.E.D**

References

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

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.

Anderson, T.W., N. Kunitomo, and Y. Matsushita (2005), "A New Light from Old Wisdoms : Alternative Estimation Method of Simultaneous Equations and Microeconometric Models," Discussion Paper CIRJE-F-321, Graduate School of Economics, University of Tokyo (http://www.e.u-tokyo.ac.jp/cirje/research/dp/2005).

Bandorff-Nielsen, O. (1998), "Processes of Normal Inverse Gaussian Type," *Finance*

and Stochastics, Vol. 2, 41-68.

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

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.

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.

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

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.

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

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

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

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

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.

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

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

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

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

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.

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

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

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

Matsui, M. and A. Takemura (2004), "Some Improvements in Numerical Evaluation of Symmetric Stable Density and its Derivatives," Discussion Paper CIRJE-F-292, Graduate School of Economics, University of Tokyo.

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

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

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

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

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

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

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.

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

Qin, J. and Lawless, J. (1994), "Empirical Likelihood and General Estimating Equa-

tions," *The Annals of Statistics*, Vol. 22, 300-325.

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

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