Ann Inst Stat Math (2017) 69:945–968
DOI 10.1007/s10463-016-0574-9

Estimation of the tail exponent of multivariate regular variation

Moosup Kim[1] · **Sangyeol Lee**[2]

Received: 25 July 2015 / Revised: 27 May 2016 / Published online: 18 July 2016

Abstract In this study, we consider the problem of estimating the tail exponent of multivariate regular variation. Since any convex combination of a random vector with a multivariate regularly varying tail has a univariate regularly varying tail with the same exponent under certain conditions, to estimate the tail exponent of the multivariate regular variation of a given random vector, we employ a weighted average of Hill's estimators obtained for all of its convex combinations, designed to reduce the variability of estimation. We investigate the asymptotic properties and evaluate the finite sample performance of the weighted average of Hill's estimators. A simulation study and real data analysis are provided for illustration.

Keywords Tail exponent · Multivariate regular variation · Hill's estimator · Empirical process theory

1 Introduction

Heavy tail phenomena are frequently observed in many applied fields and is characterized by the regularly varying tail of random observations with tail exponent, indicating the degree of hyperbolic decaying of the tail probability. The estimation of tail exponent has received much attention from researchers during the past decades and many

Electronic supplementary material The online version of this article (doi:10.1007/s10463-016-0574-9) contains supplementary material, which is available to authorized users.

✉ Sangyeol Lee
jpgrslee@yahoo.com

1 APEC Climate Center, Busan 48058, Korea

2 Department of Statistics, Seoul National University, Seoul 08826, Korea

estimators have been proposed in the literature: see, for example, Hill (1975), Hall (1982), Hall and Welsh (1985), Feuerverger and Hall (1999), Drees et al. (2000), and Gomes et al. (2008). Hill's estimator is the most popular among those estimators (cf. Hill 1975). In this study, we consider the problem of estimating the tail exponent of multivariate regular variation. It is well known that the tail exponent combined with the spectral measure characterizes the extreme behavior of multivariate observations (cf. Resnick 1987, Proposition 5.18). In the literature, little attention has been paid to the tail exponent estimation, compared with the spectral measure estimation, because the problem can be reduced to a univariate sample problem owing to the fact that the magnitudes of multivariate observations have a univariate regularly varying tail with the same tail exponent. However, the problem in the multivariate case deserves more attention because further information, unavailable in the univariate case, can be used to accurately estimate the tail exponent. This study is mainly inspired by the fact that every convex combination of a random vector having a multivariate regularly varying tail also has a univariate regularly varying tail with the same exponent under certain conditions. Using this fact, we first obtain Hill's estimator for each combination, and then, employ a weighted average of those Hill's estimators to reduce the variability of estimation. Below, we investigate the asymptotic properties and evaluate the finite sample performance of the weighted average of Hill's estimators.

The remainder of this paper is organized as follows. Section 2 formulates the weighted average of Hill's estimators and presents the relevant asymptotic results. Section 3 implements a simulation study and real data analysis for illustration. Section 4 summarizes the results obtained in this study. Section 5 provides the proofs of the theorems presented in Sect. 2.

2 Main result

Let $\{\boldsymbol{U}_i = (U_{i,1}, \ldots, U_{i,d})' : i \in \mathbb{Z}\}$ be an i.i.d. sequence of d-dimensional random vectors defined on a probability space $(\Omega, \mathcal{F}, P)$, $(d \in \mathbb{N})$. For brevity, we set $\boldsymbol{U} = \boldsymbol{U}_0$,

$$\mathbb{S}^{d-1} := \{\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_d)' \in [0,1]^d : \lambda_1 + \cdots + \lambda_d = 1\},$$

and

$$\boldsymbol{U}_i^{(\boldsymbol{\lambda})} := \boldsymbol{\lambda}'\boldsymbol{U}_i = \lambda_1 U_{i,1} + \cdots + \lambda_d U_{i,d}, \quad \text{for } \boldsymbol{\lambda} \in \mathbb{S}^{d-1}.$$

Let $F_{\boldsymbol{\lambda}}$ denote the distribution function of $\boldsymbol{U}^{(\boldsymbol{\lambda})}$ and set

$$b_{\boldsymbol{\lambda}}(x) := \inf\{y : F_{\boldsymbol{\lambda}}(y) \geq 1 - x^{-1}\},$$

which is the quantile function of $F_{\boldsymbol{\lambda}}$.

In what follows, we assume that there exist $\alpha > 0$ and a random vector $\boldsymbol{\Theta}$ such that for each $t > 0$,

$$\frac{P(|\boldsymbol{U}| > tx, \boldsymbol{U}/|\boldsymbol{U}| \in \cdot)}{P(|\boldsymbol{U}| > x)} \xrightarrow{\text{v}} t^{-\alpha} P(\boldsymbol{\Theta} \in \cdot) \quad \text{as } x \to \infty, \tag{1}$$

where $|\cdot|$ is the usual norm in $\mathbb{R}^d$ and $\xrightarrow{\mathrm{v}}$ denotes the vague convergence in the unit sphere [for vague convergence, we refer to Kallenberg (1983)]. In this case, the distribution function of $\boldsymbol{U}$ is said to have a multivariate regularly varying tail with tail exponent $-\alpha$, and the distribution of $\boldsymbol{\Theta}$ is called its spectral measure that represents the dependency among the components of $\boldsymbol{U}$ at the extreme level. Particularly, because it follows from some mild conditions that

$$\frac{P(|\boldsymbol{U}| > tx)}{P(|\boldsymbol{U}| > x)} \to t^{-\alpha}, \quad \frac{P(\boldsymbol{U}^{(\boldsymbol{\lambda})} > x)}{P(|\boldsymbol{U}| > x)} \to \alpha \int_{A_{\boldsymbol{\lambda}}} \frac{\mathrm{d}s}{s^{\alpha+1}} P(\boldsymbol{\Theta} \in d\theta) =: w(\boldsymbol{\lambda}) \qquad \text{as } x \to \infty, \tag{2}$$

where $A_{\boldsymbol{\lambda}} = \{(s, \theta) : s = |\boldsymbol{u}|, \theta = \boldsymbol{u}/|\boldsymbol{u}|, \boldsymbol{u}^{(\boldsymbol{\lambda})} > 1\}$ $(\boldsymbol{u}^{(\boldsymbol{\lambda})} = \boldsymbol{\lambda}'\boldsymbol{u})$, $F_{\boldsymbol{\lambda}}$ has a (univariate) regularly varying tail with the same tail exponent $-\alpha$ provided that $w(\boldsymbol{\lambda}) \neq 0$.

Suppose that $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_n$ are observed ($n \in \mathbb{N}$). Let

$$\mathrm{H}_n(\boldsymbol{\lambda}, y) := \frac{1}{yk} \sum_{i=1}^{n} \left(\log \boldsymbol{U}_i^{(\boldsymbol{\lambda})} - \log \boldsymbol{U}_{(yk+1)}^{(\boldsymbol{\lambda})} \right)_+, \quad \boldsymbol{\lambda} \in \mathbb{S}^{d-1}, \ y \in (0, \infty), \tag{3}$$

where $k = k_n$ is a positive integer varying according to sample size n with

$$k \to \infty \qquad \text{and} \qquad k = o(n) \quad \text{as } n \to \infty, \tag{4}$$

and $\boldsymbol{U}_{(yk+1)}^{(\boldsymbol{\lambda})}$ is the $\lfloor yk + 1 \rfloor$th largest order statistic in $\boldsymbol{U}_1^{(\boldsymbol{\lambda})}, \ldots, \boldsymbol{U}_n^{(\boldsymbol{\lambda})}$. Then, $\mathrm{H}_n(\boldsymbol{\lambda}, y)$ is called Hill's estimator. For every $\boldsymbol{\lambda} \in \mathbb{S}^{d-1}$, $\mathrm{H}_n(\boldsymbol{\lambda}, y)$ converges in probability to $1/\alpha$ as $n \to \infty$ and is asymptotically normal under certain conditions, provided that $w(\boldsymbol{\lambda}) \neq 0$ (cf. Hall 1982).

It can be anticipated that the tail exponent can be more accurately estimated by averaging $\mathrm{H}_n(\boldsymbol{\lambda}, y)$ with respect to a suitable weight function of $\boldsymbol{\lambda}$. More precisely, for $\varphi \in \Gamma$ and $y > 0$, where Γ is the class of signed measures φ on $\mathbb{S}^{d-1}$ with $\varphi(\mathbb{S}^{d-1}) = 1$, we consider

$$\mathrm{H}_n(\varphi, y) := \int_{\mathbb{S}^{d-1}} \mathrm{H}_n(\boldsymbol{\lambda}, y) \varphi(\mathrm{d}\boldsymbol{\lambda}). \tag{5}$$

In fact, Dematteo and Clémençon (2016) introduced an estimator similar to $\mathrm{H}_n(\varphi, y)$. They considered a convex linear combination of componentwise Hill's estimators, proved an asymptotic normality result for their estimator, and addressed the problem of selecting the optimal $\boldsymbol{\lambda}$ that would minimize the mean squared error.

Below, we investigate the asymptotic property of $\mathrm{H}_n(\varphi, y)$ under the regularity conditions:

A1 For every $\boldsymbol{\lambda} \in \mathbb{S}^{d-1}$, $F_{\boldsymbol{\lambda}}$ is continuous. There exist $\gamma_{\boldsymbol{\lambda}} < 0$, $C_{\boldsymbol{\lambda}} > 0$ and $D_{\boldsymbol{\lambda}} \in \mathbb{R}\backslash\{0\}$ such that

$$\bar{F}_{\boldsymbol{\lambda}}(x) := 1 - F_{\boldsymbol{\lambda}}(x) = C_{\boldsymbol{\lambda}} x^{-\alpha}\left(1 + D_{\boldsymbol{\lambda}} x^{\gamma_{\boldsymbol{\lambda}}} + x^{\gamma_{\boldsymbol{\lambda}}} \Delta_{\boldsymbol{\lambda}}(x)\right), \tag{6}$$

where $\Delta_{\boldsymbol{\lambda}}(x)$ is differentiable in x such that

$$\sup_{\boldsymbol{\lambda}\in\mathbb{S}^{d-1}} |\Delta_{\boldsymbol{\lambda}}(x)| = o(1), \quad \sup_{\boldsymbol{\lambda}\in\mathbb{S}^{d-1}} \left|\frac{\partial \Delta_{\boldsymbol{\lambda}}}{\partial x}(x)\right| = o(x^{-1}) \quad \text{as } x \to \infty.$$

Moreover, $\boldsymbol{\lambda} \mapsto D_{\boldsymbol{\lambda}}$ and $\boldsymbol{\lambda} \mapsto \gamma_{\boldsymbol{\lambda}}$ are bounded and away from 0, $\boldsymbol{\lambda} \mapsto C_{\boldsymbol{\lambda}}$ is Lipschitz-continuous, and there exists $M_{\boldsymbol{\lambda}} \in [0, \infty)$ such that

$$M_{\boldsymbol{\lambda}} = \lim_{n\to\infty} \sqrt{k}\,\{b_{\boldsymbol{\lambda}}(n/k)\}^{\gamma_{\boldsymbol{\lambda}}}. \tag{7}$$

A2 Let $0 < \underline{y} < 1 < \bar{y} < \infty$ and

$$T := \mathbb{S}^{d-1} \times [\underline{y}, \bar{y}]. \tag{8}$$

There exist real-valued functions $\mathrm{c}_1, \mathrm{c}_2, \mathrm{c}_3$ defined on $T \times T$ such that

$$\mathrm{c}_1((\boldsymbol{\lambda}_1, y_1), (\boldsymbol{\lambda}_2, y_2)) = \lim_{n\to\infty} \frac{n}{k} P\left(Y_n(\boldsymbol{\lambda}_1, y_1) > 0, Y_n(\boldsymbol{\lambda}_2, y_2) > 0\right),$$
$$\mathrm{c}_2((\boldsymbol{\lambda}_1, y_1), (\boldsymbol{\lambda}_2, y_2)) = \lim_{n\to\infty} \frac{n\alpha^2}{k} \mathrm{E}\left\{Y_n(\boldsymbol{\lambda}_1, y_1) Y_n(\boldsymbol{\lambda}_2, y_2)\right\},$$
$$\mathrm{c}_3((\boldsymbol{\lambda}_1, y_1), (\boldsymbol{\lambda}_2, y_2)) = \lim_{n\to\infty} \frac{n\alpha}{k} \mathrm{E}\left\{Y_n(\boldsymbol{\lambda}_1, y_1) \mathrm{I}\left(Y_n(\boldsymbol{\lambda}_2, y_2) > 0\right)\right\},$$

where $Y_n(\boldsymbol{\lambda}, y) := (\log \boldsymbol{U}^{(\boldsymbol{\lambda})} - \log b_{\boldsymbol{\lambda}}(n/k) + \frac{1}{\alpha}\log y)_+$.
A3 There exists $C > 0$ such that $P(|\boldsymbol{U}| > x) \sim Cx^{-\alpha}$ as $x \to \infty$.

Remark 1 (6) is an example of a second-order regular variation with a smooth remainder that is uniformly negligible in $\boldsymbol{\lambda} \in \mathbb{S}^{d-1}$. For the details of this condition, we refer to Goldie and Smith (1987), Hall (1982), Hall and Welsh (1985) and Feuerverger and Hall (1999). Under **A1**–**A3** and additional mild conditions, we have $C_{\boldsymbol{\lambda}} = w(\boldsymbol{\lambda})C$ and

$$\mathrm{c}_1((\boldsymbol{\lambda}_1, y_1), (\boldsymbol{\lambda}_2, y_2)) = \alpha \int \mathrm{I}\left(h_1(s,\theta) \wedge h_2(s,\theta) > 0\right) \frac{\mathrm{d}s}{s^{\alpha+1}} P(\boldsymbol{\Theta} \in \mathrm{d}\theta),$$
$$\mathrm{c}_2((\boldsymbol{\lambda}_1, y_1), (\boldsymbol{\lambda}_2, y_2)) = \alpha^3 \int (h_1(s,\theta))_+ (h_2(s,\theta))_+ \frac{\mathrm{d}s}{s^{\alpha+1}} P(\boldsymbol{\Theta} \in \mathrm{d}\theta),$$
$$\mathrm{c}_3((\boldsymbol{\lambda}_1, y_1), (\boldsymbol{\lambda}_2, y_2)) = \alpha^2 \int (h_1(s,\theta))_+ \mathrm{I}\left(h_2(s,\theta) > 0\right) \frac{\mathrm{d}s}{s^{\alpha+1}} P(\boldsymbol{\Theta} \in \mathrm{d}\theta),$$

where $h_i(s,\theta) = \log(s\boldsymbol{\lambda}_i'\theta) - \frac{1}{\alpha}\log w(\boldsymbol{\lambda}_i) + \frac{1}{\alpha}\log y_i$ for $i = 1, 2$. It is readily checked that

$$\mathrm{c}_i((\boldsymbol{\lambda}_1, y), (\boldsymbol{\lambda}_2, y)) = y\mathrm{c}_i((\boldsymbol{\lambda}_1, 1), (\boldsymbol{\lambda}_2, 1)) \quad \text{for } i = 1, 2, 3 \text{ and } y > 0. \tag{9}$$

More explicitly, $w(\boldsymbol{\lambda}) = \int (\boldsymbol{\lambda}'\theta)^\alpha P(\boldsymbol{\Theta} \in \mathrm{d}\theta)$, where the integration is taken over the unit sphere. If $\alpha \geq 1$, both $w(\boldsymbol{\lambda})$ and $C_{\boldsymbol{\lambda}}$ are Lipschitz-continuous.

In what follows, G denotes a Gaussian process indexed by T with zero mean and covariance function

$$\begin{aligned}\mathrm{V}((\boldsymbol{\lambda}_1, y_1), (\boldsymbol{\lambda}_2, y_2)) := \frac{1}{\alpha^2}\left\{\mathrm{c}_2((\boldsymbol{\lambda}_1, y_1), (\boldsymbol{\lambda}_2, y_2)) + \mathrm{c}_1((\boldsymbol{\lambda}_1, y_1), (\boldsymbol{\lambda}_2, y_2))\right\}\\ - \frac{1}{\alpha^2}\left\{\mathrm{c}_3((\boldsymbol{\lambda}_1, y_1), (\boldsymbol{\lambda}_2, y_2)) + \mathrm{c}_3((\boldsymbol{\lambda}_2, y_2), (\boldsymbol{\lambda}_1, y_1))\right\}.\end{aligned} \tag{10}$$

For a given metric space S, let $\ell^\infty(S)$ denote the metric space of all real-valued and bounded functions defined on S endowed with the uniform metric. For the theory of weak convergence on these function spaces and further details, we refer to van der Vaart and Wellner (1996). Below, we present a theorem that describes the asymptotic behavior of $\mathrm{H}_n(\varphi, y)$. Its proof is provided in Sect. 5.

Theorem 1 *Assume that* (1) *and* **A1–A3** *hold. Then, for* $\varphi \in \Gamma$,

$$\begin{aligned}&y\sqrt{k}\left\{\mathrm{H}_n(\varphi, y) - \frac{1}{\alpha}\right\}\\ &\rightsquigarrow \int_{\mathbb{S}^{d-1}} \mathrm{G}(\boldsymbol{\lambda}, y)\varphi(\mathrm{d}\boldsymbol{\lambda}) + \int_{\mathbb{S}^{d-1}} \frac{y^{-\frac{\gamma_{\boldsymbol{\lambda}}}{\alpha}+1}\gamma_{\boldsymbol{\lambda}} M_{\boldsymbol{\lambda}} D_{\boldsymbol{\lambda}}}{\alpha(\alpha-\gamma_{\boldsymbol{\lambda}})}\varphi(\mathrm{d}\boldsymbol{\lambda}), \quad \text{in } \ell^\infty([\underline{y}, \bar{y}]),\end{aligned} \tag{11}$$

provided that $\int_{\mathbb{S}^{d-1}} |\frac{\gamma_{\boldsymbol{\lambda}} M_{\boldsymbol{\lambda}} D_{\boldsymbol{\lambda}}}{\alpha(\alpha-\gamma_{\boldsymbol{\lambda}})}|\varphi(\mathrm{d}\boldsymbol{\lambda}) < \infty$.

Remark 2 The problem of choosing the tail sample fraction yk/n is important and difficult in estimating the tail exponent. In the univariate case, this problem is considered by several authors such as Hall and Welsh (1985), Drees et al. (2000), and Danielsson et al. (2001). For the multivariate case, an argument similar to that of Hall (1982) can be established. For simplicity, we assume that $\gamma = \gamma_{\boldsymbol{\lambda}}$ does not depend on $\boldsymbol{\lambda}$ and take $k = \lfloor n^{2\rho/(2\rho+1)} \rfloor$ with $\rho = -\gamma/\alpha$. Then, it follows from (9) and (11) that

$$\sqrt{k}\left\{\mathrm{H}_n(\varphi, y) - \frac{1}{\alpha}\right\} \rightsquigarrow \mathrm{N}\left(y^\rho \int_{\mathbb{S}^{d-1}} \frac{\gamma M_{\boldsymbol{\lambda}} D_{\boldsymbol{\lambda}}}{\alpha(\alpha-\gamma)}\varphi(\mathrm{d}\boldsymbol{\lambda}), \frac{\int\int \mathrm{V}\varphi(\mathrm{d}\boldsymbol{\lambda}_1)\varphi(\mathrm{d}\boldsymbol{\lambda}_2)}{y}\right),$$

where

$$\int\int \mathrm{V}\varphi(\mathrm{d}\boldsymbol{\lambda}_1)\varphi(\mathrm{d}\boldsymbol{\lambda}_2) = \int_{\boldsymbol{\lambda}_1\in\mathbb{S}^{d-1}} \int_{\boldsymbol{\lambda}_2\in\mathbb{S}^{d-1}} \mathrm{V}((\boldsymbol{\lambda}_1, 1), (\boldsymbol{\lambda}_2, 1))\varphi(\mathrm{d}\boldsymbol{\lambda}_1)\varphi(\mathrm{d}\boldsymbol{\lambda}_2).$$

Thus, the asymptotic mean squared error is given by

$$y^{2\rho}\left\{\int_{\mathbb{S}^{d-1}} \frac{\gamma M_{\boldsymbol{\lambda}} D_{\boldsymbol{\lambda}}}{\alpha(\alpha-\gamma)}\varphi(\mathrm{d}\boldsymbol{\lambda})\right\}^2 + \frac{\int\int \mathrm{V}\varphi(\mathrm{d}\boldsymbol{\lambda}_1)\varphi(\mathrm{d}\boldsymbol{\lambda}_2)}{y}. \tag{12}$$

Then, if $0 < \{\int_{\mathbb{S}^{d-1}} \frac{\gamma M_{\boldsymbol{\lambda}} D_{\boldsymbol{\lambda}}}{\alpha(\alpha-\gamma)}\varphi(\mathrm{d}\boldsymbol{\lambda})\}^2 < \infty$, we might be able to attain the asymptotic efficiency by finding y that minimizes the above. However, it is not feasible to estimate the optimal tail fraction owing to the difficulty in estimating several second-order parameters and integrating the function V. Here, we do not pursue to deal with this issue because it is beyond the scope of this study.

Remark 3 Since the performance of $\mathrm{H}_n(\varphi, 1)$ depends on φ, its choice is also crucial to estimate α. A simple one is the uniform measure φ_{unif} on $\mathbb{S}^{d-1}$. However, it does not guarantee the best performance. The best solution φ° may be obtained by minimizing (12) with respect to $\varphi \in \Gamma$, where y is fixed as 1. More precisely, we can write

$$\varphi^\circ := \underset{\varphi\in\Gamma}{\arg\min}\left\{\left\{\int_{\mathbb{S}^{d-1}} \frac{\gamma M_{\boldsymbol{\lambda}} D_{\boldsymbol{\lambda}}}{\alpha(\alpha-\gamma)}\varphi(\mathrm{d}\boldsymbol{\lambda})\right\}^2 + \int\int \mathrm{V}\varphi(\mathrm{d}\boldsymbol{\lambda}_1)\varphi(\mathrm{d}\boldsymbol{\lambda}_2)\right\}, \tag{13}$$

provided that the minimizer exists and is unique. Note that if $M_{\boldsymbol{\lambda}} = 0$ for every $\boldsymbol{\lambda} \in \mathbb{S}^{d-1}$, (13) is reduced to

$$\varphi^\circ = \underset{\varphi\in\Gamma}{\arg\min}\int\int \mathrm{V}\varphi(\mathrm{d}\boldsymbol{\lambda}_1)\varphi(\mathrm{d}\boldsymbol{\lambda}_2).$$

However, this method requires a full knowledge on V, as well as the second-order parameters, as seen in choosing the optimal tail fraction.

To overcome this difficulty, one can consider using some discrete signed measures, defined on a fine grid, that approximate $\varphi \in \Gamma$. For $r \in \mathbb{N}$, we set

$$\left\{\left(\frac{t_1}{r}, \ldots, \frac{t_d}{r}\right)' : t_1, \ldots, t_d \in \{0, 1, \ldots, r\} \text{ and } t_1 + \cdots + t_d = r\right\} = \{\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_m\}, \tag{14}$$

where $m = \binom{r+d-1}{d-1}$ denotes the number of the grid points, and

$$\varphi_{\text{disc}}(\cdot) = \sum_{i=1}^{m} \varphi_i \mathrm{I}(\boldsymbol{\lambda}_i \in \cdot),$$

where $\varphi_1, \ldots, \varphi_m \in \mathbb{R}$ with $\varphi_1 + \cdots + \varphi_m = 1$ (so that $\varphi_{\text{disc}} \in \Gamma$). In this case, $\mathrm{H}_n(\varphi_{\text{disc}}, 1) = \sum_{i=1}^m \varphi_i \mathrm{H}_n(\boldsymbol{\lambda}_i, 1)$. Suppose that $M_{\boldsymbol{\lambda}} = 0$ for every $\boldsymbol{\lambda} \in \mathbb{S}^{d-1}$. Then, putting

$$\Sigma = \begin{pmatrix} \mathrm{V}((\boldsymbol{\lambda}_1, 1), (\boldsymbol{\lambda}_1, 1)) & \cdots & \mathrm{V}((\boldsymbol{\lambda}_1, 1), (\boldsymbol{\lambda}_m, 1)) \\ \vdots & \ddots & \vdots \\ \mathrm{V}((\boldsymbol{\lambda}_m, 1), (\boldsymbol{\lambda}_1, 1)) & \cdots & \mathrm{V}((\boldsymbol{\lambda}_m, 1), (\boldsymbol{\lambda}_m, 1)) \end{pmatrix},$$

and $\left(\varphi_1^\circ, \ldots, \varphi_m^\circ\right)' = \{\mathbf{1}'\Sigma^{-1}\mathbf{1}\}^{-1}\Sigma^{-1}\mathbf{1}$, ($\mathbf{1} = (1, \ldots, 1)' \in \mathbb{R}^m$), provided that Σ is not singular, we get

$$\varphi_{\text{disc}}^\circ(\cdot) = \sum_{i=1}^{m} \varphi_i^\circ \mathrm{I}(\boldsymbol{\lambda}_i \in \cdot).$$

Then, the asymptotic variance of $\mathrm{H}_n(\varphi_{\text{disc}}, 1)$ is minimized at $\varphi_{\text{disc}} = \varphi^\circ_{\text{disc}}$. As an estimate of Σ, one can employ

$$\hat{\mathrm{c}}_1(\boldsymbol{\lambda}_i, \boldsymbol{\lambda}_j) = \frac{1}{k}\sum_{l=1}^{n} \mathrm{I}(\tilde{Y}_{n,l}(\boldsymbol{\lambda}_i) > 0, \tilde{Y}_{n,l}(\boldsymbol{\lambda}_j) > 0),$$
$$\hat{\mathrm{c}}_2(\boldsymbol{\lambda}_i, \boldsymbol{\lambda}_j) = \frac{\hat{\alpha}^2}{k}\sum_{l=1}^{n} \tilde{Y}_{n,l}(\boldsymbol{\lambda}_i)\tilde{Y}_{n,l}(\boldsymbol{\lambda}_j),$$
$$\hat{\mathrm{c}}_3(\boldsymbol{\lambda}_i, \boldsymbol{\lambda}_j) = \frac{\hat{\alpha}}{k}\sum_{l=1}^{n} \tilde{Y}_{n,l}(\boldsymbol{\lambda}_i)\mathrm{I}(\tilde{Y}_{n,l}(\boldsymbol{\lambda}_j) > 0),$$

where $\hat{\alpha} = 1/\mathrm{H}_n(\varphi_{\text{unif}}, 1)$ and $\tilde{Y}_{n,l}(\boldsymbol{\lambda}) = (\log \boldsymbol{U}_l^{(\boldsymbol{\lambda})} - \log \boldsymbol{U}_{(k+1)}^{(\boldsymbol{\lambda})})_+$ for $\boldsymbol{\lambda} \in \mathbb{S}^{d-1}$ and $l = 1, \ldots, n$. Then, we obtain

$$\left(\hat{\varphi}_1, \ldots, \hat{\varphi}_m\right)' := \{\mathbf{1}'\hat{\Sigma}^{-1}\mathbf{1}\}^{-1}\hat{\Sigma}^{-1}\mathbf{1}, \quad \hat{\varphi}_{\text{disc}}(\cdot) := \sum_{i=1}^{m} \hat{\varphi}_i \mathrm{I}(\boldsymbol{\lambda}_i \in \cdot). \tag{15}$$

In Sect. 3, we evaluate the performance of $\mathrm{H}_n(\varphi_{\text{unif}})$ and $\mathrm{H}_n(\hat{\varphi}_{\text{disc}})$.

Note that the r determines the grid resolution of (14). Although $\varphi^\circ_{\text{disc}}$ is closer to φ° as the r gets larger, the computation of $\hat{\varphi}_{\text{disc}}$ becomes more complicated as well. Our simulation study, not reported here in details, shows that the performance of $\mathrm{H}_n(\hat{\varphi}_{\text{disc}}, 1)$ varies a little according as the (large) r gets increased. Also, it reveals that $\mathrm{H}_n(\hat{\varphi}_{\text{disc}}, 1)$ performs reasonably when $r \le 50$ in the case of $n = 2000$ and $d \le 3$. In general, it is quite hard finding a rule to choose an optimal r owing to the complexity in computation. Considering its importance in implementation, we leave this issue as our future project.

3 Simulation study and real data analysis

3.1 Simulation study

In this subsection, we conduct a simulation study to evaluate the performance of the proposed estimator in finite samples. Let $\{\boldsymbol{U}_i\}$ be i.i.d. random vectors. Each $\boldsymbol{U}_i$ is assumed to follow an elliptical hyperbolic distribution:

$$\boldsymbol{U}_i \sim \frac{(Z_1^+, \ldots, Z_d^+)}{\sqrt{V/\alpha}}, \quad (Z_1, \ldots, Z_d) \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}), \qquad V \sim \chi^2(\alpha), \tag{16}$$

where $(Z_1, \ldots, Z_d)$ and V are independent, and $\boldsymbol{\Sigma}$ is a $d \times d$ positive definite matrix. Then, $\boldsymbol{U}_i$ have a multivariate regularly varying tail with tail exponent $-\alpha$. To evaluate the performance of $\mathrm{H}_n(\varphi, 1)$, we consider the mean squared error (MSE), $\mathrm{E}\{\log(\alpha\mathrm{H}_n(\varphi, 1))\}^2$. For comparison, we consider $\mathrm{H}_n(\varphi_{\text{unif}}, 1)$, $\mathrm{H}_n(\hat{\varphi}_{\text{disc}}, 1)$, and

$$\mathrm{H}_n^* = \frac{1}{k}\sum_{i=1}^{n}\left(\log|\boldsymbol{U}_i| - \log|\boldsymbol{U}|_{(k+1)}\right)_+, \tag{17}$$

where $|\boldsymbol{U}|_{(k+1)}$ denotes the $(k+1)$th largest order statistic of $|\boldsymbol{U}_1|, \ldots, |\boldsymbol{U}_n|$. Note that H_n^* is utilized in Mainik and Rüschendorf (2010). The MSEs are calculated in the settings as follows:

(i) $n = 2000$, $d = 2$, $\alpha = 3$, and

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}.$$

$\hat{\varphi}_{\mathrm{disc}}$ is supported by $\{(\frac{t_1}{r}, \frac{t_2}{r})' : t_1, t_2 \in \{0, 1, \ldots, r\}$ and $t_1 + t_2 = r\}$ with $r = 10$.

(ii) the same setting as in (i) except $d = 3$ and

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.5 & 0.1 \\ 0.5 & 1 & 0.5 \\ 0.1 & 0.5 & 1 \end{bmatrix}.$$

$\hat{\varphi}_{\mathrm{disc}}$ is supported by $\{(\frac{t_1}{r}, \frac{t_2}{r}, \frac{t_3}{r})' : t_1, t_2, t_3 \in \{0, 1, \ldots, r\}$ and $t_1 + t_2 + t_3 = r\}$ with $r = 10$.

Figures 1 and 2 present the square roots of MSE of the estimators and the relative efficiency of $\mathrm{H}_n(\hat{\varphi}_{\mathrm{disc}}, 1)$ in setting (i), which indicates the ratio of the MSEs of $\mathrm{H}_n(\hat{\varphi}_{\mathrm{disc}}, 1)$ (numerator) relative to those of its counterparts (denominator). It can be seen that $\mathrm{H}_n(\hat{\varphi}_{\mathrm{disc}}, 1)$ is the most efficient, among them, at $k = 50$. Figures 3 and 4 show the same in setting (ii) as Figs. 1 and 2, wherein $\mathrm{H}_n(\hat{\varphi}_{\mathrm{disc}}, 1)$ also reveals the best performance at $k = 80$. All these results confirm the validity of $\mathrm{H}_n(\hat{\varphi}_{\mathrm{disc}}, 1)$.

3.2 Real data analysis

In this subsection, we conduct a real data analysis. We analyze the negative returns

$$(X_{1,1}, X_{1,2}), \ldots, (X_{n,1}, X_{n,2}), \quad n = 2039,$$

of the stock prices of the Apple and Google from March 2005 to April 2013. Figure 5 shows that the negative returns have some conditional heteroscedasticity. Therefore, we fit a GARCH(1,1) model to each series, that is,

$$\begin{aligned} X_{i,j} &= \mu_j + Z_{i,j}\sqrt{h_{i,j}}, \\ h_{i,j} &= \omega_j + \phi_j\{X_{i-1,j} - \mu_j\}^2 + \beta_j h_{i-1,j}, \; i \in \mathbb{Z}, \; j = 1, 2, \end{aligned}$$

where $(Z_{i,1}, Z_{i,2})', i \in \mathbb{Z}$ are i.i.d. random vectors. We estimate the parameters by the quasi-MLE $\hat{\mu}_j, \hat{\omega}_j, \hat{\phi}_j, \hat{\beta}_j$, $j = 1, 2$, using a two-sided exponential likelihood function as in Berkes and Horváth (2004). Then, we obtain the (standardized) residuals:

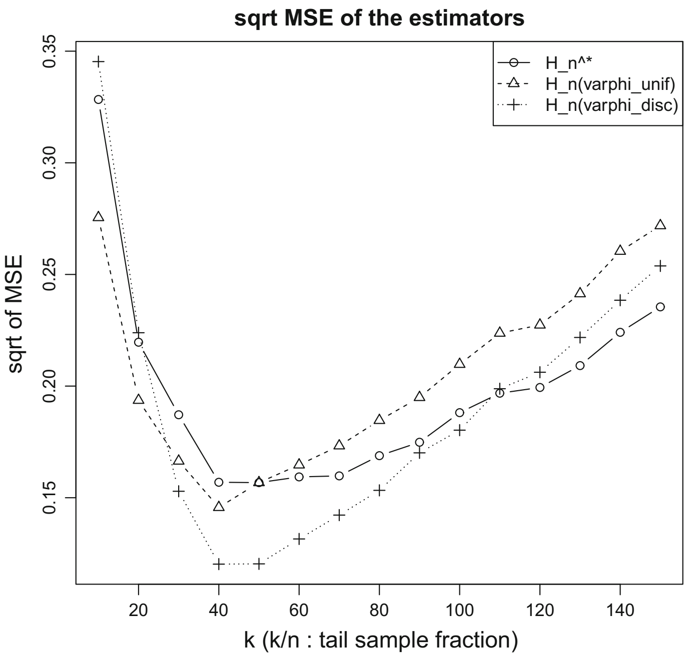

Fig. 1 The square roots of the MSEs in setting (i)

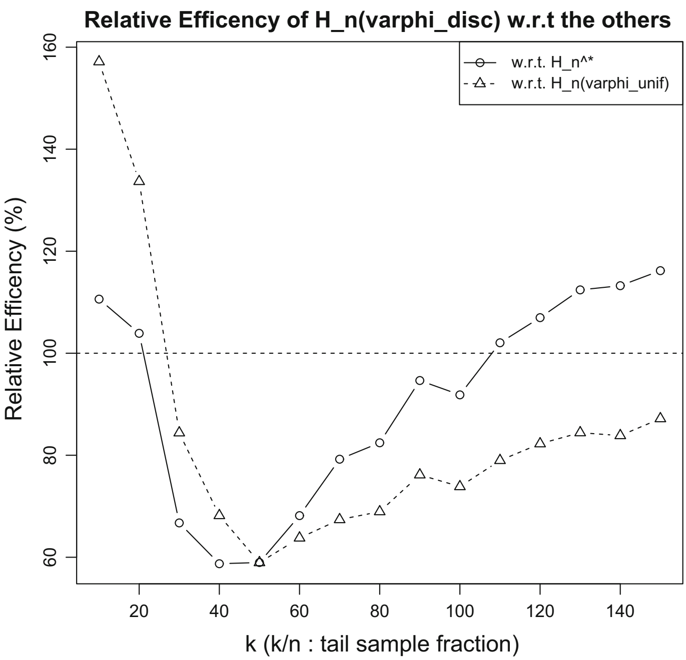

Fig. 2 The relative efficiency of $\mathrm{H}_n(\hat{\varphi}_{\mathrm{disc}}, 1)$ with respect to H_n^* and $\mathrm{H}_n(\varphi_{\mathrm{unif}}, 1)$ in setting (i)

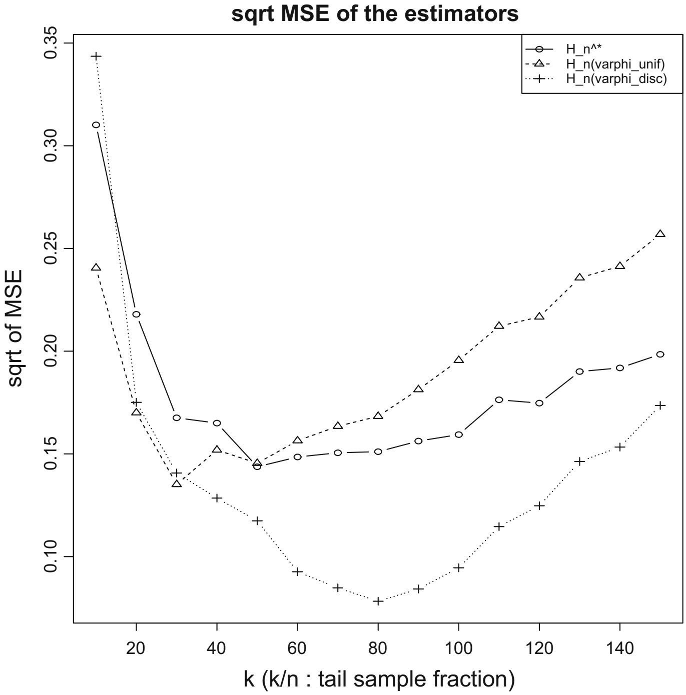

Fig. 3 The square roots of the MSEs in setting (ii)

$$\hat{Z}_{i,j} = \{X_{i,j} - \hat{\mu}_j\}/\hat{h}_{i,j}^{1/2}, \quad i = 1, \ldots, n, \ j = 1, 2,$$

where $\hat{h}_{i,j}$ are recursively obtained from the equations:

$$\hat{h}_{1,j} = \hat{\omega}_j, \quad \hat{h}_{i,j} = \hat{\omega}_j + \hat{\phi}_j\{X_{i-1,j} - \hat{\mu}_j\}^2 + \hat{\beta}_j\hat{h}_{i-1,j}, \qquad i = 2, \ldots, n, \ j = 1, 2.$$

Figure 6 presents the time series plot of the residuals, whereas Figs. 7 and 8, respectively, show their and their squares' correlogram and cross-correlogram. Since the negative returns are seemingly well filtered out, we view the residuals as i.i.d. random vectors. Figure 9 shows the scatter plot of the residuals, wherein several extreme values are observed, allowing for the existence of a multivariate regularly varying tail.

To deal with extreme financial risk, we project the residuals into the first quadrant to obtain $U_{i,j} = \hat{Z}_{i,j}^{+}$, $i = 1, \ldots, n$ and $j = 1, 2$. Now, we estimate the tail exponent of multivariate regularly varying tail of $\boldsymbol{U}_i = (U_{i,1}, U_{i,2})'$ by H_n^*, $\mathrm{H}_n(\varphi_{\mathrm{unif}}, 1)$, and $\mathrm{H}_n(\hat{\varphi}_{\mathrm{disc}}, 1)$. Here, $\hat{\varphi}_{\mathrm{disc}}$ is supported by $\{(\frac{t_1}{r}, \frac{t_2}{r})' : t_1, t_2 \in \{0, 1, \ldots, r\}$and $t_1 + t_2 = r\}$ with $r = 50$. Figure 10 exhibits the estimates against tail sample fraction k/n. The estimates appear to be relatively stable when k ranges from 20 to 70, but the values for α appear to be slightly different, that is, $1/\mathrm{H}_n(\varphi_{\mathrm{unif}}, 1)$ and $1/\mathrm{H}_n(\hat{\varphi}_{\mathrm{disc}}, 1)$ lie between 3.5 and 4 whereas $1/\mathrm{H}_n^*$ lies between 3 and 3.5. Figures 11, 12 and 13 present the confidence intervals at level 95 %. Among them, $1/\mathrm{H}_n(\hat{\varphi}_{\mathrm{disc}}, 1)$ appears to have the

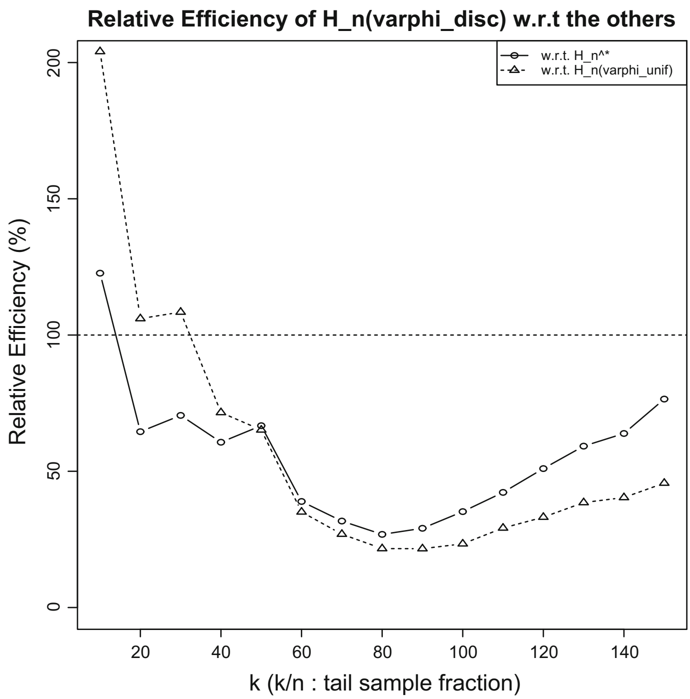

Fig. 4 The relative efficiency of $\mathrm{H}_n(\hat{\varphi}_{\mathrm{disc}}, 1)$ with respect to H_n^* and $\mathrm{H}_n(\varphi_{\mathrm{unif}}, 1)$ in setting (ii)

narrowest one, ranging from 3 to 4.5, which suggests that the negative tail exponent lies between 3 and 4.5.

4 Conclusion and discussion

In this paper, we considered the estimation of the tail exponent of multivariate regularly varying tail. As an estimator, a weighted average of Hill's estimators is proposed. We investigated its asymptotic properties and evaluated its finite sample performance through a simulation study. As a result, the proposed estimator $\mathrm{H}_n(\hat{\varphi}_{\mathrm{disc}}, 1)$ appeared to outperform the classical one in (17). Our real data analysis also confirms the validity of our method. In fact, it is well known in the literature that many other existing estimators perform better than Hill's estimator in estimating the tail exponent of univariate regularly varying tail. This suggests that weighted averages of such estimators would perform properly in the multivariate case as well. Thus, it would be naturally interesting to compare their performance with that of our estimator. We leave this issue as our future project.

5 Proofs

5.1 Preliminary results

In this section, we prove Theorem 1 presented in Sect. 2 using Propositions 1–3 below. In what follows, we assume that (1), (4), and **A1–A3** hold. Further, K denotes a generic

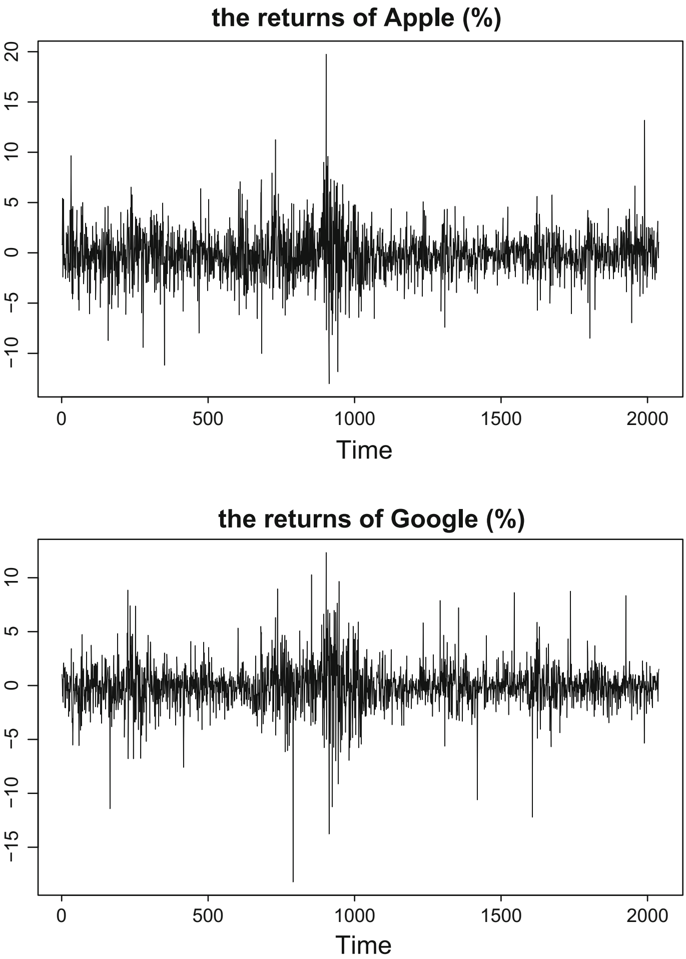

Fig. 5 Time series plots of the negative returns (%)

positive constant and $\|\cdot\|_2 = \sqrt{\mathrm{E}\{(\cdot)^2\}}$. Below are some preliminary lemmas useful to prove Theorem 1. The proofs of all the lemmas in this section are provided in the supplementary material.

Lemma 1 *As $x \to \infty$,*

$$b_{\boldsymbol{\lambda}}(x) = C_{\boldsymbol{\lambda}}^{1/\alpha} x^{1/\alpha} \left\{ 1 + \frac{C_{\boldsymbol{\lambda}}^{\gamma_{\boldsymbol{\lambda}}/\alpha} D_{\boldsymbol{\lambda}}}{\alpha} x^{\gamma_{\boldsymbol{\lambda}}/\alpha} + o\left(x^{\gamma_{\boldsymbol{\lambda}}/\alpha}\right) \right\} \tag{18}$$

uniformly in $\boldsymbol{\lambda} \in \mathbb{S}^{d-1}$. Thus, (7) holds uniformly in $\boldsymbol{\lambda} \in \mathbb{S}^{d-1}$.

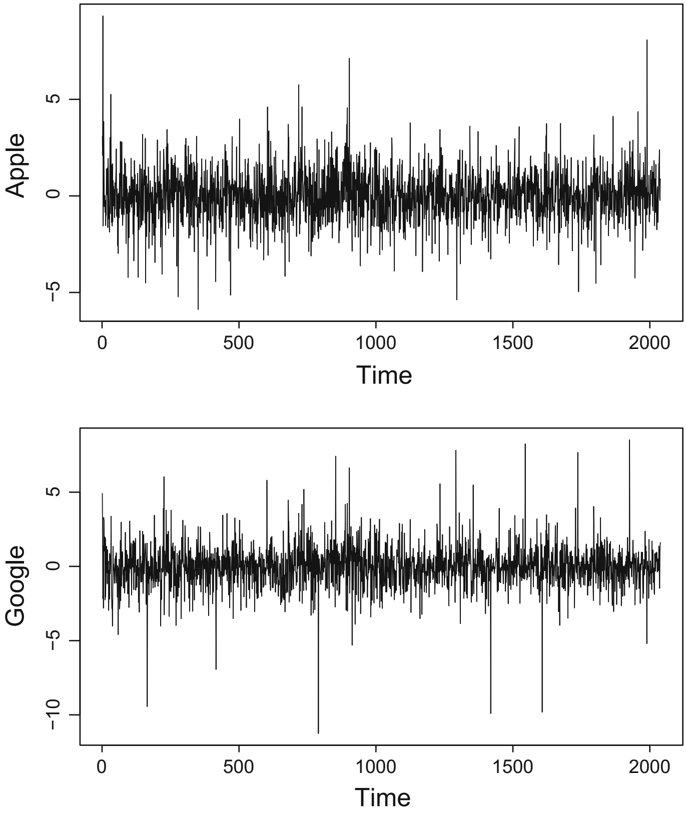

Fig. 6 Time series plots of the residuals of Apple (*top*) and Google (*bottom*)

Lemma 2 *Let* $\bar{F}_{\boldsymbol{\lambda}} = 1 - F_{\boldsymbol{\lambda}}$ *and* $Y_n(\boldsymbol{\lambda}, y) = (\log \boldsymbol{U}^{(\boldsymbol{\lambda})} - \log b_{\boldsymbol{\lambda}}(n/k) + \frac{1}{\alpha}\log y)_+$. *Then, for any* $y_0 > 0$, *as* $n \to \infty$,

$$\bar{F}_{\boldsymbol{\lambda}}(y_2^{-1/\alpha} b_{\boldsymbol{\lambda}}(n/k)) - \bar{F}_{\boldsymbol{\lambda}}(y_1^{-1/\alpha} b_{\boldsymbol{\lambda}}(n/k)) \sim \frac{k}{n}(y_2 - y_1), \tag{19}$$

$$\mathrm{E}\{Y_n(\boldsymbol{\lambda}, y_2) - Y_n(\boldsymbol{\lambda}, y_1)\}^2 \sim \frac{2k}{\alpha^2 n}\left(y_2 - y_1 - y_1 \log \frac{y_2}{y_1}\right), \tag{20}$$

$$\mathrm{E}\{Y_n(\boldsymbol{\lambda}, y_2) - Y_n(\boldsymbol{\lambda}, y_1)\}^3 \sim \frac{3k}{\alpha^3 n}\left(2y_2 - y_1 - y_1\left(\log \frac{y_2}{y_1} + 1\right)^2\right) \tag{21}$$

uniformly in $\boldsymbol{\lambda} \in \mathbb{S}^{d-1}$ *and* $0 \le y_1 < y_2 \le y_0$.

Note that (19) is a compact expression of

$$\lim_{n\to\infty} \sup_{\boldsymbol{\lambda}\in\mathbb{S}^{d-1}} \sup_{0\le y_1<y_2\le y_0} \left| \frac{n}{k} \frac{\bar{F}_{\boldsymbol{\lambda}}(y_2^{-1/\alpha} b_{\boldsymbol{\lambda}}(n/k)) - \bar{F}_{\boldsymbol{\lambda}}(y_1^{-1/\alpha} b_{\boldsymbol{\lambda}}(n/k))}{y_2 - y_1} - 1 \right| = 0.$$

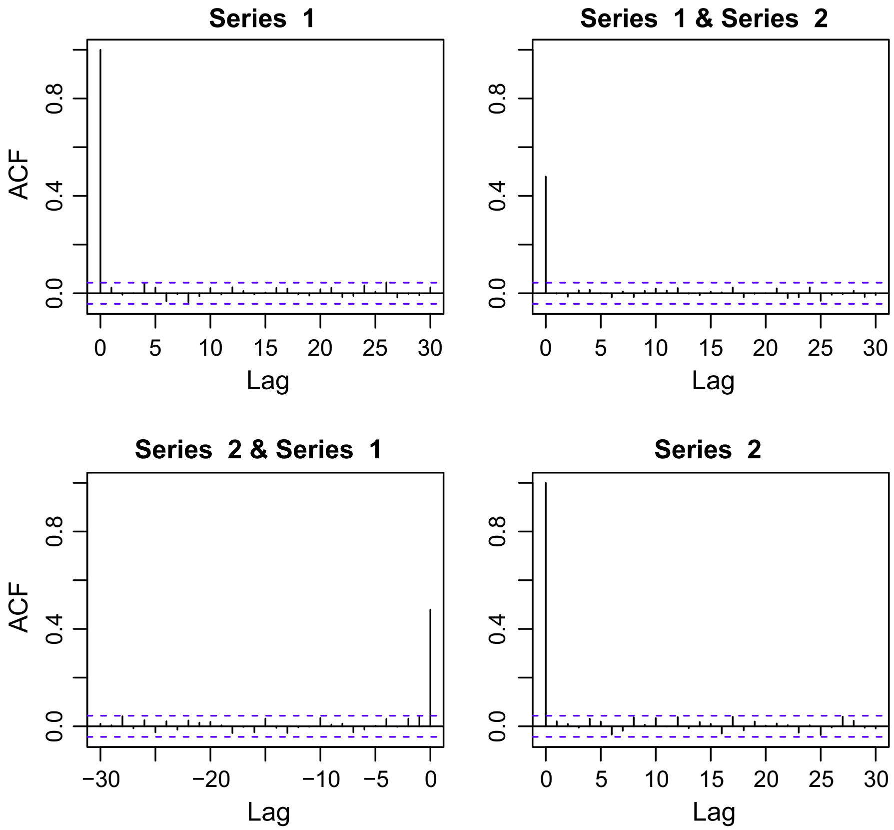

Fig. 7 The ACFs and CCFs of the residuals

Also, (20) and (21) can be rewritten similarly.

5.2 Weak convergence of the auxiliary processes

Define the $\ell^\infty(T)$-valued processes

$$L_n(\boldsymbol{\lambda}, y) := \frac{1}{\sqrt{k}} \sum_{i=1}^{n} \left\{ \left(\log \boldsymbol{U}_i^{(\boldsymbol{\lambda})} - \log y^{-\frac{1}{\alpha}} b_{\boldsymbol{\lambda}}(n/k) \right)_+ \right.$$
$$\left. - \mathrm{E} \left(\log \boldsymbol{U}_i^{(\boldsymbol{\lambda})} - \log y^{-\frac{1}{\alpha}} b_{\boldsymbol{\lambda}}(n/k) \right)_+ \right\},$$
$$M_n(\boldsymbol{\lambda}, y) := \frac{1}{\sqrt{k}} \sum_{i=1}^{n} \left\{ \mathrm{I} \left(\boldsymbol{U}_i^{(\boldsymbol{\lambda})} > y^{-\frac{1}{\alpha}} b_{\boldsymbol{\lambda}}(n/k) \right) - P \left(\boldsymbol{U}_i^{(\boldsymbol{\lambda})} > y^{-\frac{1}{\alpha}} b_{\boldsymbol{\lambda}}(n/k) \right) \right\},$$

where T is defined in (8). Below, we establish the weak convergence of these processes. To this end, we investigate the asymptotical tightness.

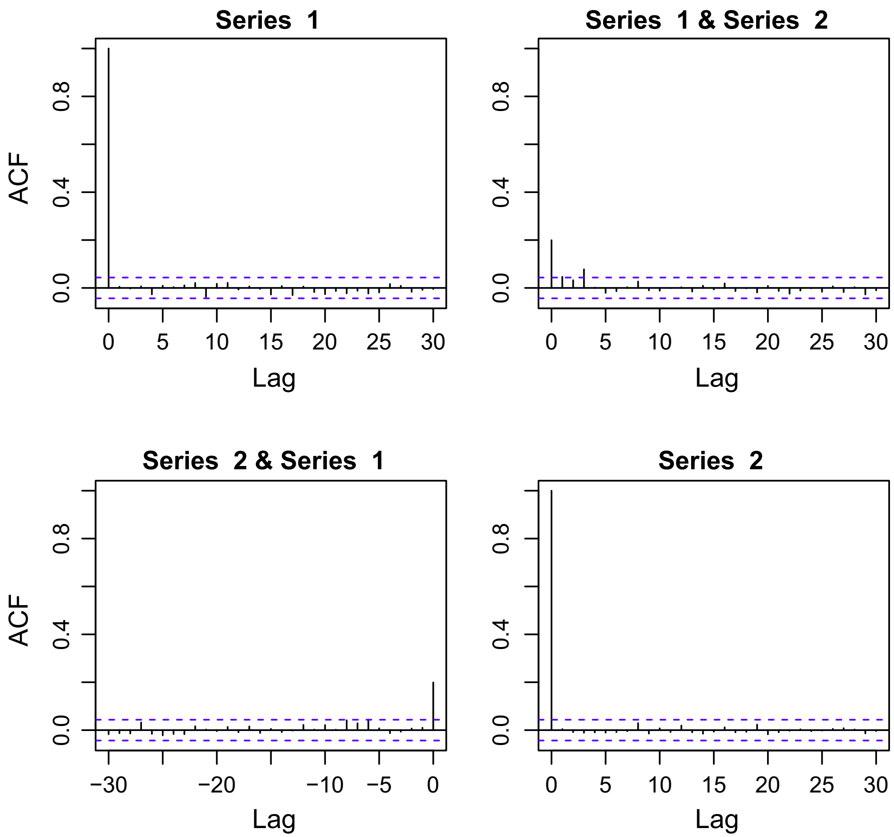

Fig. 8 The ACFs and CCFs for the squares of the residuals

Putting

$$f_n(\boldsymbol{u}; \boldsymbol{\lambda}, y) = \sqrt{\frac{n}{k}} \mathrm{I}\left(\boldsymbol{u}^{(\boldsymbol{\lambda})} > y^{-\frac{1}{\alpha}} b_{\boldsymbol{\lambda}}(n/k)\right),$$

$$g_n(\boldsymbol{u}; \boldsymbol{\lambda}, y) = \sqrt{\frac{n}{k}} \left(\log \boldsymbol{u}^{(\boldsymbol{\lambda})} - \log y^{-\frac{1}{\alpha}} b_{\boldsymbol{\lambda}}(n/k)\right)_+$$

where $\boldsymbol{u}^{(\boldsymbol{\lambda})} = \lambda_1 u_1 + \cdots + \lambda_d u_d$, we rewrite

$$L_n(\boldsymbol{\lambda}, y) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{g_n(\boldsymbol{U}_i; \boldsymbol{\lambda}, y) - \mathrm{E} g_n(\boldsymbol{U}_i; \boldsymbol{\lambda}, y)\},$$

$$M_n(\boldsymbol{\lambda}, y) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{f_n(\boldsymbol{U}_i; \boldsymbol{\lambda}, y) - \mathrm{E} f_n(\boldsymbol{U}_i; \boldsymbol{\lambda}, y)\}.$$

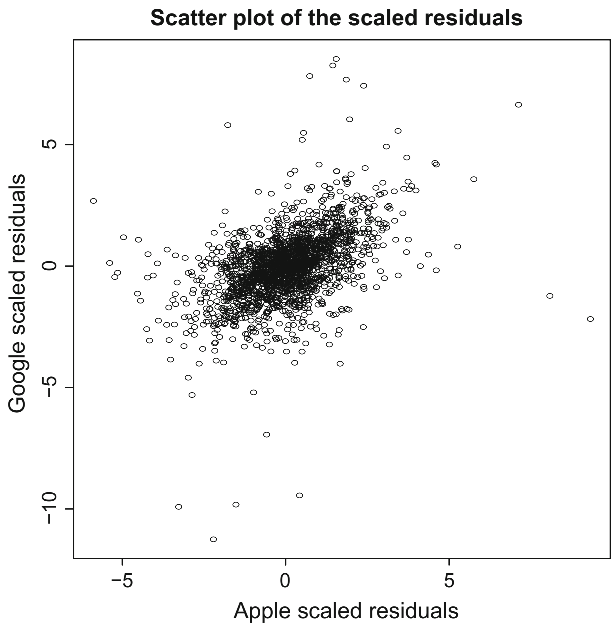

Fig. 9 The scatter plot of the residuals

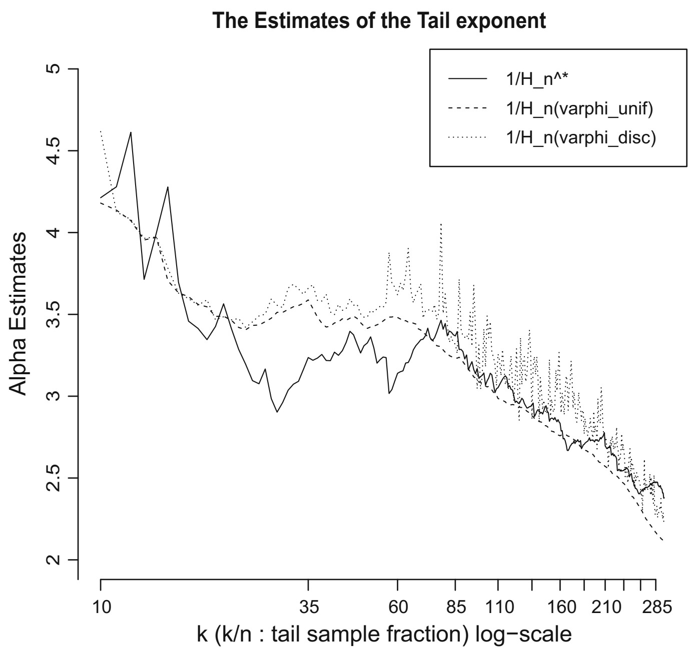

Fig. 10 The estimates of $1/\mathrm{H}_n^*$, $1/\mathrm{H}_n(\varphi_{\mathrm{unif}}, 1)$, and $1/\mathrm{H}_n(\hat{\varphi}_{\mathrm{disc}}, 1)$

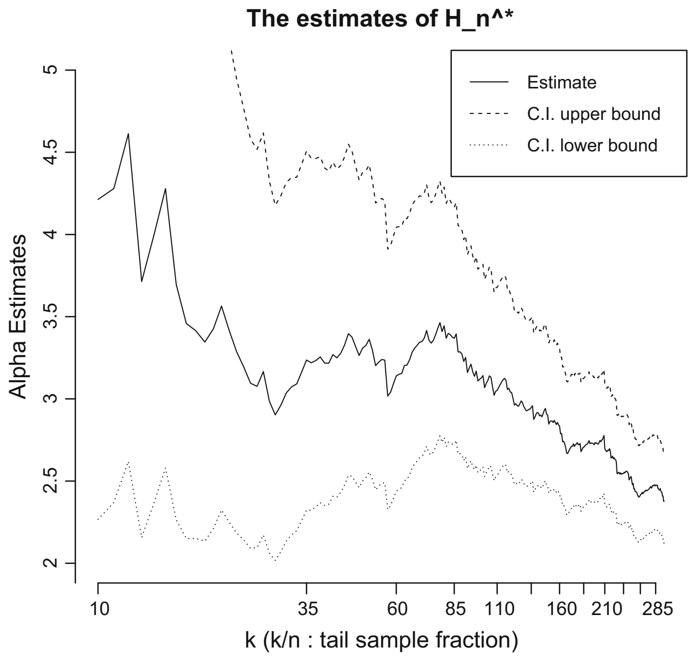

Fig. 11 The estimates of $1/\mathrm{H}_n^*$ and confidence intervals at level 95 %

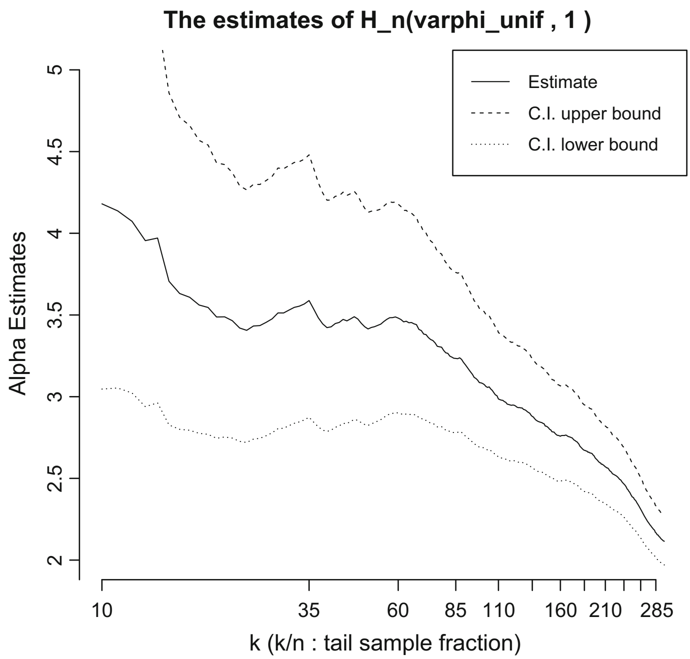

Fig. 12 The estimates of $1/\mathrm{H}_n(\varphi_{\mathrm{unif}}, 1)$ and confidence intervals at level 95 %

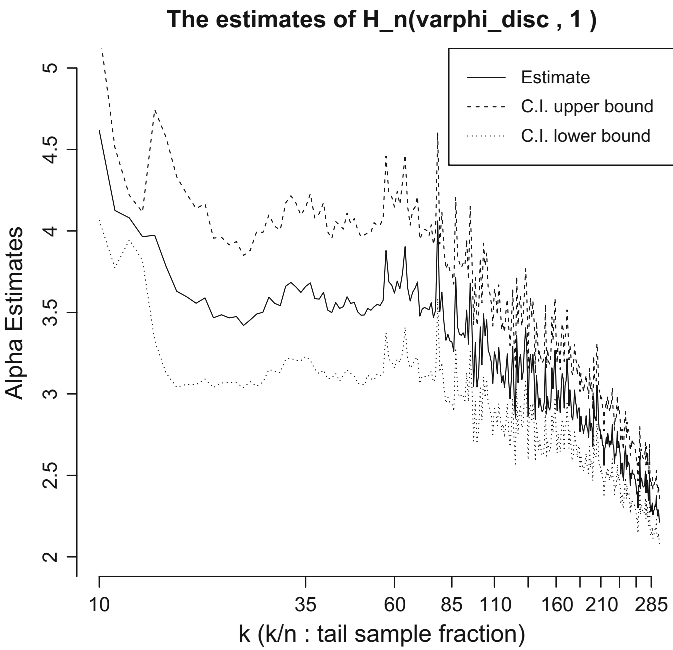

Fig. 13 The estimates of $1/\mathrm{H}_n(\hat{\varphi}_{\mathrm{disc}}, 1)$ and confidence intervals at level 95 %

Lemma 1 implies that there exist $0 < \underline{w} < \bar{w} < \infty$ and $n_0 \in \mathbb{N}$, such that

$$\underline{w} < \frac{y(n/k)}{\{b_{\boldsymbol{\lambda}}(n/k)\}^{\alpha}} < \bar{w} \quad \text{for all } y \in [\underline{y}, \bar{y}],\ \boldsymbol{\lambda} \in \mathbb{S}^{d-1} \text{ and } n \geq n_0. \tag{22}$$

We express

$$f_n^*(\boldsymbol{u}) = \sqrt{\frac{n}{k}}\mathrm{I}\left(|\boldsymbol{u}| > \left(\frac{n}{\bar{w}k}\right)^{\frac{1}{\alpha}}\right), \quad g_n^*(\boldsymbol{u}) = \sqrt{\frac{n}{k}}\left(\log|\boldsymbol{u}| - \log\left(\frac{n}{\bar{w}k}\right)^{\frac{1}{\alpha}}\right)_+,$$

and for $a > 0$,

$$\begin{aligned}
s_n(\boldsymbol{u}; \boldsymbol{\lambda}, w, a) &= \sqrt{\frac{n}{k}}\left\{\mathrm{I}\left(|\boldsymbol{u}| > a\right) + \mathrm{I}\left(|\boldsymbol{u}| \leq a,\ \boldsymbol{u}^{(\boldsymbol{\lambda})} > \left(\frac{n}{wk}\right)^{\frac{1}{\alpha}}\right)\right\},\\
l_n(\boldsymbol{u}; \boldsymbol{\lambda}, w, a) &= \sqrt{\frac{n}{k}}\mathrm{I}\left(|\boldsymbol{u}| \leq a,\ \boldsymbol{u}^{(\boldsymbol{\lambda})} > \left(\frac{n}{wk}\right)^{\frac{1}{\alpha}}\right),\\
s_n'(\boldsymbol{u}; \boldsymbol{\lambda}, w, a) &= g_n^*(\boldsymbol{u})\mathrm{I}\left(|\boldsymbol{u}| > a\right) + \sqrt{\frac{n}{k}}\mathrm{I}(|\boldsymbol{u}| \leq a)\left(\log \boldsymbol{u}^{(\boldsymbol{\lambda})} - \log\left(\frac{n}{wk}\right)^{\frac{1}{\alpha}}\right)_+,\\
l_n'(\boldsymbol{u}; \boldsymbol{\lambda}, w, a) &= \sqrt{\frac{n}{k}}\mathrm{I}(|\boldsymbol{u}| \leq a)\left(\log \boldsymbol{u}^{(\boldsymbol{\lambda})} - \log\left(\frac{n}{wk}\right)^{\frac{1}{\alpha}}\right)_+.
\end{aligned}$$

We provide several lemmas for proving the asymptotical tightness of $L_n(\boldsymbol{\lambda}, y)$ and $M_n(\boldsymbol{\lambda}, y)$. The following is used to verify Proposition 1.

Lemma 3 *For any* $\boldsymbol{\lambda} \in \mathbb{S}^{d-1}$ *and* $y \in [\underline{y}, \bar{y}]$,

$$f_n(\boldsymbol{u}; \boldsymbol{\lambda}, y) \le f_n^*(\boldsymbol{u}), \quad g_n(\boldsymbol{u}; \boldsymbol{\lambda}, y) \le g_n^*(\boldsymbol{u}) \qquad \textit{for all } \boldsymbol{u},$$

and $\lim_{n\to\infty} \| f_n^*(\boldsymbol{U})\|_2$ *and* $\lim_{n\to\infty} \| g_n^*(\boldsymbol{U})\|_2$ *exist and are finite and strictly positive. Further, for any* $\eta > 0$, *we have*

$$\lim_{n\to\infty} \mathrm{E}\{f_n^*(\boldsymbol{U})\}^2 \mathrm{I}(f_n^*(\boldsymbol{U}) > \eta\sqrt{n}) = 0 \quad \textit{and} \quad \lim_{n\to\infty} \mathrm{E}\{g_n^*(\boldsymbol{U})\}^2 \mathrm{I}(g_n^*(\boldsymbol{U}) > \eta\sqrt{n}) = 0.$$

Let $\epsilon > 0$ and $b_{|\boldsymbol{U}|}(x) = \inf\left\{z : P(|\boldsymbol{U}| > z) \le x^{-1}\right\}$. Note that for $\epsilon > 0$,

$$\frac{n}{k} P\left(|\boldsymbol{U}| > b_{|\boldsymbol{U}|}\left(\frac{2n}{\epsilon^2 k}\right)\right) \le \frac{\epsilon^2}{2}.$$

Also, we define

$$b(n, \epsilon) := b_{|\boldsymbol{U}|}\left(\frac{2n}{\epsilon^2 k}\right),$$

and, similarly, $b'(n, \epsilon)$ as the minimum constant a satisfying

$$\mathrm{E}\left\{g_n^*(\boldsymbol{U})\right\}^2 \mathrm{I}(|\boldsymbol{U}| > a) \le \frac{\epsilon^2}{2}.$$

Then, we get the following two lemmas, which will be used to verify Lemma 6.

Lemma 4 *For any* $w_0 > 0$, *there exist* $n_1 \in \mathbb{N}$ *and* $K_1 > 0$ *such that for* $\epsilon > 0$, $n \ge n_1$, $\boldsymbol{\lambda} \in \mathbb{S}^{d-1}$, *and* $0 < w_1 < w_2 < w_0$,

$$\|s_n(\boldsymbol{U}; \boldsymbol{\lambda}, w_2, b(n,\epsilon)) - l_n(\boldsymbol{U}; \boldsymbol{\lambda}, w_1, b(n,\epsilon))\|_2^2 \le K_1(w_2 - w_1) + \frac{\epsilon^2}{2},$$
$$\|s_n'(\boldsymbol{U}; \boldsymbol{\lambda}, w_2, b'(n,\epsilon)) - l_n'(\boldsymbol{U}; \boldsymbol{\lambda}, w_1, b'(n,\epsilon))\|_2^2 \le K_1(w_2 - w_1) + \frac{\epsilon^2}{2}.$$

Lemma 5 *Let* $y, w_1, w_2 > 0$, $\boldsymbol{\lambda}, \boldsymbol{\lambda}^* \in \mathbb{S}^{d-1}$ *and* $A > 0$, *such that*

$$\Delta_1 = w_1^{-\frac{1}{\alpha}} (n/k)^{\frac{1}{\alpha}} - y^{-\frac{1}{\alpha}} b_{\boldsymbol{\lambda}}(n/k) > 0, \quad \Delta_2 = y^{-\frac{1}{\alpha}} b_{\boldsymbol{\lambda}}(n/k) - w_2^{-\frac{1}{\alpha}} (n/k)^{\frac{1}{\alpha}} > 0,$$
$$|\boldsymbol{\lambda} - \boldsymbol{\lambda}^*| \le \frac{\Delta_1 \wedge \Delta_2}{A}.$$

Then, for each $\boldsymbol{u}$, *we have*

$$l_n(\boldsymbol{u}; \boldsymbol{\lambda}^*, w_1, A) \le f_n(\boldsymbol{u}; \boldsymbol{\lambda}, y) \le s_n(\boldsymbol{u}; \boldsymbol{\lambda}^*, w_2, A),$$
$$l_n'(\boldsymbol{u}; \boldsymbol{\lambda}^*, w_1, A) \le g_n(\boldsymbol{u}; \boldsymbol{\lambda}, y) \le s_n'(\boldsymbol{u}; \boldsymbol{\lambda}^*, w_2, A).$$

Define

$$[h_1, h_2]^f := \{(\boldsymbol{\lambda}, y) : h_1(\boldsymbol{u}) \le f_n(\boldsymbol{u}; \boldsymbol{\lambda}, y) \le h_2(\boldsymbol{u}) \text{ for each } \boldsymbol{u}\}$$

which is an ϵ-bracket when $\|h_2(\boldsymbol{U}) - h_1(\boldsymbol{U})\|_2 \le \epsilon$ is satisfied. Then, $N_{[]}^f(\epsilon\,; n)$ is defined as the minimum number of ϵ-brackets needed for covering T. $N_{[]}^g(\epsilon\,; n)$ is analogously defined.

The following two lemmas are needed to verify Proposition 1.

Lemma 6 *It holds that*

$$\lim_{\delta\to\infty} \limsup_{n\to\infty} \int_0^\delta \sqrt{\log N_{[]}^f(\epsilon\,; n)}\mathrm{d}\epsilon = 0 \tag{23}$$

and

$$\lim_{\delta\to\infty} \limsup_{n\to\infty} \int_0^\delta \sqrt{\log N_{[]}^g(\epsilon\,; n)}\mathrm{d}\epsilon = 0. \tag{24}$$

Lemma 7 *It holds that*

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \sup\{\|f_n(\boldsymbol{U}; \boldsymbol{\lambda}_1, y_1) - f_n(\boldsymbol{U}; \boldsymbol{\lambda}_2, y_2)\|_2 : |(\boldsymbol{\lambda}_1, y_1) - (\boldsymbol{\lambda}_2, y_2)| < \delta\} = 0 \tag{25}$$

and

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \sup\{\|g_n(\boldsymbol{U}; \boldsymbol{\lambda}_1, y_1) - g_n(\boldsymbol{U}; \boldsymbol{\lambda}_2, y_2)\|_2 : |(\boldsymbol{\lambda}_1, y_1) - (\boldsymbol{\lambda}_2, y_2)| < \delta\} = 0. \tag{26}$$

Proposition 1 (i) $\{M_n(\boldsymbol{\lambda}, y)\}$ *and* $\{L_n(\boldsymbol{\lambda}, y)\}$ *are asymptotically tight in* $\ell^\infty(T)$ *and thus converge weakly to the Gaussian processes with zero mean function and covariance functions* c_1 *and* c_2/α^2, *respectively.*

(ii) *Moreover, the weak limits have continuous versions.*

Proof From Lemmas 3, 6 and 7, we can see that all the conditions in Theorem 2.11.23 of van der Vaart and Wellner (1996) are fulfilled. Hence, (i) follows from the fact that

$$\lim_{n\to\infty} \mathrm{E}\{M_n(\boldsymbol{\lambda}_1, y_1)M_n(\boldsymbol{\lambda}_2, y_2)\} = \mathrm{c}_1((\boldsymbol{\lambda}_1, y_1), (\boldsymbol{\lambda}_2, y_2))$$

and

$$\lim_{n\to\infty} \mathrm{E}\{L_n(\boldsymbol{\lambda}_1, y_1)L_n(\boldsymbol{\lambda}_2, y_2)\} = \frac{1}{\alpha^2}\mathrm{c}_2((\boldsymbol{\lambda}_1, y_1), (\boldsymbol{\lambda}_2, y_2)).$$

To prove (ii), we retrieve the proof of Lemma 7. Suppose that $\boldsymbol{\lambda}_1 \neq \boldsymbol{\lambda}_2$. Take $\epsilon = (6K_0^3|\boldsymbol{\lambda}_1 - \boldsymbol{\lambda}_2|)^{\frac{1}{2+2/\alpha}}$. Then, there exists $K_1 > 0$, not depending on $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, y_1, y_2$, such that

$$\begin{aligned}\limsup_{n\to\infty} \| f_n(\boldsymbol{U}; \boldsymbol{\lambda}_1, y_1) - f_n(\boldsymbol{U}; \boldsymbol{\lambda}_2, y_2)\|_2^2 &\leq \epsilon^2 + \left\{\frac{5\epsilon^2}{6} + K_0 \left|\frac{y_1}{C_{\boldsymbol{\lambda}_1}} - \frac{y_2}{C_{\boldsymbol{\lambda}_2}}\right|\right\} \\ &\leq K_1 \left(|\boldsymbol{\lambda}_1 - \boldsymbol{\lambda}_2|^{\frac{2}{2+2/\alpha}} + |y_1 - y_2| + |C_{\boldsymbol{\lambda}_1} - C_{\boldsymbol{\lambda}_2}|\right).\end{aligned}$$

Thus, letting M_∞ denote the weak limit of $\{M_n\}$, we get

$$\|M_\infty(\boldsymbol{\lambda}_1, y_1) - M_\infty(\boldsymbol{\lambda}_2, y_2)\|_2^2 \leq K_1 \left(|\boldsymbol{\lambda}_1 - \boldsymbol{\lambda}_2|^{\frac{2}{2+2/\alpha}} + |y_1 - y_2| + |C_{\boldsymbol{\lambda}_1} - C_{\boldsymbol{\lambda}_2}|\right).$$

Since M_∞ is Gaussian, there exists $\nu > 0$ such that

$$\mathrm{E}\,|M_\infty(\boldsymbol{\lambda}_1, y_1) - M_\infty(\boldsymbol{\lambda}_2, y_2)|^\nu \leq K\,|(\boldsymbol{\lambda}_1, y_1) - (\boldsymbol{\lambda}_2, y_2)|^{d+1},$$

where K does not depend on $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, y_1, y_2$. If $\boldsymbol{\lambda}_1 = \boldsymbol{\lambda}_2$, we directly obtain the above inequality from (19). Hence, M_∞ has a continuous version (cf. Problem 2.2.9 of Karatzas and Shreve 1991). We can also prove the same for $\{L_n\}$ in a similar fashion. This completes the proof. □

5.3 The proof of Theorem 1

Define

$$W_n(\boldsymbol{\lambda}, y) = \sqrt{k}\left\{\log \boldsymbol{U}^{(\boldsymbol{\lambda})}_{(yk+1)} - \log b_{\boldsymbol{\lambda}}\left(\frac{n}{yk}\right)\right\}.$$

Below, we provide a lemma to show the connection between $W_n(\boldsymbol{\lambda}, y)$ and $M_n(\boldsymbol{\lambda}, y)$, which is useful to prove Propositions 2 and 3. Proposition 1 plays a crucial role to prove this lemma.

Lemma 8 *For any $K > 0$, we have that $W_n(\boldsymbol{\lambda}, y) > \zeta$ if and only if $M_n(\boldsymbol{\lambda}, y) > \alpha y \zeta + o_P(1)$ as $n \to \infty$, where the $o_P(1)$ terms are negligible uniformly in $(\boldsymbol{\lambda}, y) \in T$ and $\zeta \in [-K, K]$.*

Proposition 2 (i) *$\{W_n(\boldsymbol{\lambda}, y) : (\boldsymbol{\lambda}, y) \in T\}$ is asymptotically uniformly equicontinuous.*
(ii) *Moreover, the weak limit of $\{yW_n(\boldsymbol{\lambda}, y)\}$ is identical to that of $\{\alpha^{-1}M_n(\boldsymbol{\lambda}, y)\}$.*

Proof Since $\{M_n\}$ is asymptotically tight, owing to Lemma 8,

$$\sup_{(\boldsymbol{\lambda}, y)\in T} |yW_n(\boldsymbol{\lambda}, y)| = O_P(1). \tag{27}$$

Let $\epsilon > 0$ and $K > 0$ and take

$$-K = \zeta_0 < \zeta_1 < \cdots < \zeta_m = K$$

with $\epsilon/4 < \zeta_i - \zeta_{i-1} < \epsilon/2$ for $i = 1, \ldots, m$. Then, if we put

$$A_n(K) = \left\{ \sup_{(\boldsymbol{\lambda}, y) \in T} |y W_n(\boldsymbol{\lambda}, y)| < K \right\},$$
$$B_n(\delta) = \left\{ \sup_{\delta} |y_1 W_n(\boldsymbol{\lambda}_1, y_1) - y_2 W_n(\boldsymbol{\lambda}_2, y_2)| > \epsilon \right\}, \quad \delta > 0,$$

where $\sup_\delta$ is taken over $|(\boldsymbol{\lambda}_1, y_1) - (\boldsymbol{\lambda}_2, y_2)| < \delta$, $A_n(K) \cap B_n(\delta)$ implies

$$\bigcup_{i=1}^{m} \bigcup_{|(\boldsymbol{\lambda}_1, y_1) - (\boldsymbol{\lambda}_2, y_2)| < \delta} \{y_1 W_n(\boldsymbol{\lambda}_1, y_1) < \zeta_{i-1},\ y_2 W_n(\boldsymbol{\lambda}_2, y_2) \geq \zeta_i\}$$
$$= \bigcup_{i=1}^{m} \bigcup_{|(\boldsymbol{\lambda}_1, y_1) - (\boldsymbol{\lambda}_2, y_2)| < \delta} \{M_n\,(\boldsymbol{\lambda}_1, y_1) < \alpha\zeta_{i-1} + o_P(1),\ M_n\,(\boldsymbol{\lambda}_2, y_2)$$
$$\geq \alpha\zeta_i + o_P(1)\}$$
$$\subset \bigcup_{|(\boldsymbol{\lambda}_1, y_1) - (\boldsymbol{\lambda}_2, y_2)| < \delta} \left\{ |M_n\,(\boldsymbol{\lambda}_1, y_1) - M_n\,(\boldsymbol{\lambda}_2, y_2)| \geq \frac{\alpha\epsilon}{4} + o_P(1) \right\},$$

where the $o_P(1)$ terms are negligible uniformly in $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, y_1, y_2$. Thus, owing to the asymptotical uniformly equicontinuity of $\{M_n\}$, we get

$$\lim_{\delta \to 0} \limsup_{n \to \infty} P(B_n(\delta)) \leq \lim_{\delta \to 0} \limsup_{n \to \infty} P(B_n(\delta) \cap A_n(K)) + \limsup_{n \to \infty} \{1 - P(A_n(K))\}$$
$$= \limsup_{n \to \infty} \{1 - P(A_n(K))\}.$$

Then, by letting $K \to \infty$, we can see that (i) holds owing to (27). Since (ii) can be directly derived from Lemma 8, the proposition is established. □

The following is useful to prove Proposition 3.

Lemma 9 *For any* $K > 0$,

$$\frac{n}{yk} \mathrm{E} \left(\log \boldsymbol{U}^{(\boldsymbol{\lambda})} - \log b_{\boldsymbol{\lambda}} \left(\frac{n}{yk} \right) - \frac{\zeta}{\alpha\sqrt{k}} \right)_+ = \frac{1}{\alpha} - \frac{\zeta}{\alpha\sqrt{k}} + \frac{\gamma_{\boldsymbol{\lambda}} D_{\boldsymbol{\lambda}} M_{\boldsymbol{\lambda}}}{\sqrt{k}\alpha(\alpha - \gamma_{\boldsymbol{\lambda}})} y^{-\gamma_{\boldsymbol{\lambda}}/\alpha}$$
$$+ o\left(\frac{1}{\sqrt{k}} \right)$$

uniformly in $\zeta \in [-K, K]$ *and* $(\boldsymbol{\lambda}, y) \in T$.

Now, we are ready to provide the asymptotic property of $\mathrm{H}_n(\boldsymbol{\lambda}, y)$.

Proposition 3 *The process*

$$(\boldsymbol{\lambda}, y) \mapsto y\sqrt{k}\left\{\mathrm{H}_n(\boldsymbol{\lambda}, y) - \frac{1}{\alpha} - \frac{\gamma_{\boldsymbol{\lambda}} D_{\boldsymbol{\lambda}} M_{\boldsymbol{\lambda}}}{\sqrt{k}\alpha(\alpha - \gamma_{\boldsymbol{\lambda}})} y^{-\gamma_{\boldsymbol{\lambda}}/\alpha}\right\}$$

converges weakly to a Gaussian process with zero mean and covariance function V (*presented in* (10)) *in* $\ell^\infty(T)$.

Proof According to Lemma 9, for any $K > 0$,

$$\begin{aligned} L_n\left(\boldsymbol{\lambda}, ye^{-\zeta/\sqrt{k}}\right) = y\sqrt{k}\left\{\frac{1}{yk}\sum_{i=1}^{n}\left(\log \boldsymbol{U}_i^{(\boldsymbol{\lambda})} - \log b_{\boldsymbol{\lambda}}\left(\frac{n}{yk}\right) - \frac{\zeta}{\alpha\sqrt{k}}\right)_+ \right.\\ \left. - \frac{1}{\alpha} + \frac{\zeta}{\alpha\sqrt{k}} - \frac{\gamma D_{\boldsymbol{\lambda}} M_{\boldsymbol{\lambda}}}{\sqrt{k}\alpha(\alpha - \gamma_{\boldsymbol{\lambda}})} y^{-\gamma_{\boldsymbol{\lambda}}/\alpha}\right\} \\ + o_P(1) \end{aligned}$$

uniformly in $\zeta \in [-K, K]$ and $(\boldsymbol{\lambda}, y) \in T$. Since $\sup_{(\boldsymbol{\lambda}, y)\in T} |W_n(\boldsymbol{\lambda}, y)| = O_P(1)$ (cf. Proposition 2), it holds that

$$\begin{aligned} L_n\left(\boldsymbol{\lambda}, ye^{-\alpha W_n(\boldsymbol{\lambda}, y)/\sqrt{k}}\right) = y\sqrt{k}\left\{\mathrm{H}_n(\boldsymbol{\lambda}, y) - \frac{1}{\alpha} - \frac{\gamma_{\boldsymbol{\lambda}} D_{\boldsymbol{\lambda}} M_{\boldsymbol{\lambda}}}{\sqrt{k}\alpha(\alpha - \gamma_{\boldsymbol{\lambda}})} y^{-\gamma_{\boldsymbol{\lambda}}/\alpha}\right\} \\ + yW_n(\boldsymbol{\lambda}, y) + o_P(1) \end{aligned}$$

uniformly in $(\boldsymbol{\lambda}, y) \in T$. Moreover, since $\{L_n\}$ is asymptotically uniformly equicontinuous, we get

$$y\sqrt{k}\left\{\mathrm{H}_n(\boldsymbol{\lambda}, y) - \frac{1}{\alpha} - \frac{\gamma_{\boldsymbol{\lambda}} D_{\boldsymbol{\lambda}} M_{\boldsymbol{\lambda}}}{\sqrt{k}\alpha(\alpha - \gamma_{\boldsymbol{\lambda}})} y^{-\gamma_{\boldsymbol{\lambda}}/\alpha}\right\} = L_n(\boldsymbol{\lambda}, y) - yW_n(\boldsymbol{\lambda}, y) + o_P(1) \tag{28}$$

uniformly in $(\boldsymbol{\lambda}, y) \in T$. Now that $\{L_n(\boldsymbol{\lambda}, y) - yW_n(\boldsymbol{\lambda}, y)\}$ converges weakly to the weak limit of $\{L_n(\boldsymbol{\lambda}, y) - \alpha^{-1}M_n(\boldsymbol{\lambda}, y)\}$ due to Lemma 8, the proposition is established. □

Proof of Theorem 1 The theorem is readily established by Proposition 3 and the mapping theorem in Theorem 1.3.6 of van der Vaart and Wellner (1996). □

Acknowledgements We thank a referee for her/his valuable comments. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2015R1A2A2A01003894) (S. Lee) and Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (Grant Number: NRF-2012R1A6A3A01039754) (M. Kim).

References

Berkes, I., Horváth, L. (2004). The efficiency of the estimators of the parameters in GARCH processes. *Annals of Statistics*, *32*, 633–655.

Danielsson, J., de Haan, L., Peng, L., de Vries, C. G. (2001). Using a bootstrap method to choose the sample fraction in tail index estimation. *Journal of Multivariate Analysis*, *76*, 226–248.

Dematteo, A., Clémençon, S. (2016). On tail index estimation based on multivariate data. *Journal of Nonparametric Statistics*, *28*, 152–176.

Drees, H., de Haan, L., Resnick, S. (2000). How to make a Hill plot. *Annals of Statistics*, *28*, 254–274.

Feuerverger, A., Hall, P. (1999). Estimating a tail exponent by modelling departure from a Pareto distribution. *Annals of Statistics*, *27*, 760–781.

Goldie, C. M., Smith, R. L. (1987). Slow variation with remainder: theory and applications. *The Quarterly Journal of Mathematics*, *38*, 45–71.

Gomes, M. I., de Haan, L., Rodrigues, L. H. (2008). Tail index estimation for heavy-tailed models: Accommodation of bias in weighted log-excesses. *Journal of Royal Statistical Society Series B*, *70*, 31–52.

Hall, P. (1982). On some simple estimates of an exponent of regular variation. *Journal of Royal Statistical Society Series B*, *44*, 37–42.

Hall, P., Welsh, A. H. (1985). Adaptive estimates of parameters of regular variation. *Annals of Statistics*, *13*, 331–341.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, *3*, 1163–1174.

Kallenberg, O. (1983). *Random measures* (3rd ed.). Berlin: Akademie.

Karatzas, I., Shreve, S. E. (1991). *Brownian motion and stochastic calculus* (2nd ed.). New York: Springer.

Mainik, G., Rüschendorf, L. (2010). On optimal portfolio diversification with respect to extreme risks. *Finance Stochastics*, *14*, 593–623.

Resnick, S. I. (1987). *Extreme values, regular variation and point processes*. New York: Springer.

van der Vaart, A. W., Wellner, J. A. (1996). *Weak convergence and empirical processes with applications to statistics*. New York: Springer.