

Chapter 2

Empirical Likelihood Approaches for Financial Returns

Abstract We deal with an empirical likelihood and apply it to several financial problems. Empirical likelihood is one of the nonparametric methods of statistical inference. It allows us to use likelihood methods although we do not assume that the data comes from a known family. Consequently, the empirical likelihood has both effectiveness and flexibility of the likelihood method, and reliability of the nonparametric methods. The construction of this chapter is as follows. We briefly look at the history of the empirical likelihood in Sect. 2.1 and review its method for i.i.d. data in Sect. 2.2. The frequency domain approach of empirical likelihood for multivariate non-Gaussian linear processes is discussed in Sect. 2.3. Section 2.4 gives extensions of the empirical likelihood such as Cressie-Read power-divergence statistic and generalized empirical likelihood. Section 2.5 considers application of the generalized empirical likelihood to an inference problem for multivariate stable distributions. Technical proofs of the theorems are given in Sect. 2.6.

Keywords Empirical likelihood · Non-Gaussian linear process · Frequency domain · Portfolio · Generalized empirical likelihood · Stable distribution

2.1 Introduction

Empirical likelihood is originally proposed by Owen (1988, 1990) and its overview is found in Owen (2001). Because of its advantages, many researchers investigated the empirical likelihood. For example, Qin and Lawless (1994) linked the estimating functions and equations to empirical likelihood. Newey and Smith (2004) considered generalized empirical likelihood, which is a richer class than empirical likelihood, and gave its higher order properties. For the application of empirical likelihood to the dependent data, Kitamura (1997) proposed blockwise empirical likelihood for the weakly dependent processes. Monti (1997) gave the frequency domain approach to the empirical likelihood methods for linear processes and Ogata and Taniguchi (2010)

extended it to the multivariate one. Ogata and Taniguchi (2009) investigated the asymptotic properties of Cressie-Read divergence statistics, which encompasses the empirical likelihood, for multivariate linear processes. A frequency domain empirical likelihood for the processes including long-range dependence is considered by Nordman and Lahiri (2006).

2.2 Empirical Likelihood Method for i.i.d. Data

Given $X_1, \dots, X_n \in \mathbb{R}^m$, assumed independent with common distribution function F_0 , the nonparametric likelihood of distribution function F is defined by

$$L(F) = \prod_{i=1}^n F(\{X_i\}), \quad (2.1)$$

where $F(\{X_i\})$ is the probability of getting the value X_i in a sample from F . Apparently, only the distributions which have the positive point mass probability on each observation constitute positive nonparametric likelihoods. Therefore, we restrict F to the one having the probability $p_i = F(\{X_i\}) > 0$ on each observation X_i . By a simple calculation, we find the maximizer of the nonparametric likelihood (2.1) turns to be the empirical distribution function F_n , placing probability $1/n$ on each observation. Therefore, similar to the parametric case, nonparametric likelihood ratio of F to the maximizer F_n is defined by:

$$R(F) = \frac{L(F)}{L(F_n)} = \frac{\prod_{i=1}^n p_i}{\prod_{i=1}^n 1/n} = \prod_{i=1}^n (np_i).$$

Suppose that we are interested in a parameter $\theta \in \Theta \subset \mathbb{R}^p$ and that the true value θ_0 satisfies the following estimating equation

$$E[\mathbf{m}(X; \theta_0)] = \mathbf{0} \quad (2.2)$$

where $\mathbf{m}(X; \theta) \in \mathbb{R}^q$ is a vector-valued function, called estimating function. As examples of estimating functions, we take $\mathbf{m}(X; \theta) = X - \theta$ to indicate a vector mean by Eq. (2.2). For $\Pr(X \in A)$, we take $\mathbf{m}(X; \theta) = 1_{X \in A} - \theta$. For a continuously distributed scalar X and $\theta \in \mathbb{R}$, the function $m(X; \theta) = 1_{X < \theta} - \alpha$ defines θ as the α quantile of X . Now we define the empirical likelihood ratio function for θ by

$$\mathcal{R}(\theta) = \max_p \left\{ \prod_{i=1}^n (np_i) \left| \sum_{i=1}^n p_i \mathbf{m}(X_i; \theta) = \mathbf{0}, p_i \geq 0, \sum_{i=1}^n p_i = 1 \right. \right\}, \quad (2.3)$$

where $\mathbf{p} = (p_1, \dots, p_n)$. This is the maximum of the nonparametric likelihood ratio with the restriction that the mean of the estimating function is zero under the distribution F . $\mathcal{R}(\boldsymbol{\theta})$ is calculated by the Lagrange multiplier method as follows. Write

$$G = \sum_{i=1}^n \log(np_i) - n\boldsymbol{\xi}' \sum_{i=1}^n p_i \mathbf{m}(X_i; \boldsymbol{\theta}) + \gamma \left(\sum_{i=1}^n p_i - 1 \right),$$

where $\boldsymbol{\xi} \in \mathbb{R}^q$ and $\gamma \in \mathbb{R}$ are Lagrange multipliers. Setting $\partial G / \partial p_i = 0$ gives

$$\frac{\partial G}{\partial p_i} = \frac{1}{p_i} - n\boldsymbol{\xi}' \mathbf{m}(X_i; \boldsymbol{\theta}) + \gamma = 0.$$

Therefore, the equation $\sum_{i=1}^n p_i (\partial G / \partial p_i) = 0$ gives $\gamma = -n$. Then, we may write

$$p_i = \frac{1}{n} \frac{1}{1 + \boldsymbol{\xi}' \mathbf{m}(X_i; \boldsymbol{\theta})}$$

where the vector $\boldsymbol{\xi} = \boldsymbol{\xi}(\boldsymbol{\theta})$ satisfies q equations given by

$$\frac{1}{n} \sum_{i=1}^n \frac{\mathbf{m}(X_i; \boldsymbol{\theta})}{1 + \boldsymbol{\xi}' \mathbf{m}(X_i; \boldsymbol{\theta})} = \mathbf{0}. \quad (2.4)$$

Finally, we may write

$$\mathcal{R}(\boldsymbol{\theta}) = \prod_{i=1}^n (np_i) = \prod_{i=1}^n \frac{1}{1 + \boldsymbol{\xi}' \mathbf{m}(X_i; \boldsymbol{\theta})}. \quad (2.5)$$

The following theorem is due to Owen (2001).

Theorem 2.1 (Owen (2001, Theorem 3.4)) *Assume that $\text{Var}(\mathbf{m}(X; \boldsymbol{\theta}_0))$ is finite and has rank r . Then $-2 \log \mathcal{R}(\boldsymbol{\theta}_0) \rightarrow \chi_{(r)}^2$ in distribution as $n \rightarrow \infty$.*

2.3 Estimation with Frequency Domain Empirical Likelihood for Stationary Processes

Here we are concerned with the m -dimensional linear process $\{X(t)\}$ in (1.10) and consider the problem of estimating parameter $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$. Suppose that the information of $\boldsymbol{\theta}$ exists through a system of general estimating equations in frequency domain as follows. Let $\boldsymbol{\phi}_j(\lambda; \boldsymbol{\theta})$, ($j = 1, \dots, q$) be $m \times m$ matrix-valued continuous functions on $[-\pi, \pi]$ satisfying $\boldsymbol{\phi}_j(\lambda; \boldsymbol{\theta}) = \boldsymbol{\phi}_j(\lambda; \boldsymbol{\theta})^*$ and $\boldsymbol{\phi}_j(-\lambda; \boldsymbol{\theta}) = \boldsymbol{\phi}_j(\lambda; \boldsymbol{\theta})'$. We assume that each $\boldsymbol{\phi}_j(\lambda; \boldsymbol{\theta})$ satisfies the spectral moment condition

$$\int_{-\pi}^{\pi} \text{tr}\{\phi_j(\lambda; \theta_0) f(\lambda)\} d\lambda = 0 \quad (j = 1, \dots, q) \quad (2.6)$$

where $\theta_0 = (\theta_{10}, \dots, \theta_{p0})'$ is the true value of parameter and $f(\lambda)$ is the true spectral density matrix of $X(t)$. By taking an appropriate function for $\phi_j(\lambda; \theta)$, the θ_0 specified by Eq. (2.6) can express various important indices for time series.

As an estimating function, we consider the following quantity

$$\mathbf{m}(\lambda_t; \theta) = \left(\text{tr}\{\phi_1(\lambda_t; \theta) \mathbf{I}_n(\lambda_t)\}, \dots, \text{tr}\{\phi_q(\lambda_t; \theta) \mathbf{I}_n(\lambda_t)\} \right)' \quad (2.7)$$

where $\lambda_t = (2\pi t)/n$, $(t = 1, \dots, n)$ and $\mathbf{I}_n(\lambda)$ is the periodogram matrix. Following (2.3), we define the empirical likelihood ratio function for frequency domain by

$$\tilde{\mathcal{R}}(\theta) = \max_p \left\{ \prod_{t=1}^n (np_t) \left| \sum_{t=1}^n p_t \mathbf{m}(\lambda_t; \theta) = \mathbf{0}, p_t \geq 0, \sum_{t=1}^n p_t = 1 \right| \right\}. \quad (2.8)$$

Now we impose the following assumption.

Assumption 2.1

- (i) $\{X(t)\}$ satisfies (B) in Proposition 1.2.
- (ii) For the sequence $\{C_k\}$ defined by

$$C_k = \sup_{a_1 \dots a_k} \sum_{t_1, \dots, t_{k-1} = -\infty}^{\infty} |c_{a_1 \dots a_k}^X(t_1, \dots, t_{k-1})|,$$

it holds

$$\sum_{k=1}^{\infty} \frac{C_k z^k}{k!} < \infty$$

for z in a neighborhood of zero. Here, $c_{a_1 \dots a_k}^X(t_1, \dots, t_{k-1})$ is the k -th-order cumulant of X defined in Chap. 1.

Applying Proposition 1.2, we obtain the following theorem.

Theorem 2.2 (Ogata and Taniguchi (2010)) *Let Assumption 2.1 hold. Then*

$$-2 \log \tilde{\mathcal{R}}(\theta_0) \rightarrow (\Sigma N)'(\Sigma N),$$

where N has a q -dimensional standard normal distribution and $\Sigma = \Sigma_2^{-1/2} \Sigma_1^{1/2}$. Here Σ_1 and Σ_2 are constant $q \times q$ matrices whose (i, j) -th components are

$$\begin{aligned} & \frac{1}{\pi} \int_{-\pi}^{\pi} \text{tr}\{\mathbf{f}(\lambda)\boldsymbol{\phi}_i(\lambda; \boldsymbol{\theta}_0)\mathbf{f}(\lambda)\boldsymbol{\phi}_j(\lambda; \boldsymbol{\theta}_0)\}d\lambda \\ & + \frac{1}{2\pi} \sum_{r,t,u,v=1}^m \int_{-\pi}^{\pi} \phi_{rt}^{(i)}(\lambda_1; \boldsymbol{\theta}_0)\phi_{uv}^{(j)}(\lambda_2; \boldsymbol{\theta}_0)Q_{rtuv}^X(-\lambda_1, \lambda_2, -\lambda_2)d\lambda_1d\lambda_2 \end{aligned}$$

and

$$\begin{aligned} & \frac{1}{2\pi} \left[\int_{-\pi}^{\pi} \text{tr}\{\mathbf{f}(\lambda)\boldsymbol{\phi}_i(\lambda; \boldsymbol{\theta}_0)\mathbf{f}(\lambda)\boldsymbol{\phi}_j(\lambda; \boldsymbol{\theta}_0)\}d\lambda \right. \\ & \quad \left. + \int_{-\pi}^{\pi} \text{tr}\{\mathbf{f}(\lambda)\boldsymbol{\phi}_i(\lambda; \boldsymbol{\theta}_0)\}\text{tr}\{\mathbf{f}(\lambda)\boldsymbol{\phi}_j(\lambda; \boldsymbol{\theta}_0)\}d\lambda \right], \end{aligned}$$

respectively.

The proof of Theorem 2.2 is given in Sect. 2.6. In what follows, we give several application examples of the theorem.

Example 2.1 (Autocorrelation) Denote the autocovariance and the autocorrelation of the process $\{X_i(t)\}$ (i -th component of the process $\{\mathbf{X}(t)\}$) with lag h by $\gamma_i(h)$ and $\rho_i(h)$, respectively. Suppose that we are interested in the joint estimation of $\rho_i(h)$ and $\rho_j(k)$. Take

$$\begin{aligned} \boldsymbol{\phi}_1(\lambda; \boldsymbol{\theta}) &= \begin{cases} \cos(h\lambda) - \theta_1 & (i, i)\text{-th component} \\ 0 & \text{otherwise} \end{cases}, \\ \boldsymbol{\phi}_2(\lambda; \boldsymbol{\theta}) &= \begin{cases} \cos(k\lambda) - \theta_2 & (j, j)\text{-th component} \\ 0 & \text{otherwise} \end{cases}. \end{aligned}$$

Then (2.6) leads to $\theta_{10} = \gamma_i(h)/\gamma_i(0)$ and $\theta_{20} = \gamma_j(k)/\gamma_j(0)$, hence, then $\boldsymbol{\theta}_0 = (\theta_{10}, \theta_{20})'$ corresponds to the desired autocorrelations $\boldsymbol{\rho} = (\rho_i(h), \rho_j(k))'$. ■

Example 2.2 (Portfolio selection) Let $X_i(t)$ be the log-return of i -th asset ($i = 1, \dots, m$) at time t and suppose that the process $\{\mathbf{X}(t) = (X_1(t), \dots, X_m(t))'\}$ is stationary with zero mean. Consider the portfolio $p(t) = \sum_{i=1}^m \theta_i X_i(t)$ where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)'$ is a vector of weights. The process $\{p(t)\}$ is a linear combination of the stationary process, hence $\{p(t)\}$ is also stationary and, from Herglotz's theorem, its variance is

$$\text{Var}\{p(t)\} = \boldsymbol{\theta}'\text{Var}\{\mathbf{X}(t)\}\boldsymbol{\theta} = \boldsymbol{\theta}'\left(\int_{-\pi}^{\pi} \mathbf{f}(\lambda)d\lambda\right)\boldsymbol{\theta}.$$

Our aim is to find the weights $\boldsymbol{\theta}_0 = (\theta_{10}, \dots, \theta_{m0})'$, that minimize the variance (the risk) of the portfolio $p(t)$. The first-order condition $(\partial/\partial\boldsymbol{\theta})_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\text{Var}\{p(t)\} = \mathbf{0}$ leads to

$$\left(\int_{-\pi}^{\pi} \mathbf{f}(\lambda) + \mathbf{f}(\lambda)'d\lambda\right)\boldsymbol{\theta}_0 = \mathbf{0}. \quad (2.9)$$

Now, for $j = 1, \dots, m$, consider to take

$$\phi_j(\lambda; \theta) = \begin{cases} \theta_i & (j, i)\text{-th and } (i, j)\text{-th component, } (i = 1, \dots, m \text{ and } i \neq j) \\ 2\theta_j & (j, j)\text{-th component} \\ 0 & \text{otherwise} \end{cases}.$$

Then, (2.6) coincides with the condition (2.9), which implies that the best portfolio weights can be solved with the framework of the spectral moment condition. ■

Example 2.3 (Whittle estimation) Consider fitting a parametric spectral density model $f_\theta(\lambda)$ to the true spectral density $f(\lambda)$, and seeking the quasi-true value $\underline{\theta}$ in (1.21). Assume that the spectral density model has the form described in Assumption (NGR) (iii). The Kolmogorov's formula (cf. p. 162 of Hannan (1970)) says

$$\det \mathbf{K} = \exp \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \det \{2\pi f_\theta(\lambda)\} d\lambda \right\}.$$

This implies that, if θ is innovation-free, the quantity $\int_{-\pi}^{\pi} \log \det \{f_\theta(\lambda)\} d\lambda$ is independent of θ and (1.21) leads to

$$\left. \frac{\partial}{\partial \theta} \int_{-\pi}^{\pi} \text{tr} \{f_\theta(\lambda)^{-1} f(\lambda)\} d\lambda \right|_{\theta=\underline{\theta}} = \mathbf{0}.$$

This corresponds to (2.6) when we set

$$\phi_j(\lambda; \theta) = \frac{\partial f_\theta(\lambda)^{-1}}{\partial \theta_j} \quad (j = 1, \dots, p),$$

so the quasi-true value can be expressed by the spectral moment condition. ■

2.4 Extensions of Empirical Likelihood

Since the classical empirical likelihood method was proposed, several extensions have been considered by many researchers. Among those, we introduce Cressie-Read power-divergence (CR) statistic and generalized empirical likelihood (GEL).

As an alternative of the empirical likelihood ratio, Baggerly (1998) made use of the CR statistic, which is used at χ^2 -goodness of fit test. It is defined as:

$$\frac{2}{v(v+1)} \sum_{i=1}^k N_i \left\{ \left(\frac{N_i}{np_i} \right)^v - 1 \right\}, \quad v \in \mathbb{R},$$

where k is a number of categories, N_i is a number of the observations which fell into the i -th category and p_i is the probability that the observation falls into the i -th category (see Read and Cressie (1988)). In the case of $\nu = -1, 0$, it is define by the continuous limits:

$$2 \sum_{i=1}^k (np_i) \log \left(\frac{np_i}{N_i} \right) \quad \text{and} \quad 2 \sum_{i=1}^k N_i \log \left(\frac{N_i}{np_i} \right).$$

The CR statistic contains the user-specified parameter $\nu \in \mathbb{R}$ and encompasses several commonly used test statistics, i.e., the Neyman-modified χ^2 -statistic ($\nu = -2$), the Kullback-Leibler divergence ($\nu = -1$), the Freeman-Tukey statistic ($\nu = -1/2$) and Pearson's χ^2 -statistic ($\nu = 1$).

In the empirical likelihood framework, we consider the distribution whose support is a set of observations. This means that $k = n$ and $N_i = 1$ for all i in the above setting. Therefore, CR statistic is reduced to

$$\text{CR}_\nu(\mathbf{p}) = \frac{2}{\nu(\nu+1)} \sum_{i=1}^n \{(np_i)^{-\nu} - 1\}, \quad \nu \in \mathbb{R}. \quad (2.10)$$

The empirical likelihood statistic corresponds to the specific case of $\nu = 0$, therefore, CR statistic can be considered as an extension of the empirical likelihood. Baggerly (1998) showed the CR statistic is asymptotically chi-square distributed for i.i.d. random vectors.

For the asymptotic distribution of CR statistic under the dependent process, we first consider the same estimating function as in (2.7). Following (2.8), we define the frequency domain CR statistic:

$$\widetilde{\text{CR}}_\nu(\boldsymbol{\theta}) = \min_{\mathbf{p}} \left\{ \text{CR}_\nu(\mathbf{p}) \left| \sum_{t=1}^n p_t \mathbf{m}(\lambda_t; \boldsymbol{\theta}) = \mathbf{0}, p_t \geq 0, \sum_{t=1}^n p_t = 1 \right. \right\}. \quad (2.11)$$

Then we obtain the following theorem.

Theorem 2.3 (Ogata and Taniguchi (2009)) *Let Assumption 2.1 hold. Then $\widetilde{\text{CR}}_\nu(\boldsymbol{\theta}_0)$ has the same asymptotic distribution as in Theorem 2.2 for any $\nu \in \mathbb{R}$.*

The proof of Theorem 2.3 is given in Sect. 2.6. In what follows, we give several application examples, in which we assume that $X(t)$ is scalar, i.e., $\mathbf{X}(t) = X(t)$, for simplicity.

Example 2.4 (Prediction) The h -step prediction problem as in Hannan (1970, Chapter III Section 2) is considered. We predict $X(t)$ using a linear combination of $X(t-j)$, $j \geq h$, that is, we use $\tilde{X}(t) = \sum_{j \geq h} a(j; \boldsymbol{\theta}) X(t-j)$ as a predictor where $a(j; \boldsymbol{\theta})$ s are constants. We measure the error of the predictor by $E[|X(t) - \tilde{X}(t)|^2]$ and find the best linear predictor which minimizes this error. The spectral representations of $X(t)$ and $\tilde{X}(t)$ are

$$X(t) = \int_{-\pi}^{\pi} \exp(-it\lambda) z(d\lambda), \quad \tilde{X}(t) = \int_{-\pi}^{\pi} \exp(-it\lambda) \left\{ \sum_{j \geq h} a(j; \theta) \exp(ij\lambda) \right\} z(d\lambda)$$

where $E[|z(d\lambda)|^2] = f(\lambda) d\lambda$, $E[z(d\lambda_1) \overline{z(d\lambda_2)}] = 0$, $\lambda_1 \neq \lambda_2$. Therefore, the prediction error is written by

$$\int_{-\pi}^{\pi} \left| 1 - \sum_{j \geq h} a(j; \theta) \exp(ij\lambda) \right|^2 f(\lambda) d\lambda.$$

We find the minimizer of this error, θ_0 , by the equation

$$\frac{\partial}{\partial \theta} \int_{-\pi}^{\pi} \left| 1 - \sum_{j \geq h} a(j; \theta) \exp(ij\lambda) \right|^2 f(\lambda) d\lambda \Big|_{\theta=\theta_0} = \mathbf{0}.$$

This is exactly our problem when we take

$$\phi_j(\lambda; \theta) = \frac{\partial}{\partial \theta_j} \left| 1 - \sum_{j \geq h} a(j; \theta) \exp(ij\lambda) \right|^2$$

in (2.6).

Example 2.5 (Interpolation) Assume that the entire time series has been observed except for the time point $t = 0$. We would like to estimate $X(0)$ by a linear combination of the observed stochastic variables, that is, $\tilde{X}(0) = \sum_{j \neq 0} a(j; \theta) X(j)$. Similar to the prediction problem, the error of interpolation becomes

$$\int_{-\pi}^{\pi} \left| 1 - \sum_{j \neq 0} a(j; \theta) \exp(ij\lambda) \right|^2 f(\lambda) d\lambda.$$

We find the minimizer of this error, θ_0 , by the equation

$$\frac{\partial}{\partial \theta} \int_{-\pi}^{\pi} \left| 1 - \sum_{j \neq 0} a(j; \theta) \exp(ij\lambda) \right|^2 f(\lambda) d\lambda \Big|_{\theta=\theta_0} = \mathbf{0}.$$

This is exactly our problem when we take

$$\phi_j(\lambda; \theta) = \frac{\partial}{\partial \theta_j} \left| 1 - \sum_{j \neq 0} a(j; \theta) \exp(ij\lambda) \right|^2$$

in (2.6).

Example 2.6 (Smoothing) Consider smoothing the trajectory of $X(t)$ by a linear combination of adjacent observations, $\sum_{j=-N}^N \theta_j X(t+j)$. Then, similar to the

previous problem, the error of this smoothing is expressed as:

$$\int_{-\pi}^{\pi} \left| 1 - \sum_{j=-N}^N \theta_j \exp(ij\lambda) \right|^2 f(\lambda) d\lambda.$$

Set $\boldsymbol{\theta} = (\theta_{-N}, \dots, \theta_N)'$ and we find the minimizer of this error, $\boldsymbol{\theta}_0$, by the equation

$$\left. \frac{\partial}{\partial \boldsymbol{\theta}} \int_{-\pi}^{\pi} \left| 1 - \sum_{j=-N}^N \theta_j \exp(ij\lambda) \right|^2 f(\lambda) d\lambda \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \mathbf{0}.$$

This is exactly our problem when we take

$$\phi_j(\lambda; \boldsymbol{\theta}) = \frac{\partial}{\partial \theta_j} \left| 1 - \sum_{j=-N}^N \theta_j \exp(ij\lambda) \right|^2$$

in (2.6).

Another extension of empirical likelihood is GEL in Newey and Smith (2004). GEL can be also considered as an alternative of generalized methods of moments (GMM) and it is known that its asymptotic bias does not grow with the number of moment restrictions, while the bias of GMM often does. Following the i.i.d. GEL described in Newey and Smith (2004), Ogata (2012) investigated the asymptotic properties of frequency domain GEL.

To describe GEL, let $\rho(y)$ be a function of a scalar y that is concave on its domain, an open interval $\mathcal{Y}$ containing zero. Let $\hat{\mathcal{E}}_n(\boldsymbol{\theta}) = \{\boldsymbol{\xi} : \boldsymbol{\xi}' \mathbf{m}(\lambda_t; \boldsymbol{\theta}) \in \mathcal{Y}, t = 1, \dots, n\}$. The estimator is the solution to a saddle point problem

$$\hat{\boldsymbol{\theta}}_{\text{GEL}} = \arg \min_{\boldsymbol{\theta} \in \Theta} \sup_{\boldsymbol{\xi} \in \hat{\mathcal{E}}_n(\boldsymbol{\theta})} \sum_{t=1}^n \rho(\boldsymbol{\xi}' \mathbf{m}(\lambda_t; \boldsymbol{\theta})).$$

The empirical likelihood (EL) estimator of Qin and Lawless (1994), the exponential tilting (ET) estimator of Kitamura and Stutzer (1997) and the continuous updating estimator (CUE) of Hansen et al. (1996) are special cases with $\rho(y) = \log(1 - y)$, $\rho(y) = -e^y$ and $\rho(y) = -(1 + y)^2/2$, respectively. Let $\boldsymbol{\Omega} = E[\mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0)']$. The following assumptions and theorems are frequency domain version of those in Newey and Smith (2004), and are found in Ogata (2012).

Assumption 2.2

- (i) $\boldsymbol{\theta}_0 \in \Theta$ is the unique solution to (2.6).
- (ii) Θ is compact.
- (iii) $\mathbf{m}(\lambda_t; \boldsymbol{\theta})$ is continuous at each $\boldsymbol{\theta} \in \Theta$ with probability one.

Table 2.1 Market

Index	Market
S&P 500	NYSE
Bovespa	São Paulo stock exchange
CAC40	Bourse de Paris
AEX	Amsterdam stock exchange
ATX	Wiener Börse
HKHSI	Hong Kong exchanges and clearing
Nikkei	Tokyo stock exchange

Taken from Ogata (2012). Published with the kind permission of © Hiroaki Ogata 2012. Published under the Creative Commons Attribution License

- (iv) $E[\sup_{\theta \in \Theta} \|\mathbf{m}(\lambda_t; \theta)\|^\alpha] < \infty$ for some $\alpha > 2$.
- (v) $\boldsymbol{\Omega}$ is nonsingular.
- (vi) $\rho(y)$ is twice continuously differentiable in a neighborhood of zero.

Theorem 2.4 *Let Assumption 2.2 hold. Then $\hat{\boldsymbol{\theta}}_{\text{GEL}} \xrightarrow{P} \boldsymbol{\theta}_0$.*

Furthermore, let $\mathbf{G} = E[\partial \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0)/\partial \boldsymbol{\theta}]$.

Assumption 2.3

- (i) $\boldsymbol{\theta}_0 \in \text{int}(\Theta)$.
- (ii) $\mathbf{m}(\lambda_t; \boldsymbol{\theta})$ is continuously differentiable in a neighborhood $\mathcal{N}$ of $\boldsymbol{\theta}_0$ and

$$E \left[\sup_{\boldsymbol{\theta} \in \mathcal{N}} \left\| \frac{\partial \mathbf{m}(\lambda_t; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \right\| \right] < \infty.$$

- (iii) $\text{rank}(\mathbf{G}) = p$.

Theorem 2.5 *Let Assumptions 2.2 and 2.3 hold. Then*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{\text{GEL}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, (\mathbf{G}'\boldsymbol{\Omega}^{-1}\mathbf{G})^{-1}).$$

Now we are applying the frequency domain GEL estimation method to the portfolio selection problem introduced in Example 2.2. The sample consists of 7 weekly market indices (S&P 500, Bovespa, CAC 40, AEX, ATX, HKHSI, and Nikkei) having 800 observations each: the initial date is April 30, 1993, and the ending date is August 22, 2008. The data are from Ogata (2012). Refer to Table 2.1 for the market of each index.

The log-return of the i -th ($i = 1, \dots, 7$) index at time t is denoted by $x_i(t)$ and each process is assumed to be stationary. We estimate the best portfolio weights $\boldsymbol{\theta}_0 = (\theta_1, \dots, \theta_7)'$ described in Example 2.2 by using three types of frequency domain GEL estimators (EL, ET, and CUE). The results are shown in Table 2.2, and explain that Bovespa and ATX contribute large part in the optimal portfolio.

Table 2.2 Estimated portfolio weights

	EL	ET	CUE
S&P 500	0.0759	0.0617	0.0001
Bovespa	0.6648	0.6487	0.6827
CAC40	0.0000	0.0000	0.0000
AEX	0.0000	0.0000	0.0000
ATX	0.2593	0.2558	0.3168
HKHSI	0.0000	0.0000	0.0000
Nikkei	0.0000	0.0338	0.0004

Taken from Ogata (2012). Published with the kind permission of © Hiroaki Ogata 2012. Published under the creative commons attribution license

2.5 GEL for Multivariate Stable Distributions

It is well known that financial data often show asymmetry and heavy-tails properties which cannot be captured by simple distributions such as normal distribution. One of the families of distributions which can be accommodated to such asymmetry and heavy-tails data is the stable family. Since Mandelbrot (1963) and Fama (1965) proposed the use of stable distributions to analyze financial data, these distributions have received considerable attention; see, for example, Embrechts et al. (1997), Belkacem et al. (2000), and Rachev and Han (2000). Moreover, the stable distributions arise from a generalization of the central limit theorem in which the assumption of finite variance is relaxed, and consequently, the stable distributions are closed under summation. This property is also a strong motivation to fit stable distributions to financial data, since low frequency financial returns can be regarded as the sum of high frequency data.

In order to deal with several financial assets simultaneously, we treat multivariate stable distributions. In fact, we have multivariate stable distributions that are included in an elliptical family. Elliptical stable distributions are easier to handle than nonelliptical stable distributions but we do not consider them in this section. The reason is in Fig. 2.1, that displays density plots of bivariate normal, elliptical stable, and nonelliptical stable distributions along with contour plots of the densities. The graphs for the nonelliptical stable distribution show skewness that cannot be handled by either normal or elliptical stable distributions. Therefore, we especially consider multivariate nonelliptical stable distributions and apply the i.i.d. GEL method to estimate their parameters, following Ogata (2013).

We briefly review the multivariate stable distributions. A random vector $\mathbf{X} = (X_1, \dots, X_d)' \in \mathbb{R}^d$ is said to be stable if its characteristic function is

$$\phi(\mathbf{v}) = E[\exp\{i\langle \mathbf{X}, \mathbf{v} \rangle\}] = \exp\{-I(\mathbf{v})\},$$

where the exponent function is

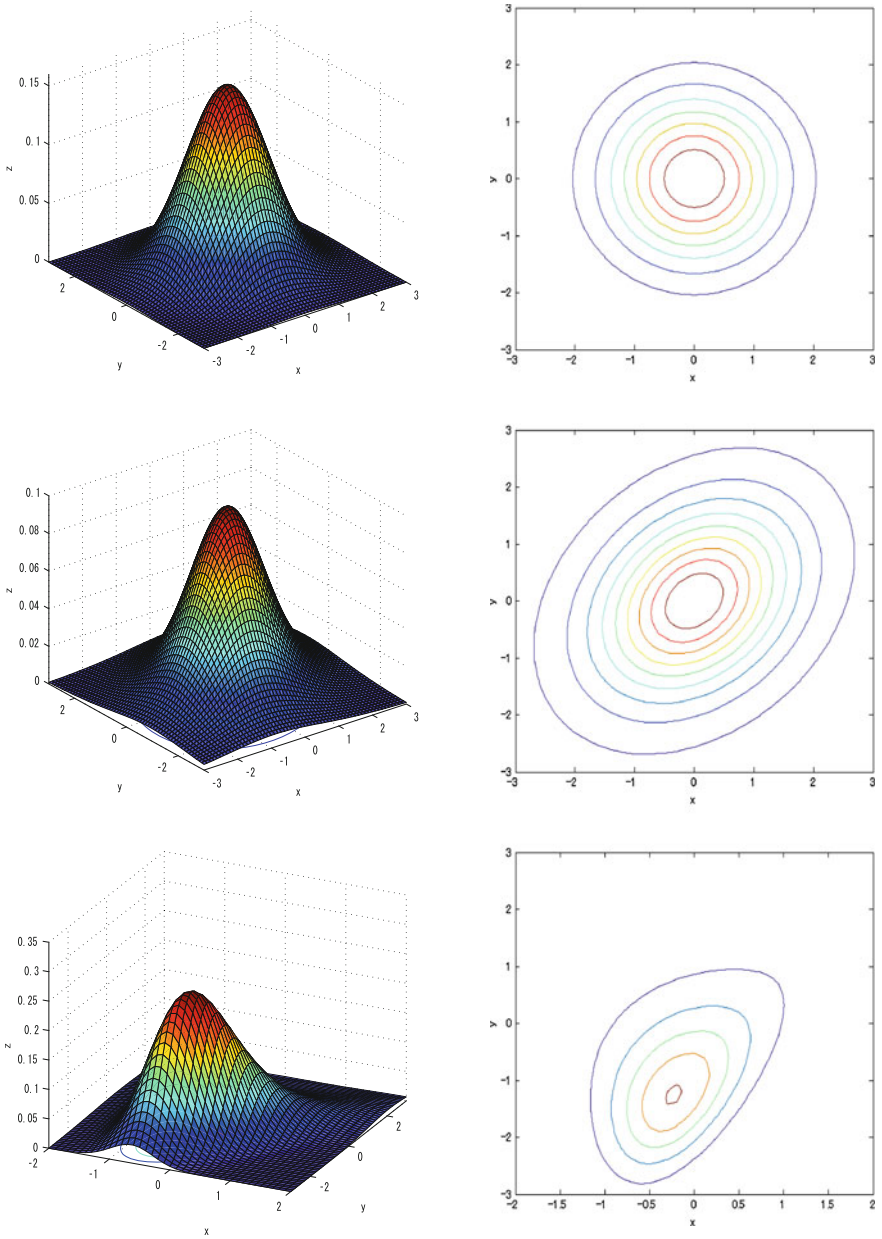


Fig. 2.1 The *left three panels* show bivariate density plots and *right three panels* show their contours. From *top to bottom*, the figures are for normal, elliptical stable, and nonelliptical stable distributions, respectively. Taken from Ogata (2013). Published with the kind permission of © Elsevier 2013. All Rights Reserved

$$I(\mathbf{v}) = \int_{S_d} \psi(\langle \mathbf{s}, \mathbf{v} \rangle) \Gamma(d\mathbf{s}) + i \langle \mathbf{v}, \boldsymbol{\mu} \rangle.$$

Here $S_d = \{\mathbf{s} : \|\mathbf{s}\| = 1\}$ is the unit sphere in $\mathbb{R}^d$, the symbol $\langle \cdot, \cdot \rangle$ denotes inner product, Γ is a finite spectral measure on S_d , $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d) \in \mathbb{R}^d$ is the location vector, and

$$\psi(u) = \begin{cases} |u|^\alpha \left(1 - i \operatorname{sign}(u) \tan \frac{\pi\alpha}{2}\right) & (\alpha \neq 1) \\ |u| \left(1 + i \frac{2}{\pi} \operatorname{sign}(u) \ln |u|\right) & (\alpha = 1), \end{cases}$$

where $\alpha \in (0, 2]$ is the characteristic exponent, which controls tail thickness. The d -dimensional stable distribution is denoted by $S^d(\alpha, \Gamma, \boldsymbol{\mu})$.

Our purpose is to estimate the spectral measure Γ , the location vector $\boldsymbol{\mu}$, and the characteristic exponent α , based on an i.i.d. sample $\mathbf{X}_1, \dots, \mathbf{X}_n$ of d -dimensional random vectors drawn from this distribution. However, difficulties arise even in the univariate case, $d = 1$. Foremost is the complexity of the density function. Except for a few cases, no simple explicit form of the density exists, which is an obstacle to implementing the usual MLE method. Moreover, the stable distributions do not necessarily have second or even first moments, which impede using the ordinary method of moments.

One remedy for this difficulty is to use the empirical characteristic function

$$\hat{\phi}_n(\mathbf{v}) = \frac{1}{n} \sum_{j=1}^n \exp\{i \langle \mathbf{X}_j, \mathbf{v} \rangle\}.$$

Following Nolan et al. (2001), we first consider a discrete approximation of the spectral measure:

$$\Gamma^* = \sum_{\ell=1}^L \gamma_\ell \delta_{s_\ell}, \quad (2.12)$$

where $\gamma_\ell = \Gamma(A_\ell)$, $\ell = 1, \dots, L$, are weights and δ_{s_ℓ} is a point mass at $s_\ell \in S_d$. Here, A_ℓ , $\ell = 1, \dots, L$, are patches that partition the sphere S_d , where A_ℓ has some “center” s_ℓ . The estimation of the parameters is done via an estimating function based on theoretical and empirical characteristic functions. Many authors have investigated the use of empirical characteristic functions for estimation; for a summary and extensive references, see Yu (2004). In the context of multivariate stable distributions, suppose that $\{\mathbf{X}_j\}_{j=1}^n$ is an i.i.d. sequence from $S^d(\alpha, \Gamma^*, \boldsymbol{\mu})$, where Γ^* is a discrete spectral measure defined in (2.12). Denote the parameters by $\boldsymbol{\theta} = (\alpha, \boldsymbol{\gamma}', \boldsymbol{\mu}')' \in (0, 2] \times \mathbb{R}_+^L \times \mathbb{R}^d$, and define the estimating function as

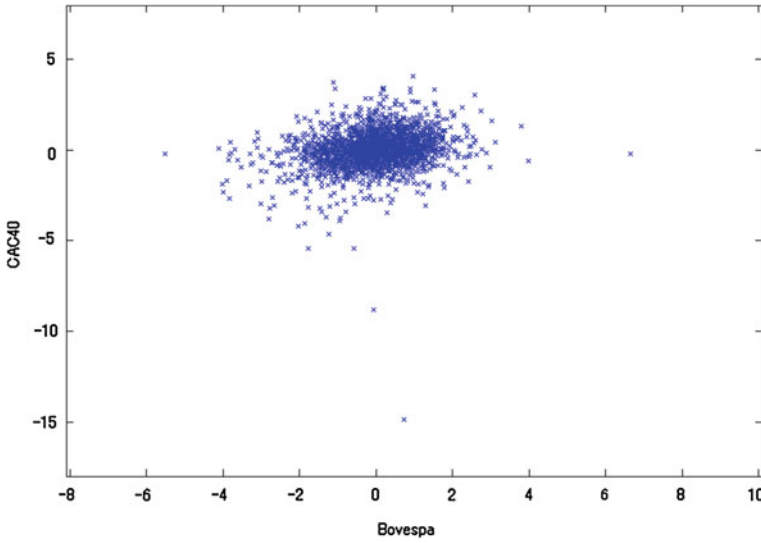


Fig. 2.2 Scatter plot of the pair of Bovespa and CAC40. Taken from Ogata (2013). Published with the kind permission of © Elsevier 2013. All Rights Reserved

$$h(\mathbf{v}, X_j, \boldsymbol{\theta}) = \exp\{i\langle \mathbf{v}, X_j \rangle\} - \phi_{\boldsymbol{\theta}}(\mathbf{v}),$$

where $\phi_{\boldsymbol{\theta}}$ is the theoretical characteristic function for model parameter $\boldsymbol{\theta}$. Denoting true parameter value by $\boldsymbol{\theta}_0$, we have $E[h(\mathbf{v}, X_j, \boldsymbol{\theta}_0)] = 0$ for any frequency $\mathbf{v} \in \mathbb{R}^d$. After choosing some frequencies $\mathbf{v}_1, \dots, \mathbf{v}_K \in \mathbb{R}^d$, we redefine the estimating function as

$$\mathbf{g}(X_j, \boldsymbol{\theta}) = (\Re[h(\mathbf{v}_1, X_j, \boldsymbol{\theta})], \dots, \Re[h(\mathbf{v}_K, X_j, \boldsymbol{\theta})], \Im[h(\mathbf{v}_1, X_j, \boldsymbol{\theta})], \dots, \Im[h(\mathbf{v}_K, X_j, \boldsymbol{\theta})])', \quad (2.13)$$

where $\Re[\cdot]$ and $\Im[\cdot]$ are the real and imaginary parts of a complex number. Obviously, we still have $E[\mathbf{g}(X_j, \boldsymbol{\theta}_0)] = \mathbf{0}$. Using this estimating function, we estimate the parameters by the GEL method described in the previous section.

Now we are applying the GEL estimation method for multivariate stable distributions to real data. The sample consists of two log returns of daily market indexes, Bovespa (São Paulo Stock Exchange) and CAC40 (Bourse de Paris), having 2536 observations in each.¹ The initial date is January 4, 2000, and the ending date is September 22, 2009. Figure 2.2 shows a scatter plot of their returns. In this plot, several points are located far from the origin, and the data are skewed heavily downward. Thus, multivariate stable distributions are more appropriate than multivariate normal or elliptical distributions.

¹ This data is a part of the data set which was used in Dominicy and Veredas (2013).

Table 2.3 Estimated parameters and VaRs : σ_{BB} and σ_{CC} are the variances for Bovespa and CAC40, and σ_{BC} is the covariance

Model	Stable				Normal	
	α	1.757				
Estimated parameters	γ_1	0.409	γ_5	0.598		
	γ_2	0.000	γ_6	0.000	σ_{BB}	0.999
	γ_3	0.000	γ_7	0.000	σ_{BC}	0.219
	γ_4	0.226	γ_8	0.729	σ_{CC}	0.988
VaR _B (0.01)						-2.960
VaR _C (0.01)						-1.603

VaR_B(0.01) and VaR_C(0.01) stand for the 1 % VaRs for Bovespa and CAC40, respectively. Taken from Ogata (2013). Published with the kind permission of © Elsevier 2013. All Rights Reserved

We model the pair (Bovespa, CAC40) using the bivariate stable distribution $S^2(\alpha, \Gamma^*, \mathbf{0})$. Here, Γ^* is a discrete spectral measure defined in (2.12), and we take $L = 8, s_\ell = (\cos \theta_\ell, \sin \theta_\ell)'$ for $\theta_\ell = (\ell - 1)\pi/4, \ell = 1, \dots, 8$. For comparison, the bivariate normal distribution $N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{BB} & \sigma_{BC} \\ \sigma_{BC} & \sigma_{CC} \end{pmatrix}\right)$ is also fitted to the data. We use ET estimators for the parameters of the stable distribution, $(\alpha, \gamma_1, \dots, \gamma_8)'$, and MLEs for the parameters of the normal distribution, $(\sigma_{BB}, \sigma_{BC}, \sigma_{CC})$. Based on the estimated parameters for both models, we calculate the 1 % VaRs for the Bovespa and the CAC40. Table 2.3 shows the estimation results and the VaRs.

Both VaRs from the fitted stable model are smaller than those from the fitted normal model, which means fitting the stable distribution evaluates the risk to be higher than that from fitting the normal distribution. The estimated marginal density functions for Bovespa and CAC40 are displayed in Fig. 2.3.

Last, we make a comment about the data. Perhaps, heavy tails may be caused by GARCH effects rather than by unconditional stable distributions. To remove the effect of the dynamic conditional volatility, Dominicy and Veredas (2013) adjusted the original data by using the AR(2)-GARCH(1,1) model. The analysis here is also based on this adjusted data.

2.6 Appendix

In this section, we provide the proofs of Theorems 2.2 and 2.3. For the simplicity of expression, we introduce

$$P_n = \frac{1}{\sqrt{n}} \sum_{t=1}^n m(\lambda_t; \theta_0),$$

$$S_n = \frac{1}{n} \sum_{t=1}^n m(\lambda_t; \theta_0) m(\lambda_t; \theta_0)',$$

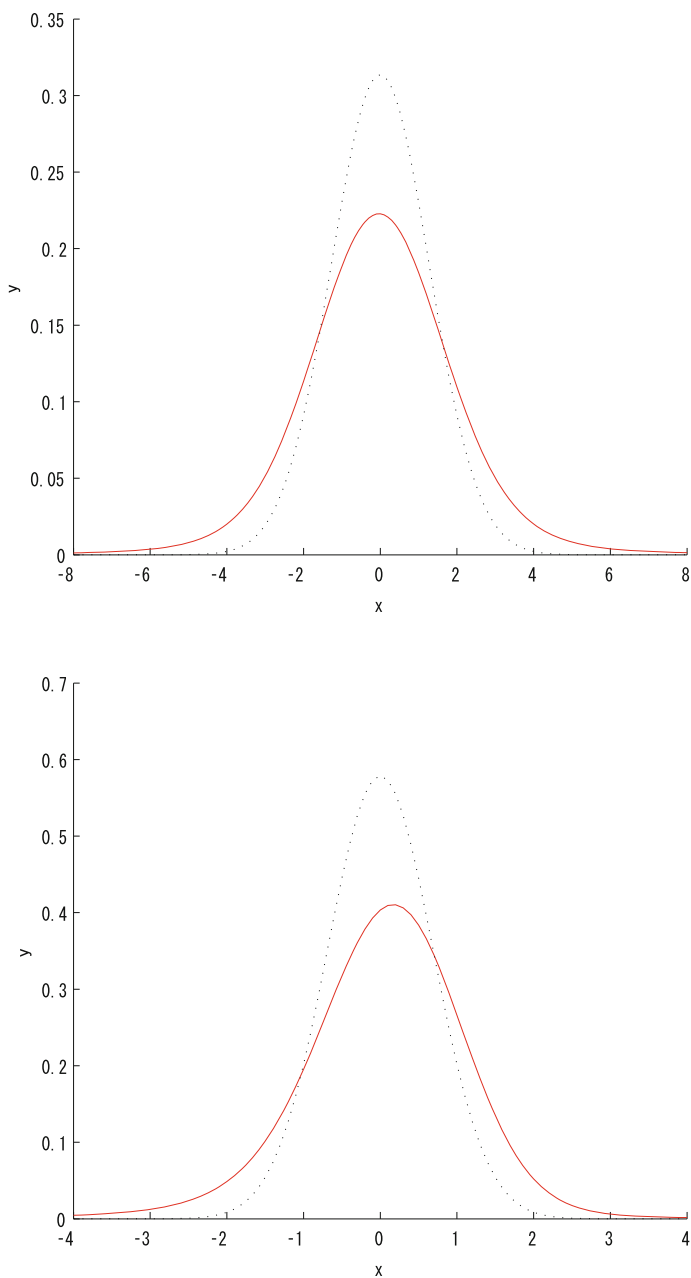


Fig. 2.3 The estimated probability density functions are displayed. The *upper figure* is for Bovespa while *lower* is for CAC40. The *real line* is for stable fitting and *dotted line* is for normal fitting. Taken from Ogata (2013). Published with the kind permission of © Elsevier 2013. All Rights Reserved

$$Z_n = \max_{1 \leq t \leq n} \|\mathbf{m}(\lambda_t; \boldsymbol{\theta}_0)\|.$$

Before giving the proofs of the theorems, we prepare Lemmas 2.1, 2.2 and 2.3 whose proofs are omitted. Lemma 2.1 is proved by use of Theorem 5.10.2 and Lemma P5.1 of Brillinger (2001), and Proposition 1.2 in Chap. 1. Lemmas 2.2 and 2.3 are proved by Theorems 7.2.2 and 4.5.1 of Brillinger (2001), respectively.

Lemma 2.1 *Let Assumption 2.1 hold. Then*

$$\mathbf{P}_n \xrightarrow{d} N_q(\mathbf{0}, \boldsymbol{\Sigma}_1)$$

where $\boldsymbol{\Sigma}_1$ is the same matrix as in Theorem 2.2.

Lemma 2.2 *Let Assumption 2.1 hold. Then*

$$\mathbf{S}_n \xrightarrow{p} \boldsymbol{\Sigma}_2$$

where $\boldsymbol{\Sigma}_2$ is the same matrix as in Theorem 2.2.

Lemma 2.3 *Let Assumption 2.1 hold. Then*

$$Z_n = O(\log n).$$

2.6.1 Proof of Theorem 2.2

Following the way of leading to (2.5), we may write

$$\tilde{\mathcal{R}}(\boldsymbol{\theta}_0) = \prod_{t=1}^n \frac{1}{1 + \boldsymbol{\xi}' \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0)}$$

where the vector $\boldsymbol{\xi} = \boldsymbol{\xi}(\boldsymbol{\theta}_0)$ satisfies q equations given by

$$\mathbf{J}(\boldsymbol{\xi}) \equiv \frac{1}{n} \sum_{t=1}^n \frac{\mathbf{m}(\lambda_t; \boldsymbol{\theta}_0)}{1 + \boldsymbol{\xi}' \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0)} = \mathbf{0}. \quad (2.14)$$

Put $Y_t = \boldsymbol{\xi}' \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0)$ and write $\boldsymbol{\xi} = \|\boldsymbol{\xi}\| \mathbf{u}$ where $\mathbf{u} \in U$ is a unit vector. By substituting $1/(1 + Y_t) = 1 - Y_t/(1 + Y_t)$ into $\mathbf{u}' \mathbf{J}(\boldsymbol{\xi}) = 0$, we have

$$\begin{aligned}
\mathbf{u}' \left\{ \frac{1}{n} \sum_{t=1}^n \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) \left(1 - \frac{Y_t}{1 + Y_t} \right) \right\} &= 0 \\
\mathbf{u}' \left(\frac{1}{n} \sum_{t=1}^n \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) \right) &= \mathbf{u}' \left(\frac{1}{n} \sum_{t=1}^n \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) \frac{\boldsymbol{\xi}' \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0)}{1 + Y_t} \right) \\
\mathbf{u}' \left(\frac{1}{n} \sum_{t=1}^n \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) \right) &= \|\boldsymbol{\xi}\| \mathbf{u}' \left(\frac{1}{n} \sum_{t=1}^n \frac{\mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0)'}{1 + Y_t} \right) \mathbf{u}.
\end{aligned}$$

Then, we obtain

$$\begin{aligned}
\|\boldsymbol{\xi}\| \mathbf{u}' S_n \mathbf{u} &\leq \|\boldsymbol{\xi}\| \mathbf{u}' \left(\frac{1}{n} \sum_{t=1}^n \frac{\mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0)'}{1 + Y_t} \right) \mathbf{u} \cdot (1 + \max_t Y_t) \\
&\leq \|\boldsymbol{\xi}\| \mathbf{u}' \left(\frac{1}{n} \sum_{t=1}^n \frac{\mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0)'}{1 + Y_t} \right) \mathbf{u} \cdot (1 + \|\boldsymbol{\xi}\| Z_n) \\
&= n^{-1/2} \mathbf{u}' \mathbf{P}_n (1 + \|\boldsymbol{\xi}\| Z_n)
\end{aligned}$$

and so

$$\|\boldsymbol{\xi}\| \left(\mathbf{u}' S_n \mathbf{u} - n^{-1/2} Z_n \mathbf{u}' \mathbf{P}_n \right) \leq n^{-1/2} \mathbf{u}' \mathbf{P}_n.$$

From Lemmas 2.1–2.3, we have

$$\|\boldsymbol{\xi}\| \{O_p(1) - O(\log n) O_p(n^{-1/2})\} \leq O_p(n^{-1/2}),$$

and hence,

$$\|\boldsymbol{\xi}\| = O_p(n^{-1/2}). \quad (2.15)$$

After we have established an order of $\boldsymbol{\xi}$, we have from Lemma 2.3 and (2.15) that

$$\max_{1 \leq t \leq n} |Y_t| = O_p(n^{-1/2}) O(\log n). \quad (2.16)$$

Now, by (2.14),

$$\begin{aligned}
\mathbf{0} &= \frac{1}{n} \sum_{t=1}^n \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) \left(1 - Y_t + \frac{Y_t^2}{1 + Y_t} \right) \\
&= n^{-1/2} \mathbf{P}_n - S_n \boldsymbol{\xi} + \frac{1}{n} \sum_{t=1}^n \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) \frac{Y_t^2}{1 + Y_t}
\end{aligned} \quad (2.17)$$

Noting that

$$\frac{1}{n} \sum_{t=1}^n \|\mathbf{m}(\lambda_t; \boldsymbol{\theta}_0)\|^3 \leq Z_n \|\mathbf{S}_n\| = O(\log n),$$

the final term of (2.17) has a norm bounded by

$$\frac{1}{n} \sum_{t=1}^n \|\mathbf{m}(\lambda_t, \boldsymbol{\theta}_0)\|^3 \|\boldsymbol{\xi}\|^2 |1 + Y_t|^{-1} = O(\log n) O_p(n^{-1}) O_p(1) = O_p(n^{-1} \log n).$$

Hence we can write

$$\boldsymbol{\xi} = \frac{1}{\sqrt{n}} \mathbf{S}_n^{-1} \mathbf{P}_n + \boldsymbol{\epsilon}$$

where $\boldsymbol{\epsilon} = O_p(n^{-1} \log n)$. By (2.16), we can write

$$\log(1 + Y_t) = Y_t - \frac{1}{2} Y_t^2 + \eta_t$$

where, for some finite constant $C > 0$,

$$\Pr(|\eta_t| \leq C|Y_t|^3, 1 \leq t \leq n) \rightarrow 1 \quad \text{as } n \rightarrow \infty.$$

Now we may write

$$\begin{aligned} -2 \log \tilde{\mathcal{R}}(\boldsymbol{\theta}_0) &= -2 \sum_{t=1}^n \log(n w_t) = -2 \sum_{t=1}^n \log(1 + Y_t) = 2 \sum_{t=1}^n Y_t - \sum_{t=1}^n Y_t^2 + 2 \sum_{t=1}^n \eta_t \\ &= \mathbf{P}_n' \mathbf{S}_n^{-1} \mathbf{P}_n - n \boldsymbol{\epsilon}' \mathbf{S}_n \boldsymbol{\epsilon} + 2 \sum_{t=1}^n \eta_t. \end{aligned}$$

Here it is seen that

$$\begin{aligned} n \boldsymbol{\epsilon}' \mathbf{S}_n \boldsymbol{\epsilon} &= n O_p(n^{-1} \log n) O_p(1) O_p(n^{-1} \log n) = O_p(n^{-1} (\log n)^2), \\ 2 \sum_{t=1}^n \eta_t &\leq C \|\boldsymbol{\xi}\|^3 \sum_{t=1}^n \|\mathbf{m}(\lambda_t, \boldsymbol{\theta}_0)\|^3 = O_p(n^{-3/2}) O(n \log n) = O_p(n^{-1/2} \log n), \end{aligned}$$

and finally, from Lemmas 2.1 and 2.2, we can show that

$$\mathbf{P}_n' \mathbf{S}_n^{-1} \mathbf{P}_n \xrightarrow{d} \left(\boldsymbol{\Sigma}_2^{-1/2} \boldsymbol{\Sigma}_1^{1/2} \boldsymbol{\Sigma}_1^{-1/2} \mathbf{P}_n \right)' \left(\boldsymbol{\Sigma}_2^{-1/2} \boldsymbol{\Sigma}_1^{1/2} \boldsymbol{\Sigma}_1^{-1/2} \mathbf{P}_n \right) \stackrel{d}{=} (\boldsymbol{\Sigma} N)' (\boldsymbol{\Sigma} N). \quad \square$$

2.6.2 Proof of Theorem 2.3

Let $\nu \neq 0, -1$. To find the minimizing weights $\mathbf{p} = (p_1, \dots, p_n)'$ in (2.11), we proceed by the Lagrange multiplier method. Write

$$G = \text{CR}_\nu(\mathbf{p}) + \boldsymbol{\xi}' \sum_{t=1}^n p_t \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) + \gamma \left(\sum_{t=1}^n p_t - 1 \right),$$

where $\boldsymbol{\xi} \in \mathbb{R}^q$ and $\gamma \in \mathbb{R}$ are Lagrange multipliers. Since the objective function $\text{CR}_\nu(\mathbf{p})$ is convex with respect to $\mathbf{p}$, the solution of the first-order condition

$$(i) \quad \frac{\partial G}{\partial \mathbf{p}} = \mathbf{0} \quad (ii) \quad \frac{\partial G}{\partial \boldsymbol{\xi}} = \mathbf{0} \quad (iii) \quad \frac{\partial G}{\partial \gamma} = 0 \quad (2.18)$$

gives a global minimum point. The first condition of (2.18) gives

$$p_t = \frac{1}{n} \left\{ \frac{(\nu + 1)\gamma}{2n} \right\}^{-1/(1+\nu)} \left\{ 1 + \frac{\boldsymbol{\xi}'}{\gamma} \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) \right\}^{-1/(1+\nu)} \quad (t = 1, \dots, n).$$

With a simpler notation, we rewrite p_t as

$$p_t = C^* \left\{ 1 + \boldsymbol{\xi}' \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) \right\}^{-1/(1+\nu)} \quad (t = 1, \dots, n) \quad (2.19)$$

where C^* and $\boldsymbol{\xi}$ are certain constant and vector, respectively. From the third condition of (2.18), the constant should be

$$C^* = \left[\sum_{t=1}^n \left\{ 1 + \boldsymbol{\xi}' \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) \right\}^{-1/(1+\nu)} \right]^{-1}. \quad (2.20)$$

Put $Y_t = \boldsymbol{\xi}' \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0)$ and write $\boldsymbol{\xi} = \|\boldsymbol{\xi}\| \mathbf{u}$ where $\mathbf{u} \in U$ is a unit vector. From the second condition of (2.18), we have

$$\frac{1}{n} \mathbf{u}' \sum_{t=1}^n p_t \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) = 0,$$

and by substituting (2.19) and (2.20) into this equation, we have

$$\begin{aligned} 0 &= \frac{1}{n} \mathbf{u}' \sum_{t=1}^n \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) (1 + Y_t)^{-1/(1+\nu)} \\ &= \frac{1}{n} \mathbf{u}' \sum_{t=1}^n \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) \left\{ 1 - \frac{(1 + \eta_t Y_t)^{-1/(1+\nu)-1}}{1 + \nu} Y_t \right\} \quad (0 < \eta_t < 1). \end{aligned}$$

We rewrite this as

$$\begin{aligned} \|\xi\| &= \left[\mathbf{u}' \left\{ \frac{1}{n} \sum_{t=1}^n \{1 + \eta_t \|\xi\| \mathbf{u}' \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0)\}^{-1/(1+\nu)-1} \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0)'\right\} \mathbf{u} \right]^{-1} \\ &\quad \times (1 + \nu) \mathbf{u}' \left\{ \frac{1}{n} \sum_{t=1}^n \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) \right\}. \end{aligned} \quad (2.21)$$

From Lemmas 2.1–2.3, there exists a sufficiently large C and an integer n_0 such that if $n > n_0$ and $\|\xi\| < Cn^{-1/2}$, the right hand side of (2.21) is less than $Cn^{-1/2}$. Applying the Brouwer fixed point theorem to the right hand side of (2.21), we can see that a solution of Eq. (2.21) exists in the region $\|\xi\| < Cn^{-1/2}$. Hence it is shown that

$$\|\xi\| = O_p(n^{-1/2}). \quad (2.22)$$

Now, from the second condition of (2.18), we obtain

$$\begin{aligned} \mathbf{0} &= \frac{1}{n} \sum_{t=1}^n \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) (1 + Y_t)^{-1/(1+\nu)} \\ &= \frac{1}{n} \sum_{t=1}^n \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) \left\{ 1 - (1 + \nu)^{-1} Y_t + O_p(Y_t^2) \right\} \\ &= n^{-1/2} \mathbf{P}_n - (1 + \nu)^{-1} \mathbf{S}_n \xi + \boldsymbol{\beta}. \end{aligned}$$

With some constant C , the norm of the last term $\boldsymbol{\beta}$ is evaluated as

$$\|\boldsymbol{\beta}\| = \frac{C}{n} \sum_{t=1}^n \|\mathbf{m}(\lambda_t; \boldsymbol{\theta}_0)\|^3 \|\xi\|^2 \leq CZ_n \mathbf{S}_n \|\xi\|^2 = O(\log n) O_p(n^{-1})$$

and we have

$$\xi = (1 + \nu) n^{-1/2} \mathbf{S}_n^{-1} \mathbf{P}_n + \boldsymbol{\beta}.$$

Now it is shown that

$$\sum_{t=1}^n Y_t = \xi' \sum_{t=1}^n \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) = (1 + \nu) \mathbf{P}_n' \mathbf{S}_n^{-1} \mathbf{P}_n + O(\log n) O_p(n^{-1/2}), \quad (2.23)$$

$$\begin{aligned} \sum_{t=1}^n Y_t^2 &= \xi' \left(\sum_{t=1}^n \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0)' \right) \xi \\ &= (1 + \nu)^2 \mathbf{P}_n' \mathbf{S}_n^{-1} \mathbf{P}_n + O(\log n) O_p(n^{-1/2}) \end{aligned} \quad (2.24)$$

$$\begin{aligned}
\sum_{t=1}^n Y_t^3 &\leq \max_{1 \leq t \leq n} |Y_t| \sum_{t=1}^n Y_t^2 = \|\xi\| \max_{1 \leq t \leq n} \|\mathbf{m}(\lambda_t; \boldsymbol{\theta}_0)\| \sum_{t=1}^n Y_t^2 \\
&= O(\log n) O_p(n^{-1/2}).
\end{aligned} \tag{2.25}$$

From (2.19) and (2.20), we can write

$$\begin{aligned}
CR_v(\mathbf{p}) &= \frac{2}{v(v+1)} \sum_{t=1}^n \{(np_t)^{-v} - 1\} \\
&= \frac{2}{v(v+1)} \sum_{t=1}^n \left[\left\{ \frac{n(1+Y_t)^{-1/(1+v)}}{\sum_{t=1}^n (1+Y_t)^{-1/(1+v)}} \right\}^{-v} - 1 \right] \\
&= \frac{2}{v(v+1)} \left[\left\{ \frac{1}{n} \sum_{t=1}^n (1+Y_t)^{-1/(1+v)} \right\}^v \left\{ \sum_{t=1}^n (1+Y_t)^{v/(1+v)} \right\} - n \right].
\end{aligned} \tag{2.26}$$

Taylor expansion leads to

$$\begin{aligned}
&\left\{ \frac{1}{n} \sum_{t=1}^n (1+Y_t)^{-1/(1+v)} \right\}^v \\
&= \left\{ 1 - \frac{1}{n(1+v)} \sum_{t=1}^n Y_t + \frac{2+v}{2n(1+v)^2} \sum_{t=1}^n Y_t^2 + \frac{1}{n} \sum_{t=1}^n O_p(Y_t^3) \right\}^v \\
&= 1 - \frac{v}{n(1+v)} \sum_{t=1}^n Y_t + \frac{v(2+v)}{2n(1+v)^2} \sum_{t=1}^n Y_t^2 + \frac{1}{n} \sum_{t=1}^n O_p(Y_t^3) + O_p(n^{-2})
\end{aligned} \tag{2.27}$$

and

$$\sum_{t=1}^n (1+Y_t)^{v/(1+v)} = n + \frac{v}{1+v} \sum_{t=1}^n Y_t - \frac{v}{2(1+v)^2} \sum_{t=1}^n Y_t^2 + \sum_{t=1}^n O_p(Y_t^3). \tag{2.28}$$

Using (2.23)–(2.25), (2.27) and (2.28), we can show that

$$(\text{A.13}) = \frac{1}{(1+v)^2} \sum_{t=1}^n Y_t^2 + O(\log n) O_p(n^{-1/2}) = \mathbf{P}'_n \mathbf{S}_n^{-1} \mathbf{P}_n + O(\log n) O_p(n^{-1/2}).$$

Finally, from Lemmas 2.1 and 2.2, we can show that

$$\mathbf{P}'_n \mathbf{S}_n^{-1} \mathbf{P}_n \xrightarrow{d} \left(\boldsymbol{\Sigma}_2^{-1/2} \boldsymbol{\Sigma}_1^{1/2} \boldsymbol{\Sigma}_1^{-1/2} \mathbf{P}_n \right)' \left(\boldsymbol{\Sigma}_2^{-1/2} \boldsymbol{\Sigma}_1^{1/2} \boldsymbol{\Sigma}_1^{-1/2} \mathbf{P}_n \right) \stackrel{d}{=} (\boldsymbol{\Sigma} \mathbf{N})' (\boldsymbol{\Sigma} \mathbf{N})$$

This is the desired result.

When $\nu = 0$, Theorem 2.3 is reduced to Theorem 2.2. When $\nu = -1$, the proof is similar. $\square$

References

- Baggerly, K.A.: Empirical likelihood as a goodness-of-fit measure. *Biometrika* **85**, 535–547 (1998)
- Belkacem, L., Vélh, J.L., Walter, C.: Capm, risk and portfolio selection in alpha-stable markets. *Fractals* **8**, 99–115 (2000)
- Brillinger, D.R.: *Time Series: Data Analysis and Theory*, expanded edn. Holden-Day, San Francisco (2001)
- Brockwell, P.J., Davis, R.A.: *Time Series: Theory and Methods*, 2nd edn. Springer, Berlin (1991)
- Dominicy, Y., Veredas, D.: The method of simulated quantiles. *J. Econometrics* **172**, 235–247 (2013)
- Embrechts, P., Klüppelberg, C., Mikosch, T.: *Modelling Extreme Events for Insurance and Finance*. Springer-Verlag, Berlin (1997)
- Fama, E.F.: The behavior of stock-market prices. *J. Bus.* **38**, 34–105 (1965)
- Hannan, E.J.: *Multiple Time Series*. John Wiley, New York (1970)
- Hansen, L.P., Heaton, J., Yaron, A.: Finite-sample properties of some alternative GMM estimators. *J. Bus. Econ. Statist.* **14**, 262–280 (1996)
- Kitamura, Y.: Empirical likelihood methods with weakly dependent processes. *Ann. Statist.* **25**, 2084–2102 (1997)
- Kitamura, Y., Stutzer, M.: An information-theoretic alternative to generalized method of moments estimation. *Econometrica* **65**, 861–874 (1997)
- Mandelbrot, B.: The variation of certain speculative prices. *J. Bus.* **36**, 394–419 (1963)
- Monti, A.C.: Empirical likelihood confidence regions in time series models. *Biometrika* **80**, 329–338 (1997)
- Nolan, J.P., Panorska, A.K., McCulloch, J.H.: Estimation of stable spectral measures. *Math. Comput. Model.* **34**, 1113–1122 (2001)
- Nordman, D.J., Lahiri, S.N.: A frequency domain empirical likelihood for short- and long-range dependence. *Ann. Statist.* **34**, 3019–3050 (2006)
- Newey, W.K., Smith, R.J.: Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica* **72**, 219–255 (2004)
- Ogata, H.: Optimal portfolio estimation for dependent financial returns with generalized empirical likelihood. *Adv. Decis. Sci.* **2012**, 8pages, Article ID 973173 (2012).
- Ogata, H.: Estimation for multivariate stable distributions with generalized empirical likelihood. *J. Econometrics* **172**, 248–254 (2013). doi:[10.1016/j.jeconom.2012.08.017](https://doi.org/10.1016/j.jeconom.2012.08.017)
- Ogata, H., Taniguchi, M.: Cressie-Read power-divergence statistics for non Gaussian vector stationary processes. *Scand. J. Stat.* **36**, 141–156 (2009)
- Ogata, H., Taniguchi, M.: An empirical likelihood approach for non-Gaussian vector stationary processes and its application to minimum contrast estimation. *Aust. N. Z. J. Stat.* **52**, 451–468 (2010)
- Owen, A.B.: Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**, 237–249 (1988)
- Owen, A.B.: Empirical likelihood confidence regions. *Ann. Statist.* **18**, 90–120 (1990)
- Owen, A.B.: *Empirical likelihood*. Chapman & Hall/CRC, Boca Raton, London, New York, Washington, DC (2001)
- Qin, J., Lawless, J.: Empirical likelihood and general estimating equations. *Ann. Statist.* **22**, 300–325 (1994)
- Rachev, S.T., Han, S.: Portfolio management with stable distributions. *Math. Method. Oper. Res.* **51**, 341–352 (2000)

- Read, T.R.C., Cressie, N.A.C.: Goodness-of-fit statistics for discrete multivariate data. Springer-Verlag, New York (1988)
- Yu, J.: Empirical characteristic function estimation and its applications. *Econometric Rev.* **23**, 93–123 (2004)

Statistical Inference for Financial Engineering

Taniguchi, M.; Amano, T.; Ogata, H.; Taniai, H.

2014, X, 118 p. 15 illus., 6 illus. in color., Softcover

ISBN: 978-3-319-03496-6