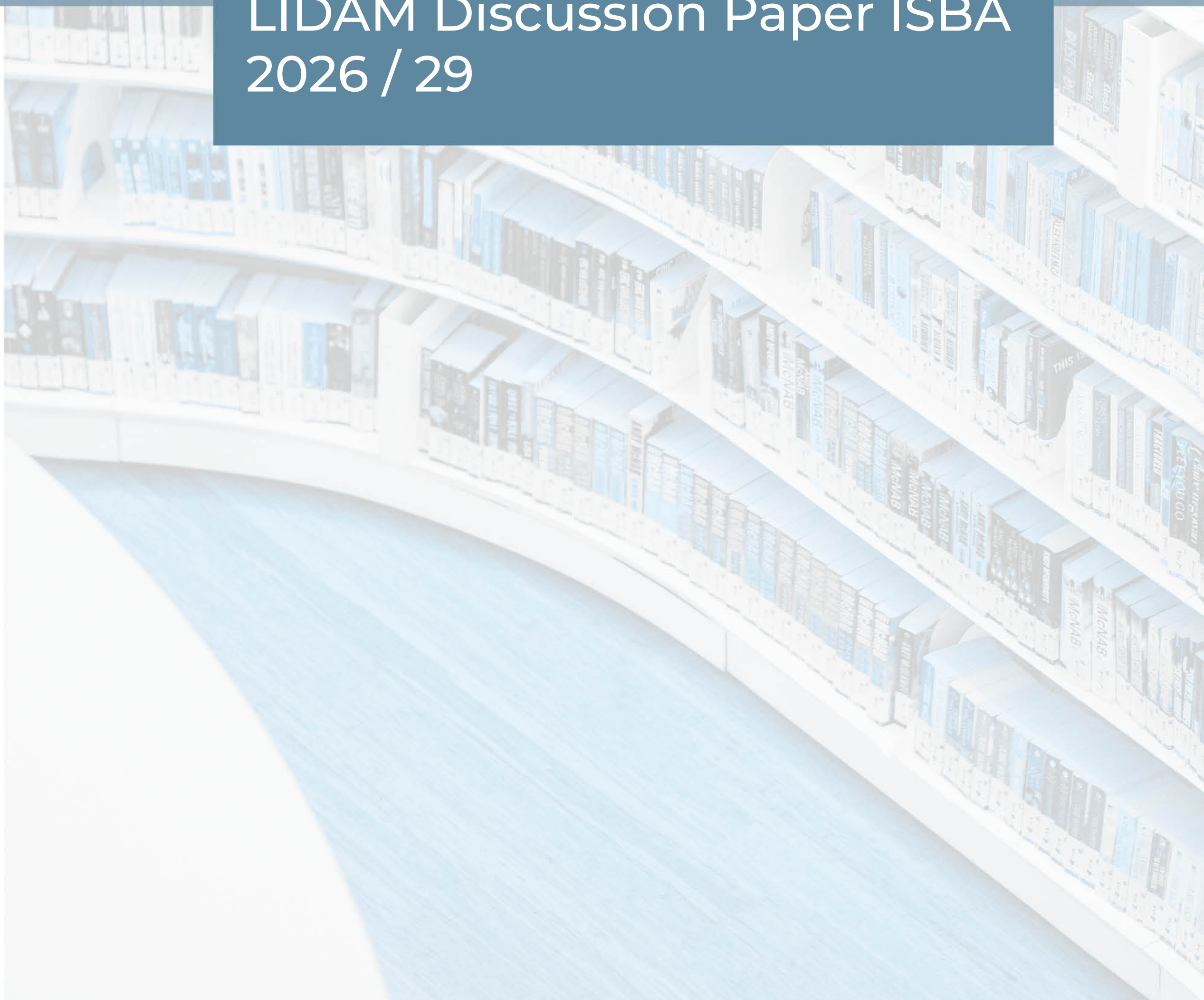
ESTIMATING THE HÜSLER–REISS VARIOGRAM MATRIX BY CLIPPED MOMENTS

Shuang Hu, Johan Segers

LIDAM Discussion Paper ISBA
2026 / 29



UCLouvain

LiDAM

ISBA | Louvain Institute of Data Analysis and Modeling in economics and statistics

ISBA
Voie du Roman Pays 20 - L1.04.01
B-1348 Louvain-la-Neuve
Email : lidam-library@uclouvain.be
https://uclouvain.be/en/research-institutes/lidam/isba/publication

Estimating the Hüsler–Reiss variogram matrix by clipped moments

Shuang Hu*[1] and Johan Segers†[2,3]

[1]School of Mathematics and Statistics/Key Laboratory of Intelligent Analysis and Decision on Complex Systems, Chongqing University of Posts and Telecommunications, 400715 Chongqing, China
[2]Department of Mathematics, KU Leuven, B3001 Heverlee, Belgium
[3]LIDAM/ISBA, UCLouvain, B1348 Louvain-la-Neuve, Belgium

July 23, 2026

Abstract

In multivariate extreme value analysis, the tail dependence between some of the risk variables at hand may be weak, even when other variables do tend to become large simultaneously. Weak tail dependence may induce a substantial bias in estimation procedures based on the limiting multivariate (generalized) Pareto distribution of excesses over high thresholds. We consider a Hüsler–Reiss multivariate generalized Pareto model and, motivated by this issue, propose first- and second-order moment estimators of its variogram matrix constructed from a lower-tail-clipped version of the underlying random vector. The asymptotic normality of the proposed estimators is established. We demonstrate by simulation studies that they have lower bias than the empirical variogram estimator in certain cases, particularly when the dependence between components is weak. The estimators are applied to flood discharge data from the Danube river basin and the US flight delay data, showing that the tail dependence structure implied by the fitted model based on the first-order clipped moment estimator aligns more closely with the empirical tail dependence of the data than that based on the empirical variogram estimator.

Keywords. Hüsler–Reiss distribution; Variogram matrix; Moment estimator

MSC 2020 Subject Classification. 60G70; 62F10

1 Introduction

In many areas such as finance, environmental science, and engineering, extreme events can have disproportionately large impacts. Accurate modeling of such rare occurrences is

*Corresponding author: hushuang@cqupt.edu.cn
†E-mail: jjjsegers@kuleuven.be

essential for effective risk assessment and decision making. While univariate extreme value theory has been well-developed, practical problems often involve multiple interconnected variables, necessitating a multivariate framework that can capture both marginal extremes and their joint dependence structure.

Multivariate extreme value theory addresses this need by characterizing the limiting distribution of componentwise maxima through multivariate max-stable distributions or multivariate generalized Pareto (MGP) distributions (MGPD). To enable flexible and interpretable modeling of extremal dependence, various parametric families have been introduced in the literature (see Beirlant et al., 2004, Chapter 9; Davison et al., 2012; Hanson et al., 2017; and references therein). Among these models, the Hüsler–Reiss distribution parameterized by the so-called variogram matrix stands out for its analytical tractability and practical relevance. It arises as the limiting distribution of componentwise maxima in Gaussian triangular arrays (Hüsler and Reiss, 1989), and also appears as the finite-dimensional marginal distribution of the Brown–Resnick process (Engelke et al., 2015).

Owing to its modeling flexibility, the Hüsler–Reiss distribution has garnered significant attention in the literature. Its theoretical aspects, such as the tail dependence properties and convergence results where it appears as a limit, have been extensively studied; see, for example, Frick and Reiss (2010), Liao and Peng (2014), and Hashorva et al. (2016). Beyond these theoretical developments, the widespread use of this model in practice highlights the importance of reliable statistical inference. Various estimation methods developed for multivariate max-stable distributions or multivariate (generalized) Pareto distributions can be adapted to the estimation of the variogram matrix, such as the composite likelihood estimation method in Padoan et al. (2010), the M-estimator in Einmahl et al. (2012) and the weighted least squares estimator in Einmahl et al. (2018). Specifically for the Hüsler–Reiss family, Huser and Davison (2013) investigated the composite likelihood estimation approach in the context of the Brown–Resnick process, while Engelke et al. (2015) proposed several new estimators for the Hüsler–Reiss distribution based on the peaks-over-threshold approach. Of particular relevance to our work, Engelke and Volgushev (2022) proposed an empirical estimator of the variogram matrix for a general multivariate Pareto distribution, which can be directly applied to the Hüsler–Reiss model as a special case. Due to its simplicity and broad applicability, this estimator has been widely used in applications.

In practice, extremes often involve only a subset of variables becoming large while others remain moderate. Such configurations, referred to as extreme clusters (Chiapino et al., 2020) or extreme directions (Mourahib et al., 2025), motivate the need for flexible models that can capture both strong and weak tail dependence across different subsets of components. The Hüsler–Reiss model is a natural choice in this context, as its variogram parameter matrix provides a parsimonious description of the entire dependence structure, including partial extremal behavior. However, estimation procedures based on empirical variograms, the widely adopted method in practice, can be severely biased when the dependence between certain components is weak. This bias arises because weak dependence increases the likelihood that some components take relatively small values, even when others are extreme. As a result, observations exceeding the threshold in only a subset of components often include non-extreme values in the remaining components, which weakens the estimation of the dependence structure implied by the limiting MGPD. Consequently, empirical variogram estimators, which rely on the assumption that the asymptotic model provides a good approximation in the tail region, can yield inaccurate estimates of the elements of the variogram matrix.

This observation motivates us to employ a clipping method in the sense that the standardized MGPD observations $\boldsymbol{X}$ below a threshold c are modified through a deterministic transformation $\boldsymbol{Y} = \boldsymbol{X} \vee c = (X_i \vee c, i \in V)$, to mitigate the bias of the empirical variogram estimator introduced by non-extreme components. Based on this clipping, we propose two moment-type estimators for the elements of the variogram matrix and establish the asymptotic normality of these estimators. Simulation studies and empirical analyses indicate that the proposed estimators significantly mitigate bias in the Hüsler–Reiss model when the dependence among certain components is weak. Furthermore, we apply the Hüsler–Reiss model to extreme flood discharges from the Danube River basin and the US flight delay data, and estimate the parameters using the proposed estimators. The resulting fitted model based on the first-order clipped moment estimator better captures the empirical tail dependence structure than that based on the empirical variogram estimator.

In the remainder of this paper, we first review the necessary background in Section 2, and then describe the construction method of the moment estimators in Section 3. The main theoretical results are presented in Section 4. A simulation study is conducted in Section 5, followed by a case study on daily discharges from the Danube and the US flight delay data in Section 6. Section 7 concludes. All proofs are deferred to Section A. Additional simulation results are provided in Section B.

Notation. Throughout the paper, bold symbols refer to multivariate quantities. Comparison and arithmetic operators between two vectors or between a vector and a real number are understood componentwise. For instance, for any d-dimensional vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, we write "$\boldsymbol{x} \le \boldsymbol{y}$" if "$x_i \le y_i$" for all $1 \le i \le d$. For a real number T, "$\boldsymbol{x} \le T$" means "$x_i \le T$" for all $i \in \{1, \dots, d\}$. Furthermore, "$\boldsymbol{x} \vee \boldsymbol{y}$" denotes the vector whose i-th component is $\max(x_i, y_i)$. The relation $X \stackrel{d}{=} Y$ indicates that the random variables X and Y are identically distributed. For sets U and V, the Cartesian product is denoted by $U \times V$. Vectors are treated as column vectors by default.

2 Background

2.1 Multivariate generalized Pareto distributions

The MGPD models the extremal dependence of a random vector through threshold exceedances. To make abstraction of the univariate marginal distributions and concentrate on the extremal dependence, we introduce the MGPD on a canonical exponential scale. The MGPD with other margins can be obtained via monotone marginal transformations.

Let $d > 1$ and $V = \{1, \ldots, d\}$. Consider a random vector $\tilde{\boldsymbol{X}} = (\tilde{X}_i,\, i \in V)$ with standard exponential margins, i.e.,

$$\mathbb{P}(\tilde{X}_i \le x) = 1 - \exp(-x), \quad x \ge 0.$$

Define $\mathcal{L} = \{\boldsymbol{y} \in [-\infty, \infty)^d : \boldsymbol{y} \nleq \boldsymbol{0}\}$. A random vector $\boldsymbol{Y} = (Y_v, v \in V)$ is said to follow a *standard multivariate generalized Pareto distribution* if it can arise as the limit

$$\mathbb{P}(\boldsymbol{Y} \le y) = \lim_{u \to \infty} \mathbb{P}(\tilde{\boldsymbol{X}} - u \le \boldsymbol{y} \mid \tilde{\boldsymbol{X}} \nleq u), \quad \boldsymbol{y} \in \mathcal{L}. \tag{2.1}$$

In this case, the random vector $\tilde{\boldsymbol{X}}$ is said to belong to the domain of attraction of $\boldsymbol{Y}$. General MGPDs with generalized Pareto margins can be obtained through componentwise transformations

$$\sigma_i \frac{\exp(\xi_i Y_i) - 1}{\xi_i}, \quad i \in V,$$

where $\sigma_i \in (0, \infty)$ and $\xi_i \in \mathbb{R}$. We refer to Resnick (2008), Beirlant et al. (2004) and Naveau and Segers (2024) for more details of the MGPD and the multivariate extreme value theory. Since we are mainly interested in the dependence structure, we will focus on the standard MGPD in the following.

For a standard MGPD on the exponential scale, it admits a useful stochastic representation. For $m \in V$, let $\boldsymbol{Y}^{(m)}$ denote the conditional distribution of the exceedance vector given that the m-th component exceeds the threshold zero, i.e., $\boldsymbol{Y}^{(m)} = \boldsymbol{Y} \mid Y_m > 0$, with support $\mathcal{L}^m = \{\boldsymbol{y} \in [-\infty, \infty)^d \,|\, y_m > 0\}$. Then,

$$\boldsymbol{Y}^{(m)} \overset{d}{=} E + \boldsymbol{Z}^{(m)} = \left(E + Z_1^{(m)}, \ldots, E + Z_d^{(m)}\right), \tag{2.2}$$

where $\boldsymbol{Z}^{(m)} = (Z_v^{(m)}, v \in V)$ is a d-dimensional random vector with $Z_m^{(m)} = 0$ almost surely and where E is a standard exponential random variable, with distribution function $\mathbb{P}(E \le x) = 1 - \exp(-x)$, $x \ge 0$, which is independent of $\boldsymbol{Z}^{(m)}$ (see Engelke and Volgushev (2022, Eq. (6)); Naveau and Segers (2024, Definition 2.1)). The representation separates the extremal magnitude and dependence structure. The common exponential component E describes the size of the extreme event, while $\boldsymbol{Z}^{(m)}$ determines the dependence among components in the tail region.

2.2 Hüsler–Reiss MGPD

A prominent and tractable subclass of MGPD is the Hüsler–Reiss MGPD. In this model, the dependence structure is fully characterized by a matrix $\boldsymbol{\Gamma} = (\gamma_{ij})_{i,i \in V}$ called *variogram matrix*, which is a $d \times d$ dimensional symmetric, conditionally negative definite matrix with elements $\gamma_{ij} \in [0, \infty)$ satisfying $\gamma_{ii} = 0$ for $i \in V$. That is, $\boldsymbol{\Gamma} \in \mathcal{M}$ with

$$\mathcal{M} = \left\{ \boldsymbol{\Gamma} \in [0, \infty)^{d \times d} : \boldsymbol{\Gamma} = \boldsymbol{\Gamma}^\top,\ \mathrm{diag}(\boldsymbol{\Gamma}) = 0,\ \text{and } \boldsymbol{u}^\top \boldsymbol{\Gamma} \boldsymbol{u} < 0,\ \forall \boldsymbol{u} \neq \boldsymbol{0} \in \mathbb{R}^d,\ \sum_{i \in V} u_i = \boldsymbol{0} \right\}. \tag{2.3}$$

A Hüsler–Reiss MGPD can be defined through the stochastic representation of the standard MGPD. Specifically, a random vector $\boldsymbol{Y}$ is said to follow a *Hüsler–Reiss MGPD* with variogram matrix $\mathbf{\Gamma} = (\gamma_{ij}) \in \mathcal{M}$ if, for any $m \in V$, the conditional excess vector admits the representation (2.2), i.e.,

$$\boldsymbol{Y}^{(m)} \overset{d}{=} E + \boldsymbol{Z}^{(m)},$$

where E is a standard exponential random variable independent of $\boldsymbol{Z}^{(m)}$, and $\boldsymbol{Z}^{(m)}$ is a (possibly degenerate) Gaussian random vector satisfying

$$Z_m^{(m)} = 0, \quad \text{a.s.}$$

with mean vector $\boldsymbol{\mu}^{(m)}$ and covariance matrix $\mathbf{\Sigma}^{(m)}$ given by

$$\mu_i^{(m)} = -\frac{1}{2}\gamma_{im}, \quad \Sigma_{ij}^{(m)} = \frac{1}{2}(\gamma_{im} + \gamma_{jm} - \gamma_{ij}).$$

Since $\mathbf{\Gamma}$ is symmetric, i.e.,

$$\gamma_{ij} = \gamma_{ji}, \quad i, j \in V,$$

the marginal distributions of $Z_i^{(j)}$ and $Z_j^{(i)}$ are identical, with

$$Z_j^{(i)} \overset{d}{=} Z_i^{(j)} \sim N(-\gamma_{ij}/2, \gamma_{ij}).$$

2.3 Empirical variogram estimator

The variogram matrix is not only a fundamental parameter for the Hüsler–Reiss MGPD, indeed, it has been extended to general MGPDs by Engelke and Volgushev (2022). To facilitate statistical inference, they proposed the empirical variogram estimator based on data in the domain of attraction of the MGPD. Although it was originally defined using the multivariate Pareto representation, we reformulate it on the exponential scale here to maintain consistency with the notation and framework adopted in this paper.

Consider a random vector $\boldsymbol{Y}$ following a standard MGPD on the exponential scale. Let $\boldsymbol{X} = (X_i, i \in V)$ denote a d–variate random vector with continuous marginal functions F_i, $i \in V$. Denote $F(\boldsymbol{X}) = (F_i(X_i), i \in V)$. Assume that, after marginal transformation

$$\boldsymbol{X} \mapsto \tilde{\boldsymbol{X}} = -\log(1 - F(\boldsymbol{X})), \tag{2.4}$$

the standardized random vector $\tilde{\boldsymbol{X}}$, with standard exponential margins, belongs to the domain of attraction of $\boldsymbol{Y}$. Suppose $\boldsymbol{X}_t = (X_{t1}, \ldots, X_{td})$, $t = 1, \ldots, n$, are independent copies of $\boldsymbol{X}$.

Let $k = k(n)$ be an integer sequence satisfying $k \to \infty$ and $k/n \to q \in [0, 1]$ as $n \to \infty$. Denote the logarithmic transformation to the empirical Pareto exceedance by

$$\hat{Y}_{ti} = \log\left\{\frac{k}{n(1 - \hat{F}_i(X_{ti}))}\right\}, \; i \in V, \; t = 1, \ldots, n, \tag{2.5}$$

where

$$\hat{F}_i(x) = \frac{1}{n+1}\sum_{t=1}^{n} \mathbb{1}\{X_{ti} \le x\}$$

is the (adjusted) empirical distribution function of the i-th variable. The *empirical variogram estimator rooted at* m, denoted by $\hat{\mathbf{\Gamma}}^{(m)} = \left(\hat{\gamma}_{n,ij}^{(m)}\right)$, proposed by Engelke and Volgushev (2022) is given as

$$\hat{\gamma}_{n,ij}^{(m)} = \widehat{\text{var}}\left\{(\hat{Y}_{ti} - \hat{Y}_{tj}) : \hat{F}_m(X_m) \geq 1 - k/n\right\}, \; i, j, m \in V. \tag{2.6}$$

where $\widehat{\text{var}}$ denotes the sample variance of $\hat{Y}_{ti} - \hat{Y}_{tj}$ for those $t \in \{1, \ldots, n\}$ such that $\hat{F}_m(X_{tm}) \geq 1 - k/n$.

For a Hüsler–Reiss MGPD with variogram matrix $\mathbf{\Gamma}$, it turns out that $\mathbf{\Gamma}$ can be estimated by any $\hat{\mathbf{\Gamma}}^{(m)}$, $m \in V$, so does the averaged empirical estimator introduced also by Engelke and Volgushev (2022) (hereafter referred to as the empirical variogram estimator)

$$\hat{\gamma}_{n,ij}^{\text{EMP}} = \frac{1}{d} \sum_{m \in V} \hat{\gamma}_{n,ij}^{(m)}, \tag{2.7}$$

with $\hat{\gamma}_{n,ij}^{(m)}$ given by (2.6).

Owing to its computational simplicity and ease of implementation, the empirical estimator has been widely adopted in practice. However, in heterogeneous dependence structures, it may suffer from bias due to the inclusion of non-extreme observations. Specifically, for fixed $k \in \{1, \ldots, n\}$ and $(i, j) \in V \times V$ with $i \neq j$, the empirical variogram is the average of $\hat{\gamma}_{n,ij}^{(m)}$ over $m \in V$, where $\hat{\gamma}_{n,ij}^{(m)}$ is based on the points $(\hat{Y}_{ti}, \hat{Y}_{tj})$ for all $t \in \{1, \ldots, n\}$ such that $\hat{Y}_{tm} \geq 0$. For $m \notin \{i, j\}$, if variable m is only weakly dependent with variables i and/or j, then the empirical variogram also uses data points for which the observations of variables i and j are not large at all. Although the bias issue arises for general MGPDs, in this work we concentrate on the Hüsler–Reiss MGPD case. To mitigate this bias, we introduce a moment method that employs a clipping strategy on the lower tail of the Hüsler–Reiss MGPD vector, as described in the next section.

3 Moment-based variogram estimators

3.1 Moments and sample moments of clipped MGP random vectors

To eliminate the bias caused by the non-extreme values of the sample in the estimation of the variogram matrix, we propose an approach based on the clipped MGP random vector.

For a d-variate random vector $\boldsymbol{Y}$ with an arbitrary MGPD, we fix a constant $c \geq 0$ and consider the random vector clipped at level $-c$ given by

$$\left(Y_i^{(j)} \vee (-c), \; Y_j^{(j)}\right)$$

for each pair $(i, j) \in V \times V$ and $i \neq j$. Under the stochastic representation in (2.2), the pair above is distributed as

$$\left((E + Z_i^{(j)}) \vee (-c), \; E\right).$$

Assume that $\boldsymbol{Y}$ follows a parametric model with parameter vector $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$. Write $x_+ = x \vee 0$, and note that $(x \vee (-c)) + c = (x+c)_+$. We focus on the first- and second-order moments of the clipped vector shifted by c defined as

$$e_{ij}^{(\ell)}(\boldsymbol{\theta}, c) := \mathbb{E}\left\{\left(Y_i^{(j)} + c\right)_+^{\ell}\right\}, \qquad i, j \in V,\ \ell = 1, 2. \tag{3.1}$$

These clipped moments always exist for any $i, j \in V$ (see Lemma A.1 in Section A.2).

Recall that $\boldsymbol{X}_t = (X_{t1}, \ldots, X_{td})$, $t = 1, \ldots, n$, are independent copies of the random vector $\boldsymbol{X}$, where the marginally transformed vector $\tilde{\boldsymbol{X}}$ of $\boldsymbol{X}$ in (2.4) belongs to the domain of attraction of $\boldsymbol{Y}$ in the sense of Eq. (2.1). Let $k = k(n) \in \{1, \ldots, n\}$ be an intermediate sequence satisfying

$$k \to \infty,\ k/n \to 0,\ \text{as } n \to \infty.$$

Recall that $\hat{Y}_{ti}$ is the logarithmic transformation to the empirical Pareto exceedance defined in (2.5). Motivated by the convergence in (2.1), for $\ell = 1, 2$ and any $i, j \in V$, the sample versions of $e_{ij}^{(\ell)}(\boldsymbol{\theta}, c)$ in (3.1) can be constructed as

$$\begin{aligned}\hat{e}_{n,ij}^{(\ell)}(k, c) &= \frac{1}{k}\sum_{t=1}^{n}\left(\log\left\{\frac{k}{n(1-\hat{F}_i(X_{ti}))}\right\} + c\right)_+^{\ell} \mathbb{1}\left\{\hat{F}_j(X_{tj}) \geq 1 - \frac{k}{n}\right\} \\ &= \frac{1}{k}\sum_{t=1}^{n}\left(\hat{Y}_{ti} + c\right)_+^{\ell} \mathbb{1}\left\{\hat{F}_j(X_{tj}) \geq 1 - \frac{k}{n}\right\}.\end{aligned} \tag{3.2}$$

Hence, a straightforward idea is that we can estimate the parameter $\boldsymbol{\theta}$ by letting

$$\hat{e}_{n,ij}^{(\ell)}(k, c) = e_{ij}^{(\ell)}(\boldsymbol{\theta}, c), \qquad i, j \in V,\ \ell = 1, 2, \tag{3.3}$$

provided that the corresponding moment equations have a unique solution within the valid parameter space.

If the variogram matrix of a general MGPD is continuous in $\boldsymbol{\theta}$ for $m \in V$, then $\boldsymbol{\theta}$ can first be estimated by $\hat{\boldsymbol{\theta}}$ via the (generalized) method of moments, and the corresponding estimate of the variogram matrix can subsequently be obtained from $\hat{\boldsymbol{\theta}}$. This estimation idea applies, theoretically, to a general MGPD. However, in the following, we focus on its implementation for the Hüsler–Reiss MGPD.

3.2 Moment variogram estimators for a Hüsler–Reiss MGPD

In the following, we assume that $\boldsymbol{Y} = (Y_v, v \in V)$ is a d-variate random vector following a Hüsler–Reiss MGPD with parameter $\boldsymbol{\theta} = \boldsymbol{\Gamma}$, where $\boldsymbol{\Gamma} \in \mathcal{M}$ is a variogram matrix (cf. (2.3)). By a straightforward calculation, the explicit expressions of the moment functions defined in (3.1) for the Hüsler–Reiss MGPD can be derived, as stated in the lemma below.

Lemma 3.1 (clipped moment functions for a Hüsler–Reiss MGPD). *Assume the random vector $\boldsymbol{Y}$ follows a Hüsler–Reiss MGPD with variogram matrix $\boldsymbol{\Gamma} = (\gamma_{ij})_{i,j\in V}$. Then, for fixed $c \in [0, \infty)$ and each pair of $i, j \in V$, we have*

$$\mathbb{E}\left\{\left(Y_i^{(j)} + c\right)_+^{\ell}\right\} = e^{(\ell)}(\gamma_{ij}, c),$$

with

$$e^{(1)}(\gamma, c) = \exp(c)\Phi\left(-\frac{c+\gamma/2}{\sqrt{\gamma}}\right) + \sqrt{\gamma}\,\phi\left(\frac{c-\gamma/2}{\sqrt{\gamma}}\right) + \left(1 - \frac{\gamma}{2} + c\right)\Phi\left(\frac{c-\gamma/2}{\sqrt{\gamma}}\right), \tag{3.4}$$

and

$$\begin{aligned} e^{(2)}(\gamma, c) = {} & 2\exp(c)\Phi\left(-\frac{c+\gamma/2}{\sqrt{\gamma}}\right) - \sqrt{\gamma}\left(\frac{\gamma}{2} - c - 2\right)\phi\left(\frac{c-\gamma/2}{\sqrt{\gamma}}\right) \\ & + \left\{\frac{1}{4}\gamma^2 - \gamma c + (1+c)^2 + 1\right\}\Phi\left(\frac{c-\gamma/2}{\sqrt{\gamma}}\right), \end{aligned} \tag{3.5}$$

where $e^{(\ell)}(0; c)$ *is defined as the limit* $e^{(\ell)}(0, c) = \lim_{\gamma\to 0} e^{(\ell)}(\gamma, c)$ *with*

$$e^{(1)}(0, c) = \mathbb{E}\,(E + c) = 1 + c, \qquad e^{(2)}(0, c) = \mathbb{E}\,\{(E + c)^2\} = (1 + c)^2 + 1.$$

Moreover, both $e^{(1)}(\gamma, c)$ *and* $e^{(2)}(\gamma, c)$ *are strictly decreasing in* γ *on* $(0, \infty)$.

Note that for fixed c, the moment functions allow for the construction of moment estimators. In consideration of the symmetry of the variogram matrix, i.e., the property that $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}^\top$, we propose to estimate the entry γ_{ij} for any $i, j \in V$ and $i \neq j$ based on (3.3) by the moment estimators $\hat{\gamma}_{n,ij}^{\mathrm{M},(\ell)}$ defined by

$$e^{(\ell)}\left(\hat{\gamma}_{n,ij}^{\mathrm{M},(\ell)}(k, c), c\right) = \frac{1}{2}\left\{\hat{e}_{n,ij}^{(\ell)}(k, c) + \hat{e}_{n,ji}^{(\ell)}(k, c)\right\}, \quad \ell = 1, 2. \tag{3.6}$$

Here, the estimators are defined only when $i \neq j$, since the diagonal elements of $\boldsymbol{\Gamma}$ are equal to zero by definition. In practice, the choice for c is restricted to $[0, -\log(k/n)]$, since if $c \geq -\log(k/n)$, the clipping has no effect at all on the value of $\hat{e}_{n,ij}^{(\ell)}(k, c)$.

As the choice of the clipping parameter c is not the focus of this work, all subsequent discussions proceed under the assumption that c is a given number. To simplify the notation, for fixed c, we write $e^{(\ell)}(\gamma) := e^{(\ell)}(\gamma, c)$, and we suppress (k, c) in $\hat{e}_{n,ij}^{(\ell)}(k, c)$ and $\hat{\gamma}_{n,ij}^{\mathrm{M},(\ell)}(k, c)$ occasionally whenever no ambiguity arises.

4 Consistency and asymptotic normality

In this section, we present the theoretical properties of the proposed estimators. We start by establishing the weak consistency of the moment estimators defined in (3.6). The following proposition shows that the empirical moments $\hat{e}_{n,ij}^{(\ell)}(k, c)$ converge to their true counterparts for an arbitrary MGPD, providing a crucial step toward proving the weak consistency result stated in the subsequent theorem. Before stating the main result, we collect the conditions on $k(n)$ in the following assumption.

Assumption 4.1. Assume $k = k(n)$ is an intermediate sequence such that $k \to \infty$ and $k/n \to 0$ as $n \to \infty$.

Proposition 4.2. *Let $\boldsymbol{Y}$ be an arbitrary MGP distributed random vector, and let $\boldsymbol{X}_t = (X_{t1}, \ldots, X_{td})$, $t = 1, \ldots, n$, denote independent copies of a random vector $\boldsymbol{X}$, which has continuous margins and the marginally transformed vector of $\boldsymbol{X}$ in* (2.4) *lies in the domain of attraction of $\boldsymbol{Y}$. For $c \geq 0$ and for $i, j \in V$ with $i \neq j$, let $\hat{e}^{(\ell)}_{n,ij}(k, c)$ with $\ell = 1, 2$ be defined as in* (3.2)*. Then, under Assumption 4.1, we have*

$$\hat{e}^{(\ell)}_{n,ij}(k, c) \xrightarrow{\mathbb{P}} \mathbb{E}\left\{\left(Y_i^{(j)} + c\right)_+^{\ell}\right\}, \quad \text{as } n \to \infty.$$

For a Hüsler–Reiss MGPD, we show in Theorem 4.4 that the proposed moment estimators for the variogram are well-defined and converge weakly to the true values. The standing assumption is stated as follows.

Assumption 4.3. Assume $\boldsymbol{Y}$ is a Hüsler–Reiss MGP distributed random vector parameterized by a $d \times d$ dimensional variogram matrix $\boldsymbol{\Gamma} \in \mathcal{M}$ stated in (2.3). Suppose $\boldsymbol{X}_t = (X_{t1}, \ldots, X_{td})$, $t = 1, \ldots, n$, are independent copies of a random vector $\boldsymbol{X}$, which has continuous margins and the marginally transformed vector of $\boldsymbol{X}$ in (2.4) lies in the domain of attraction of $\boldsymbol{Y}$ in the sense of (2.1).

Theorem 4.4. *Suppose that Assumptions 4.1 and 4.3 hold. For $i, j \in V$ and $i \neq j$, the moment estimators in* (3.6) *are well-defined with probability tending to one, and $\hat{\gamma}^{\mathrm{M},(\ell)}_{n,ij} \xrightarrow{\mathbb{P}} \gamma_{ij}$ as $n \to \infty$ for $\ell = 1, 2$.*

Recall that F_i is the marginal distribution of X_i, $i \in V$. Let

$$U_i = 1 - F_i(X_i), \quad i = 1, \ldots, d,$$

and denote the joint distribution function of $\boldsymbol{U} = (U_1, \ldots, U_d)$ by $C(\boldsymbol{x})$. The convergence of multivariate threshold exceedances in (2.1) is equivalent to the existence of the limit

$$\lim_{q \to 0} q^{-1}\,\mathbb{P}\left(1 - F_i(X_i) \leq q x_i, i \in V\right) = \lim_{q \to 0} q^{-1} C(q\boldsymbol{x}) = R(\boldsymbol{x}) \tag{4.1}$$

for $\boldsymbol{x} \in [0, \infty]^d \setminus \{(\infty, \ldots, \infty)\}$, where $R(\boldsymbol{x})$ is called the *tail copula* of $\boldsymbol{X}$, see, e.g., Schmidt and Stadtmüller (2006) and Chiapino et al. (2019).

For any non-empty set $I \subset V$ and vector $\boldsymbol{x}_I = (x_i, i \in I) \in [0, \infty]^{|I|} \setminus \{(\infty, \ldots, \infty)\}$, define $R_I(\boldsymbol{x}_I)$ as the value of the function $R(\boldsymbol{x})$ evaluated at the point $\boldsymbol{x}$ whose components are x_i for $i \in I$ and ∞ for $i \in V \setminus I$. Let

$$\dot{R}_I(\boldsymbol{x}_I) = (\dot{R}^i_I(\boldsymbol{x}_I), i \in I) = (\partial R_I(\boldsymbol{x}_I)/\partial x_i, i \in I)$$

be the vector of its first-order partial derivatives.

Assume W is a mean-zero Gaussian process on $[0, \infty]^d \setminus \{(\infty, \ldots, \infty)\}$ with continuous trajectories and covariance function

$$\mathbb{E}\left(W(\boldsymbol{x})W(\boldsymbol{y})\right) = R(\boldsymbol{x} \wedge \boldsymbol{y}), \qquad \boldsymbol{x}, \boldsymbol{y} \in [0, \infty]^d \setminus \{(\infty, \ldots, \infty)\}, \tag{4.2}$$

with $\boldsymbol{x} \wedge \boldsymbol{y} = (x_i \wedge y_i, i \in V)$. For any nonempty set $I \subseteq V$ and $\boldsymbol{x}_I \in [0,\infty]^{|I|} \setminus \{(\infty,\dots,\infty)\}$, define

$$W_I(\boldsymbol{x}_I) = W(\boldsymbol{y}),$$

where $\boldsymbol{y} \in [0,\infty]^d \setminus \{(\infty,\dots,\infty)\}$ is a vector such that $y_i = x_i$ for $i \in I$ and $y_i = \infty$ for $i \notin I$. In particular, $W_i(x_i) = W(\infty,\dots,\infty,x_i,\infty,\dots,\infty)$, where x_i appears in the i-th component. Define

$$B_I(\boldsymbol{x}_I) = W_I(\boldsymbol{x}_I) - \sum_{i \in I} \dot{R}_I^i(\boldsymbol{x}_I) W_i(x_i), \tag{4.3}$$

which is a zero-mean stochastic process on $\boldsymbol{x}_I \in [0,\infty]^{|I|} \setminus \{(\infty,\dots,\infty)\}$.

Next, we show the asymptotic normality of the empirical moments in Proposition 4.7. Based on this result and using the delta method, the asymptotic normality of the moment estimators can be established, as shown in Theorem 4.8. A second-order condition, as stated in Assumption 4.5, is required to control the convergence rate in (2.1). Since the subsequent results rely on the continuity of the partial derivatives of the tail copula, we restrict our attention to the case $\gamma_{ij} > 0$ for all $i,j \in V$ with $i \neq j$, thereby excluding the degenerate case $\gamma_{ij} = 0$, in accordance with the assumption in Assumption 4.6.

Assumption 4.5. There exist constants $\xi, K \in (0,\infty)$ such that for any $I \subseteq V$ with $|I| = 2$ and $q \in (0,1)$, we have

$$\sup_{x_I \in [1,\infty]^{|I|}} \left| \mathbb{P}\left(\boldsymbol{F}_I(\boldsymbol{X}_I) \le 1 - q/\boldsymbol{x}_I \mid \boldsymbol{F}_I(\boldsymbol{X}_I) \nleq 1 - q\right) - \mathbb{P}(\boldsymbol{Y}_{(I)} \le \boldsymbol{x}) \right| \le K q^{\xi},$$

where $\boldsymbol{Y}_{(I)}$ is the random vector obtained from (2.1) with $\boldsymbol{X}$ replaced by $\boldsymbol{X}_I$.

Assumption 4.6. The off-diagonal elements of the variogram matrix $\boldsymbol{\Gamma}$ associated with the Hüsler–Reiss MGPD in Assumption 4.3 are strictly positive, that is, $\gamma_{ij} > 0$ for all $i,j \in V$ and $i \neq j$.

Proposition 4.7. *If Assumptions 4.1, 4.3, 4.5 and 4.6 hold with $k = o\left(n^{\xi/(\xi+\frac{1}{2})}\right)$, then*

$$\begin{aligned}&\left(\sqrt{k}\left\{\hat{e}^{(\ell)}_{n,ij}(k,c) - e^{(\ell)}(\gamma_{ij},c)\right\},\, i,j \in V,\, i \neq j,\, \ell = 1,2\right)\\ &\qquad\xrightarrow{d} \left(\int_0^{\exp(c)} \frac{B_{ij}(x,1)\cdot \ell \cdot (-\log x + c)^{\ell-1}}{x}\,\mathrm{d}x,\, i,j \in V,\, i \neq j,\, \ell = 1,2\right)\end{aligned}$$

as $n \to \infty$, where for each pair $(i,j) \in V \times V$ such that $i \neq j$, $B_{ij}(x,1)$ is the stochastic process defined in (4.3) *and given by*

$$B_{ij}(x,1) = W_{ij}(x,1) - \dot{R}^i_{ij}(x,1)W_i(x) - \dot{R}^j_{ij}(x,1)W_j(1). \tag{4.4}$$

Theorem 4.8. *Under the assumptions of Proposition 4.7, we have*

$$\left(\sqrt{k}(\hat{\gamma}^{\mathrm{M},(\ell)}_{n,ij} - \gamma_{ij});\; i,j \in V,\, i \neq j\right)$$

$$\xrightarrow{d} \left(\frac{1}{2e^{(\ell),\prime}(\gamma_{ij})} \int_0^{\exp(c)} \frac{(B_{ij}(x,1) + B_{ji}(x,1)) \cdot \ell \cdot (-\log x + c)^{\ell-1}}{x} \,\mathrm{d}\,x;\, i,j \in V,\, i \neq j \right)$$

as $n \to \infty$*, where* $B_{ij}(x,1)$ *is given in* (4.4)*, and where* $e^{(\ell),\prime}(\gamma)$ *is the derivative of* $e^{(\ell)}(\gamma)$ *in* (3.4) *and* (3.5) *with respect to* γ*, given by*

$$e^{(1),\prime}(\gamma) = -\frac{1}{2}\,\Phi\left(\frac{c-\frac{\gamma}{2}}{\sqrt{\gamma}}\right), \qquad e^{(2),\prime}(\gamma) = \left(\frac{\gamma}{2} - c\right)\Phi\left(\frac{c-\frac{\gamma}{2}}{\sqrt{\gamma}}\right) - \sqrt{\gamma}\,\phi\left(\frac{c-\frac{\gamma}{2}}{\sqrt{\gamma}}\right).$$

5 Simulation study

In this section, we study the finite-sample behavior of the proposed estimators on simulated data. To investigate the performance of the moment estimator in (3.6), we compare its finite-sample behavior with that of the empirical variogram estimator in (2.7). For convenience, we set $c = -\log a$ with $a \in (0,1]$. All plots in the simulation study and empirical analysis are presented in terms of a, since the range $(0,1]$ is easier for selecting values.

We first generate $n = 1000$ independent samples $\boldsymbol{X}_1^\star, \ldots, \boldsymbol{X}_n^\star$ from two Hüsler–Reiss max-stable models, a 10-dimensional model with randomly generated variogram matrix $\boldsymbol{\Gamma}_1$ (see Section B), and a 5-dimensional model with variogram matrix

$$\boldsymbol{\Gamma}_2 = \begin{pmatrix} 0 & 3 & 4 & 7 & 6 \\ 3 & 0 & 3 & 8 & 5 \\ 4 & 3 & 0 & 7 & 8 \\ 7 & 8 & 7 & 0 & 11 \\ 6 & 5 & 8 & 11 & 0 \end{pmatrix}.$$

Both models are in the domains of attraction of the corresponding Hüsler–Reiss MGPDs with the same variogram matrices. The matrix $\boldsymbol{\Gamma}_1$ is obtained by first generating a positive definite matrix $\boldsymbol{\Sigma}$ and then computing the variogram matrix using the function `Sigma2Gamma` from the `graphicalExtremes` package in software `R`. The second distribution is the example considered in Engelke et al. (2026) with the name of *non-faithful Hüsler–Reiss distribution*, since Y_2 and Y_4 are conditionally independent (in the sense of extremal conditional independence defined in Engelke and Hitz (2020)) given Y_1 and Y_3. In order to perturb the samples, we add standard normally distributed noise. To be precise, we set

$$\boldsymbol{X}_t = \boldsymbol{X}_t^\star + |\boldsymbol{\epsilon}_t|, \qquad t = 1, \ldots, n,$$

where $\boldsymbol{\epsilon}_t = (\epsilon_{ti},\, i \in V)$, for $t = 1, \ldots, n$, are independent standard normal random variables, independent of $\boldsymbol{X}_t^\star$, $t = 1, \ldots, n$.

Under these simulation settings, we investigate the finite-sample performance of the proposed two moment estimators and compare it with the empirical variogram estimator. The performance of the estimators is evaluated by computing the average relative squared error between the estimates $\hat{\gamma}_{n,ij}$ and their true counterparts γ_{ij} of the variogram matrix $\mathbf{\Gamma}$, defined as

$$L(\hat{\mathbf{\Gamma}}, \mathbf{\Gamma}) = \left\{ \frac{2}{d(d-1)} \sum_{i,j \in V, i<j} (\hat{\gamma}_{n,ij}/\gamma_{ij} - 1)^2 \right\}^{1/2},$$

where $\hat{\mathbf{\Gamma}} = (\hat{\gamma}_{n,ij})_{i,j \in V}$ denotes the estimates of $\mathbf{\Gamma}$ based on one of the aforementioned estimators. In addition, we also measure the average distance between the model-based and empirical (data-based) tail dependence coefficients, namely,

$$D(\hat{\boldsymbol{\chi}}, \hat{\boldsymbol{\chi}}^{\mathrm{EST}}) = \frac{2}{d(d-1)} \sum_{i,j \in V, i<j} |\hat{\chi}_{ij} - \hat{\chi}_{ij}^{\mathrm{EST}}| \tag{5.1}$$

where $\hat{\boldsymbol{\chi}} = (\hat{\chi}_{ij})_{i,j \in V}$ and $\hat{\boldsymbol{\chi}}^{\mathrm{EST}} = (\hat{\chi}_{ij}^{\mathrm{EST}})_{i,j \in V}$ denote the empirical and model-based tail dependence coefficient matrix, respectively. More specifically, for each pair $(i,j) \in V^2$ with $i \neq j$, the empirical tail dependence coefficient is calculated by

$$\hat{\chi}_{ij} = \frac{1}{k} \sum_{t=1}^{n} \mathbb{1}\left\{ 1 - \hat{F}_i(X_{ti}) \leq \frac{k}{n}, 1 - \hat{F}_j(X_{tj}) \leq \frac{k}{n} \right\},$$

with the same k used in the estimation of the variogram matrix, and the model-based tail dependence coefficients are obtained via

$$\hat{\chi}_{ij}^{\mathrm{EST}} = 2\left\{ 1 - \Phi\left(\sqrt{\hat{\gamma}_{n,ij}}/2 \right) \right\}.$$

The simulation results for samples from the 10−dimensional Hüsler–Reiss model are shown in Figures 1–4. The impact of the clipping level c (or a equivalently) on the performance of the moment estimators $\hat{\gamma}_{n,ij}^{\mathrm{M},(\ell)}(k,c)$, $\ell = 1,2$, is examined first. Figures 1 and 2 present the mean L and D corresponding to the first-and second-order moment estimator with $a = \exp(-c) = 0.25, 0.5$ and $k = 50, 75, \ldots, 250$, respectively. For estimator $\hat{\gamma}_{n,ij}^{\mathrm{M},(1)}(k, -\log a)$, we see that smaller values of a yield better performance for small k, whereas larger values of a become preferable as k increases. This indicates a trade-off between a and k, and suggests the existence of an optimal choice of k in terms of L. In contrast, the metric D exhibits a different trend, where smaller a always results in smaller deviations. For estimator $\hat{\gamma}_{n,ij}^{\mathrm{M},(2)}(k, -\log a)$ in Figure 2, both L and D consistently indicate that smaller a values yield lower estimation bias.

To compare the performance of the estimators, we set $a = 0.25$ for $\hat{\gamma}_{n,ij}^{\mathrm{M},(\ell)}(k, -\log a)$, $\ell = 1,2$, and show the boxplots of L and D values based on 300 replications for each of the three estimators in Figure 3. It suggests that the first-order moment estimator $\hat{\gamma}_{n,ij}^{\mathrm{M},(1)}(k, -\log a)$ outperforms the empirical variogram estimator, especially when k is small, whereas the second-order moment estimator only shows superior performance for large k, in terms of L. In Figure 4, we further analyze the averaged bias, variance and MSE over 300 replications of the three estimators $\hat{\gamma}_{n,ij}^{\mathrm{EMP}}(k)$, $\hat{\gamma}_{n,ij}^{\mathrm{M},(1)}(k, -\log a)$ and $\hat{\gamma}_{n,ij}^{\mathrm{M},(2)}(k, -\log a)$ for the element $\gamma_{12} = 0.89$ of $\mathbf{\Gamma}_1$, where $a = 0.25$ and $k = 50, 75, \ldots, 250$. The results indicate that, the empirical estimator tends to have a larger bias but smaller variance. The proposed first-order moment estimator is particularly advantageous for small k, and in some cases it uniformly dominates the empirical estimator in terms of MSE. Additional results for other elements of $\mathbf{\Gamma}_1$ can be found in Figures 11–13 in Section B.

The corresponding results for random samples from the 5-dimensional Hüsler–Reiss distribution with variogram matrix $\mathbf{\Gamma}_2$ are illustrated in Figures 5–8. The estimation error L varies with a in a manner similar to those of $\hat{\gamma}_{n,ij}^{\mathrm{M},(1)}(k, -\log a)$ observed for the 10-dimensional Hüsler–Reiss distribution. Moreover, the value of D exhibits greater stability with respect to changes in k. It could be noted in Figure 7 that, under such circumstances, namely, when the extremal dependence between certain components is relatively weak or asymptotically conditional independent, the moment estimators yield improved performance. We see that both moment estimators $\hat{\gamma}_{n,ij}^{\mathrm{M},(\ell)}$, $\ell = 1, 2$, have much lower estimation error than the empirical variogram estimator. In this case, Figure 8 indicates that $\hat{\gamma}_{n,ij}^{\mathrm{M},(2)}$ achieves the smallest bias and MSE for large values of k, while the empirical variogram estimator still performs best in terms of variance. For more details on other elements, see Figures 14–16 in Section B.

Indeed, what distinguishes the two matrices $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$ is that the dependence in $\mathbf{\Gamma}_2$ is weaker. The simulations show that, for the Hüsler–Reiss models with stronger dependence, the moment estimator tends to achieve better performance with relatively small values of a (or c equivalently). However, such a pattern is not clear for weakly dependent models. Moreover, in both cases, the clipped moment estimators could provide improvements over the empirical variogram estimator, especially when the dependence is weak. Overall, the first-moment estimator with $a = 0.25$ comes out as the best one: in the two L and D-plots, it has the lowest error in almost all cases. In addition, compared with the empirical variogram estimator, the two moment estimators appear to be less sensitive to the choice of k in view of D-plots, when the dependence is weak.

6 Application

6.1 Danube discharges data

To illustrate the practical performance of the proposed method, we fit a Hüsler–Reiss model to the Danube discharge dataset. The variogram matrix $\mathbf{\Gamma}$ is estimated using the proposed moment estimators, and the resulting estimates are compared with those obtained from the empirical variogram estimator in Engelke and Volgushev (2022). This dataset includes the average daily discharges recorded at 31 gauging stations in the upper Danube basin covering parts of Germany, Austria and Switzerland, available in the supplementary material of Asadi et al. (2015). The series at individual stations has lengths from 54 to 113 years, with 51 years of data for all stations from 1960 to 2010. We follow Asadi et al. (2015) and use the pre-processed data containing $n = 428$ observations after a declustering of the time series to perform the estimation, considering the declustered samples as independent observations from a 31-variate random vector $\boldsymbol{X}$ in the domain of attraction of a Hüsler–Reiss distribution with variogram matrix $\mathbf{\Gamma}$.

To investigate how the proportion of samples used for estimation affects the results, we

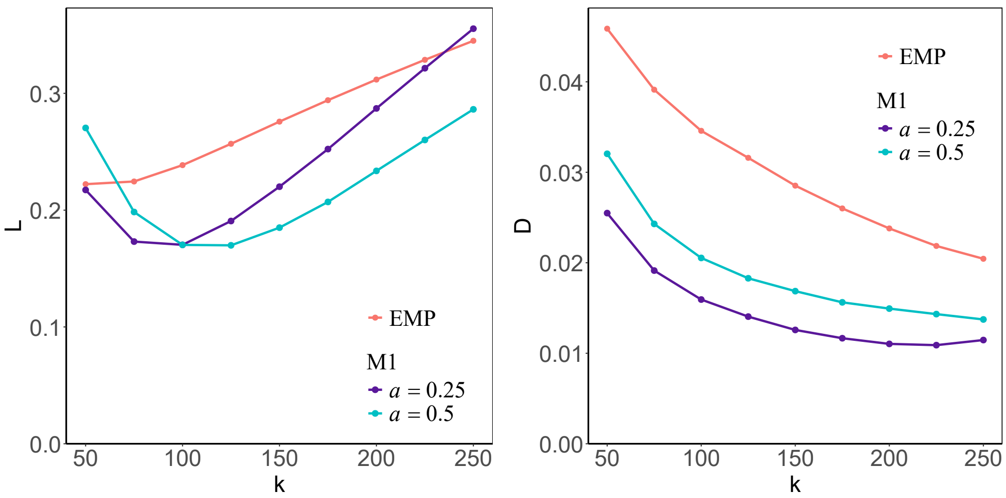


Fig. 1 The mean L (left) and D value (right) based on first-order moment estimator $\hat{\gamma}_{n,ij}^{\mathrm{M},(1)}(k, -\log a)$ (M1) with $a = 0.25, 0.5$ and the empirical estimator $\hat{\gamma}_{n,ij}^{\mathrm{EMP}}$ (EMP) over 300 replications, with random samples of size $n = 1000$ drawn from the 10-dimensional Hüsler–Reiss distribution with variogram matrix $\mathbf{\Gamma}_1$.

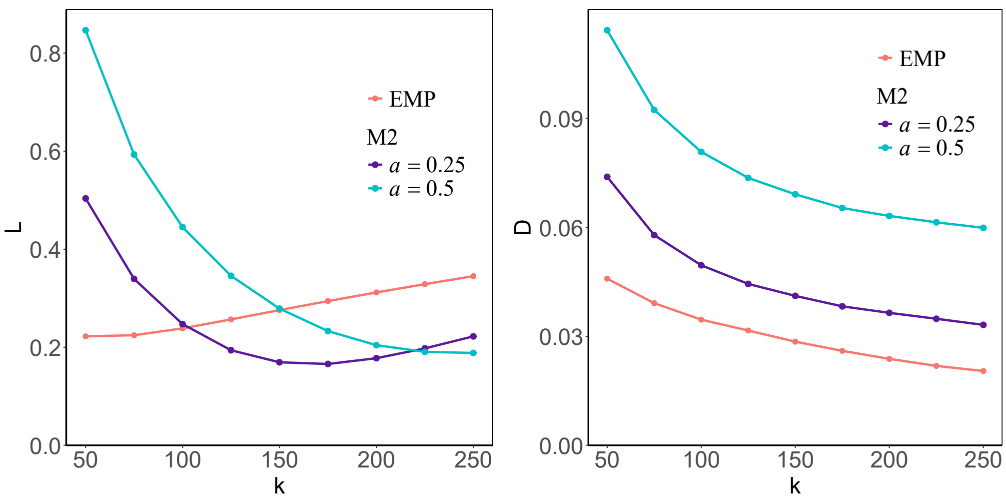


Fig. 2 The mean L (left) and D value (right) based on second-order moment estimator $\hat{\gamma}_{n,ij}^{\mathrm{M},(2)}(k, -\log a)$ (M2) with $a = 0.25, 0.5$ and the empirical variogram estimator $\hat{\gamma}_{n,ij}^{\mathrm{EMP}}$ (EMP) over 300 replications, with random samples of size $n = 1000$ drawn from the 10-dimensional Hüsler–Reiss distribution with variogram matrix $\mathbf{\Gamma}_1$.

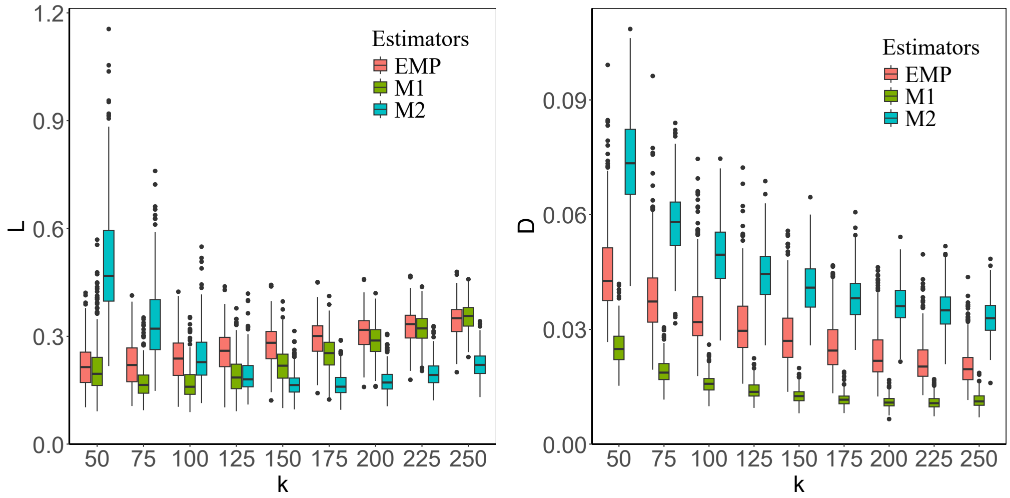


Fig. 3 The L value (left) and distance D of all bivariate tail dependence coefficients (right) based on the empirical variogram estimator $\hat{\gamma}_{n,ij}^{\mathrm{EMP}}$ (EMP), first-order moment estimator $\hat{\gamma}_{n,ij}^{\mathrm{M},(1)}(k, -\log a)$ (M1) and second-order moment estimator $\hat{\gamma}_{n,ij}^{\mathrm{M},(2)}(k, -\log a)$ (M2) with $a = 0.25$ in 300 replications. The random samples with size $n = 1000$ are drawn from the 10–dimensional Hüsler–Reiss distribution with variogram matrix $\mathbf{\Gamma}_1$.

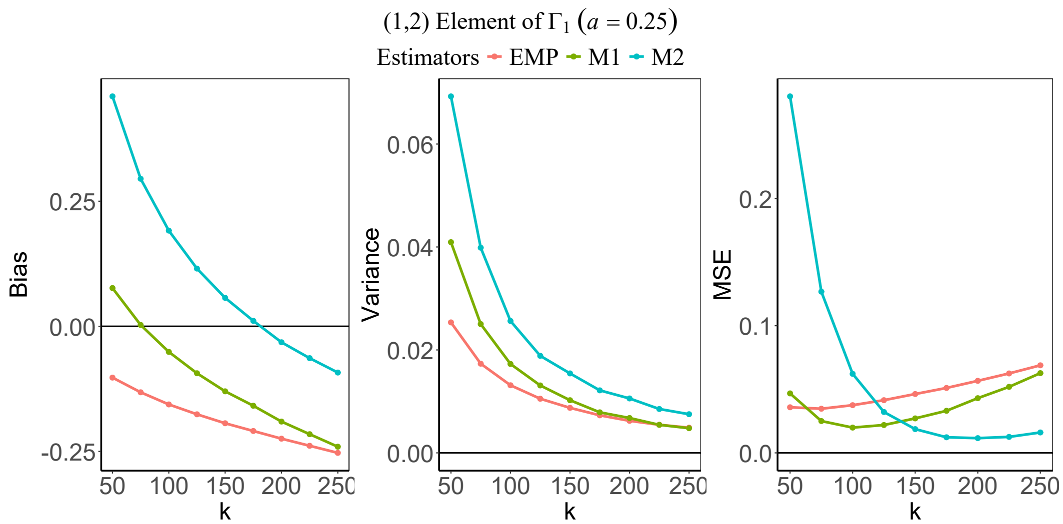


Fig. 4 The mean estimation bias, variance and MSE for $\gamma_{12} = 0.89$ of $\mathbf{\Gamma}_1$, based on the empirical variogram estimator $\hat{\gamma}_{n,ij}^{\mathrm{EMP}}$ (EMP), first-order moment estimator $\hat{\gamma}_{n,ij}^{\mathrm{M},(1)}(k, -\log a)$ (M1) and second-order moment estimator $\hat{\gamma}_{n,ij}^{\mathrm{M},(2)}(k, -\log a)$ (M2) with $a = 0.25$ in 300 replications. The random samples with size $n = 1000$ are drawn from the 10–dimensional Hüsler–Reiss distribution with variogram matrix $\mathbf{\Gamma}_1$.

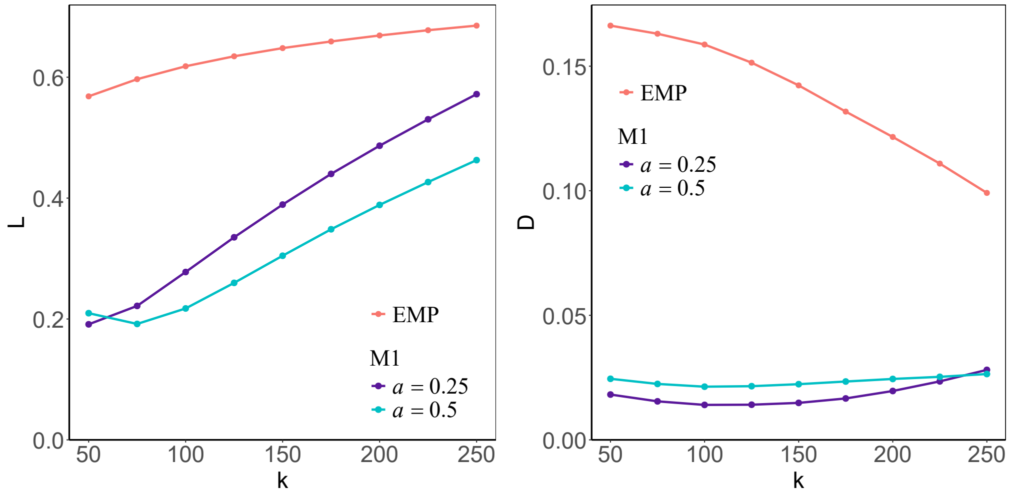


Fig. 5 The mean L (left) and D value (right) based on the first-order moment estimator $\hat{\gamma}_{n,ij}^{\mathrm{M},(1)}(k, -\log a)$ (M1) and the empirical variogram estimator $\hat{\gamma}_{n,ij}^{\mathrm{EMP}}$ (EMP) with $a = 0.25, 0.5$ over 300 replications, with random samples of size $n = 1000$ drawn from the 5-dimensional Hüsler–Reiss distribution with variogram matrix $\mathbf{\Gamma}_2$.

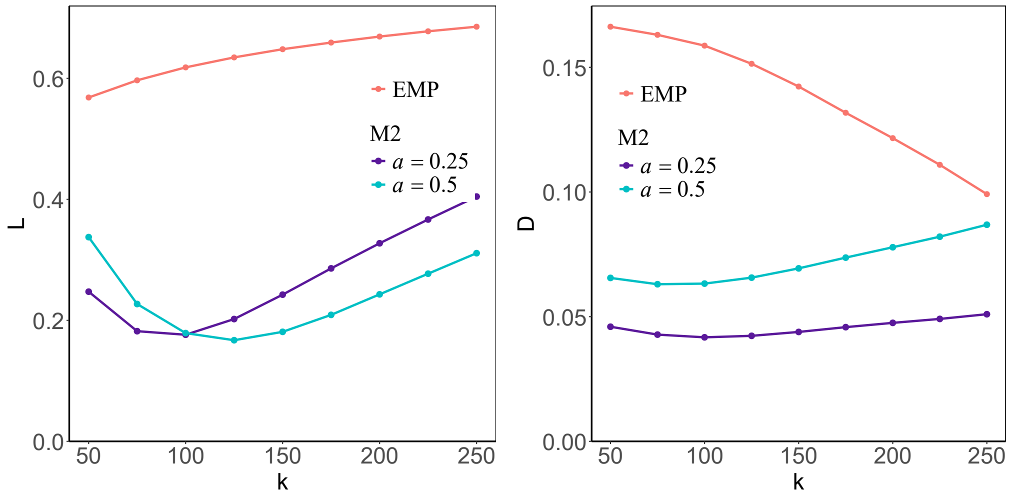


Fig. 6 The mean L (left) and D value (right) based on the second-order moment estimator $\hat{\gamma}_{n,ij}^{\mathrm{M},(2)}(k, -\log a)$ (M2) with $a = 0.25, 0.5$ and the empirical estimator $\hat{\gamma}_{n,ij}^{\mathrm{EMP}}$ (EMP) over 300 replications, with random samples of size $n = 1000$ drawn from the 5-dimensional Hüsler–Reiss distribution with variogram matrix $\mathbf{\Gamma}_2$.

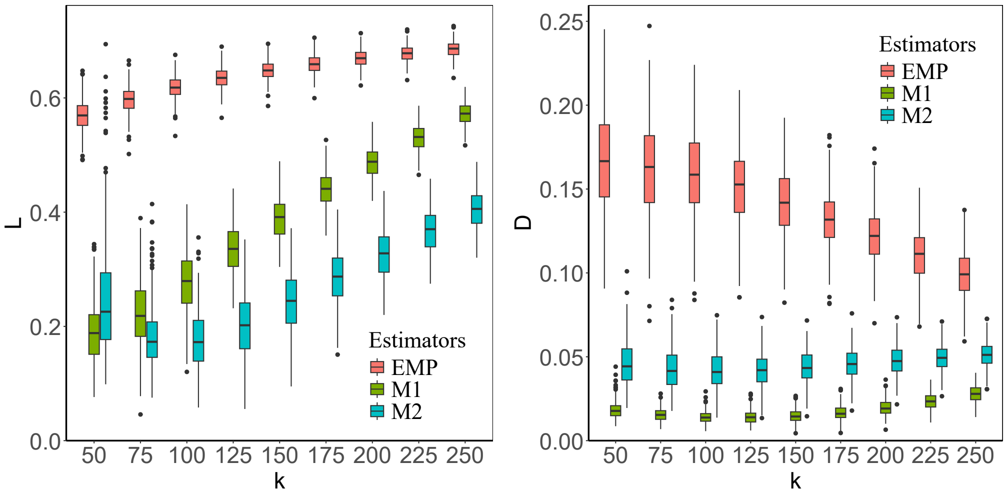


Fig. 7 The L value (left) and distance D of all bivariate tail dependence coefficients (right) based on the empirical variogram estimator $\hat{\gamma}_{n,ij}^{\mathrm{EMP}}$ (EMP), first-order moment estimator $\hat{\gamma}_{n,ij}^{\mathrm{M},(1)}(k,-\log a)$ (M1) and second-order moment estimator $\hat{\gamma}_{n,ij}^{\mathrm{M},(2)}(k,-\log a)$ (M2) with $a = 0.25$ in 300 replications. The random samples with size $n = 1000$ are drawn from the 5-dimensional Hüsler–Reiss distribution with variogram matrix $\mathbf{\Gamma}_2$.

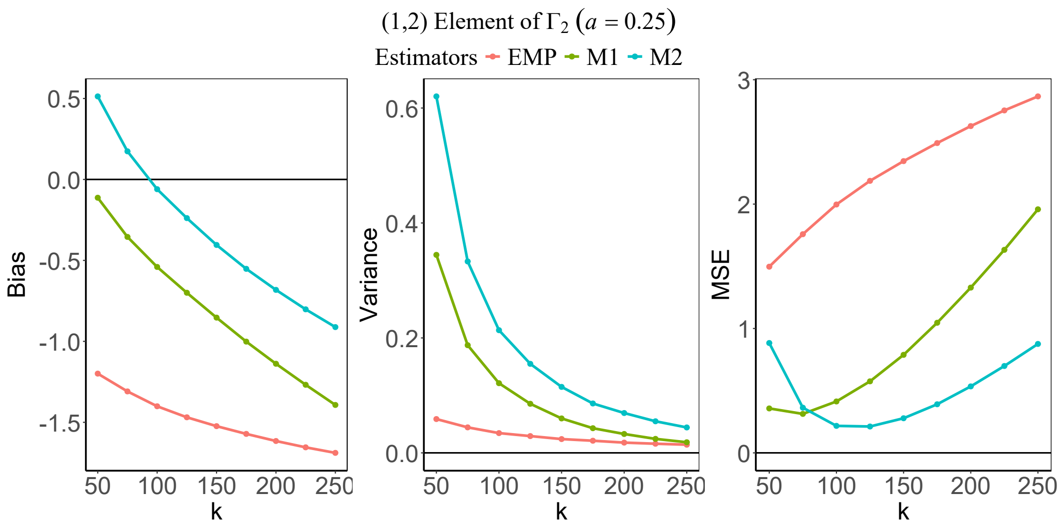


Fig. 8 The mean bias, variance and MSE for $\gamma_{12} = 3$, based on the empirical variogram estimator $\hat{\gamma}_{n,ij}^{\mathrm{EMP}}$ (EMP), first-order moment estimator $\hat{\gamma}_{n,ij}^{\mathrm{M},(1)}(k,-\log a)$ (M1) and second-order moment estimator $\hat{\gamma}_{n,ij}^{\mathrm{M},(2)}(k,-\log a)$ (M2) with $a = 0.25$ in 300 replications. The random samples with size $n = 1000$ are drawn from the 5–dimensional Hüsler–Reiss distribution with variogram matrix $\mathbf{\Gamma}_2$.

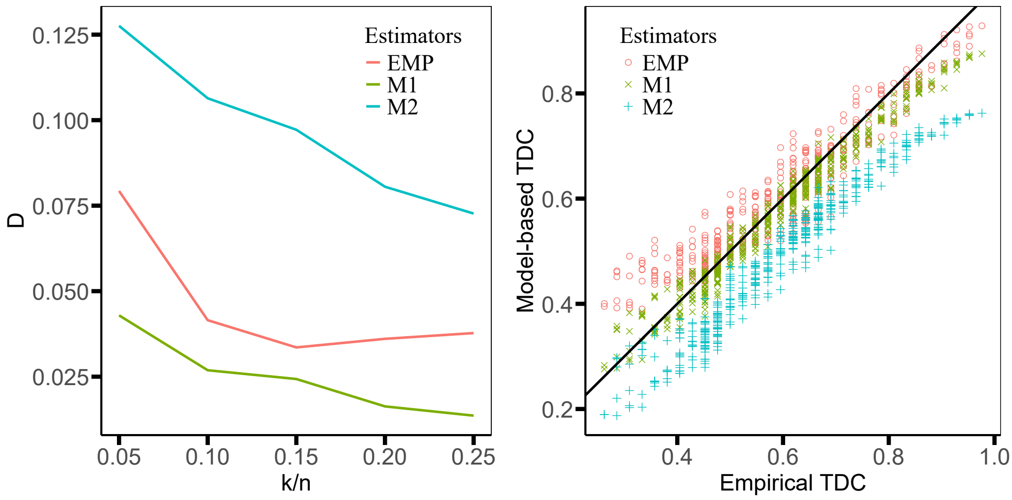


Fig. 9 The D values (left) and scatter plot (right) of empirical and model-based tail dependence coefficients (TDC) for the Danube discharges data, computed using the empirical variogram estimator $\hat{\gamma}_{n,ij}^{\mathrm{EMP}}$ (EMP), the moment estimator $\hat{\gamma}_{n,ij}^{\mathrm{M},(1)}$ (M1) and $\hat{\gamma}_{n,ij}^{\mathrm{M},(2)}$ (M2), with clipping level $a = 0.25$, ratio $k/n \in \{0.05, 0.10, 0.15, 0.20, 0.25\}$ for the line plot and $k/n = 0.10$ for the scatter plot.

set $k/n \in \{0.05, 0.10, 0.15, 0.20, 0.25\}$. The absolute distance D between the model-based and empirical tail dependence coefficients in (5.1) as a function of k/n is plotted in the left panel of Figure 9, where $a = \exp(-c)$ is fixed to be 0.25. It can be seen from the plot that, compared to the empirical variogram estimator, the absolute distance D for the first moment estimators is less sensitive to the choice of k and has consistently smaller D values than that of the empirical variogram estimator $\hat{\gamma}_{n,ij}^{\mathrm{EMP}}$.

To further assess the performance of the two estimation methods, we take $k/n = 0.10$, and present the scatter plot of $(\hat{\lambda}_{ij}, \hat{\lambda}_{ij}^{\mathrm{EST}})$ for $i, j \in V$ and $i \neq j$, where $\lambda_{ij}^{\mathrm{EST}}$ is the model-based tail dependence coefficient with the parameter matrix estimated by the moment estimators $\hat{\gamma}_{n,ij}^{\mathrm{M},(1)}(k, -\log a)$ and $\hat{\gamma}_{n,ij}^{\mathrm{M},(2)}(k, -\log a)$ with $a = 0.25$, as well as the empirical variogram estimator $\hat{\gamma}_{n,ij}^{\mathrm{EMP}}$. Figure 9 shows that the models based on the proposed estimators $\hat{\gamma}_{n,ij}^{\mathrm{M},(1)}$ provide a better fit to the empirical tail dependence coefficients than those of the other two estimators.

6.2 US flight delay data

In this subsection, we apply the proposed estimators to the U.S. flight delay dataset, which is a widely used benchmark in the analysis of multivariate extremes (see, e.g., Hentschel et al., 2025; Kiriliouk et al., 2025). The dataset contains records of domestic flights in the United States operated by major carriers (airlines with at least 1% market share) and involving airports that account for at least 1% of domestic enplanements. Hentschel et al. (2025) preprocessed the dataset by restricting attention to airports located in the contiguous United States with at least 1000 flights per year. For each airport, daily accumulated positive flight delays (in minutes) are computed by summing arrival and departure delays. This results in a dataset consisting of 5601 observations of daily accumulated flight delays over a 16−year period across 170 airports. Based on this dataset, they further apply a K-medoids clustering approach using a tail dependence coefficient distance matrix at probability level $p = 0.85$, yielding six clusters. The original data are publicly available from the U.S. Bureau of Transportation Statistics. Preprocessed versions of the dataset can be obtained from the GitHub repository of Manuel Hentschel or through the `R` package `graphicalExtremes` (Hentschel et al., 2025).

Here, we focus on the Texas cluster for model comparison purposes. This cluster consists of $d = 29$ airports and contains $n = 3603$ observation days over the period 2005–2020. We model the resulting data using the Hüsler–Reiss model to capture tail dependence among extreme flight delays. The variogram matrix is estimated by the proposed moment estimators and the empirical variogram estimator.

The D values based on the clipped moments estimator and the empirical variogram estimator, as a function of $k/n \in \{0.05, 0.10, 0.15, 0.20, 0.25\}$, are presented in the left panel of Figure 10. The right panel shows scatter plots comparing the model-based tail dependence coefficients obtained from the three estimators with the empirical tail dependence coefficients, where $k/n = 0.1$ and $a = 0.25$. We observe that, in this setting, the clipped moment estimators exhibit greater stability with respect to the choice of k. In particular, the improvement of the first-order clipped moment estimator $\hat{\gamma}_{n,ij}^{\mathrm{M},(1)}(k, -\log a)$ is more pronounced. Compared with the Danube dataset, the dependence structure in this dataset appears to be weaker. This leads to a more evident improvement of the first-order clipped moment estimator.

7 Concluding remarks

The empirical variogram estimator of Engelke and Volgushev (2022) is the natural estimator of the variogram matrix of a Hüsler–Reiss MGPD and is widely used in practice. It relies, however, on the limiting MGPD being an accurate approximation to the joint tail of the data. When the tail dependence between some of the components is weak, this approximation deteriorates: an observation that is extreme in some of the components is then likely to be non-extreme in the weakly dependent ones, and the resulting bias can be substantial. Motivated by this, we have proposed first- and second-order moment estimators of the elements of the variogram matrix, constructed from a lower-tail-clipped version of the standardized variables, and we have established their consistency and asymptotic normality.

The simulation study and the two case studies point in the same direction. The clipped moment estimators trade variance for bias: the empirical variogram estimator attains the smaller variance throughout, but the larger bias, and the moment estimators are the more accurate ones precisely in the regime that motivated them, that is, when the dependence is weak. Among the estimators considered, the first-order moment estimator with clipping level $a = 0.25$ comes out best overall, having the lowest error in almost all

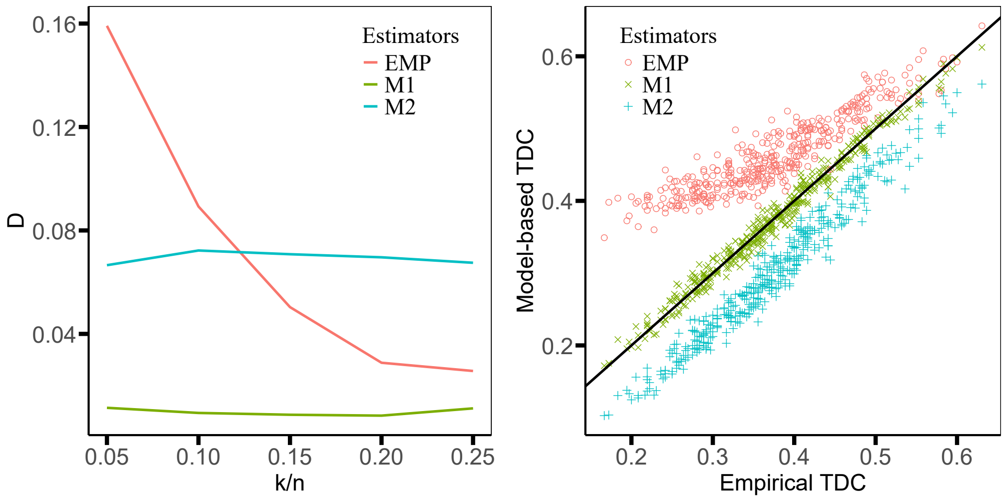


Fig. 10 The D values (left) and scatter plot (right) of empirical and model-based tail dependence coefficients (TDC) for the flight delay data, computed using the empirical variogram estimator $\hat{\gamma}_{n,ij}^{\mathrm{EMP}}$ (EMP), the moment estimator $\hat{\gamma}_{n,ij}^{\mathrm{M},(1)}$ (M1) and $\hat{\gamma}_{n,ij}^{\mathrm{M},(2)}$ (M2), with clipping level $a = 0.25$, ratio $k/n \in \{0.05, 0.10, 0.15, 0.20, 0.25\}$ for the line plot and $k/n = 0.10$ for the scatter plot.

cases in terms of both L and D. Its advantage is most pronounced for small k, and both moment estimators are less sensitive to the choice of k than the empirical variogram estimator. The applications confirm the picture: for the US flight delay data, whose dependence structure is weaker than that of the Danube discharge data, the improvement brought by the first-order clipped moment estimator is the more evident of the two.

The choice of the moment order ℓ and of the clipping level c is left open. It is tempting to select the two by minimizing the asymptotic variance of Theorem 4.8, for which Section A.4 provides an explicit expression. This route, however, is not viable: the asymptotic variance is minimal for $\ell = 1$ and in the limit $c \to \infty$, that is, $a \to 0$, which amounts to no clipping at all. The purpose of clipping is to reduce the bias, so any meaningful choice of ℓ and c must weigh bias against variance. We have no theoretical expression for the bias of the moment estimators, nor a way of estimating it, and the trade-off is therefore out of reach. This is why we have fixed $a = 0.25$ throughout, guided by the simulation study rather than by theory. A tractable handle on the bias appears to us to be the key to any principled, data-driven choice of ℓ and c.

Finally, one may ask whether removing the clipping recovers the empirical variogram estimator. It does not. The moment functions $e^{(\ell)}(\gamma, c)$ in (3.4) and (3.5) diverge as $c \to \infty$, on account of the shift by c, so that the limit cannot be taken directly in (3.6). The clipped variable itself does converge, $Y_i^{(j)} \vee (-c) \to Y_i^{(j)}$, and its first two moments are

$$\mathbb{E}\left(Y_i^{(j)}\right) = 1 - \frac{\gamma_{ij}}{2}, \qquad \mathbb{E}\left\{\left(Y_i^{(j)}\right)^2\right\} = \frac{1}{4}\gamma_{ij}^2 + 2.$$

Equating these to their empirical counterparts yields moment estimators of γ_{ij} based on the unclipped variables. These are not the empirical variogram estimator: the latter is built on the variance of the differences $Y_i^{(m)} - Y_j^{(m)}$, whereas the former rest on the moments of $Y_i^{(j)}$ alone. The clipping level thus does not interpolate between the moment estimators proposed here and the empirical variogram estimator; the two constructions remain distinct.

Statements and Declarations

Competing interests. The authors declare no competing interests.

Author contributions. Both authors contributed equally to the ideas and to the theory. The simulation study and the case study are largely the work of Shuang Hu.

Data availability. Both datasets analyzed in this paper are publicly available. The Danube discharge data are available in the supplementary material of Asadi et al. (2015). The US flight delay data are publicly available from the U.S. Bureau of Transportation Statistics; the preprocessed version used here can be obtained through the `R` package `graphicalExtremes` (Hentschel et al., 2025).

The code used for the simulation studies and data analysis is available at `https://github.com/HUSHHuShuang/HR_variogram_estimation_by_clipped_moments`.

References

Asadi, P., Davison, A.C., Engelke, S.: Extremes on river networks. Annals of Applied Statistics **9**(4), 2023–2050 (2015)

Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J.L.: Statistics of Extremes: Theory and Applications. John Wiley & Sons, Chichester (2004)

Chiapino, M., Clémençon, S., Feuillard, V., Sabourin, A.: A multivariate extreme value theory approach to anomaly clustering and visualization. Computational Statistics **35**(2), 607–628 (2020)

Cai, J.-J., Einmahl, J.H., Haan, L., Zhou, C.: Estimation of the marginal expected shortfall: the mean when a related variable is extreme. Journal of the Royal Statistical Society Series B: Statistical Methodology **77**(2), 417–442 (2015)

Chiapino, M., Sabourin, A., Segers, J.: Identifying groups of variables with the potential of being large simultaneously. Extremes **22**(2), 193–222 (2019)

Davison, A.C., Padoan, S.A., Ribatet, M.: Statistical modeling of spatial extremes. Statistical Science **27**(2), 161–186 (2012)

Einmahl, J.H.J., Haan, L., Li, D.: Weighted approximations of tail copula processes with application to testing the bivariate extreme value condition. The Annals of Statistics **34**(4), 1987–2014 (2006)

Engelke, S., Hitz, A.S.: Graphical models for extremes. Journal of the Royal Statistical Society Series B: Statistical Methodology **82**(4), 871–932 (2020)

Einmahl, J.H., Krajina, A., Segers, J.: An M-estimator for tail dependence in arbitrary dimensions. The Annals of Statistics **105**(489), 1764–1793 (2012)

Einmahl, J.H., Kiriliouk, A., Segers, J.: A continuous updating weighted least squares estimator of tail dependence in high dimensions. Extremes **21**, 205–233 (2018)

Engelke, S., Lalancette, M., Volgushev, S.: Learning extremal graphical structures in high dimensions. The Annals of Statistics **54**(3), 1205–1231 (2026)

Engelke, S., Malinowski, A., Kabluchko, Z., Schlather, M.: Estimation of Hüsler–Reiss distributions and Brown–Resnick processes. Journal of the Royal Statistical Society Series B: Statistical Methodology **77**(1), 239–265 (2015)

Einmahl, J.H., Segers, J.: Empirical tail copulas for functional data. The Annals of Statistics **49**(5), 2672–2696 (2021)

Engelke, S., Volgushev, S.: Structure learning for extremal tree models. Journal of the Royal Statistical Society Series B: Statistical Methodology **84**(5), 2055–2087 (2022)

Frick, M., Reiss, R.-D.: Limiting distributions of maxima under triangular schemes. Journal of Multivariate Analysis **101**(10), 2346–2357 (2010)

Huser, R., Davison, A.C.: Composite likelihood estimation for the Brown–Resnick process. Biometrika **100**(2), 511–518 (2013)

Hanson, T.E., Carvalho, M., Chen, Y.: Bernstein polynomial angular densities of multivariate extreme value distributions. Statistics & Probability Letters **128**, 60–66 (2017)

Hentschel, M., Engelke, S., Segers, J.: Statistical inference for Hüsler–Reiss graphical models through matrix completions. Journal of the American Statistical Association **120**(550), 909–921 (2025)

Hashorva, E., Peng, Z., Weng, Z.: Higher-order expansions of distributions of maxima in a Hüsler-Reiss model. Methodology and Computing in Applied Probability **18**, 181–196 (2016)

Hüsler, J., Reiss, R.-D.: Maxima of normal random vectors: Between independence and complete dependence. Statistics & Probability Letters **7**(4), 283–286 (1989)

Kiriliouk, A., Lee, J., Segers, J.: X-vine models for multivariate extremes. Journal of the Royal Statistical Society Series B: Statistical Methodology **87**(3), 579–602 (2025)

Liao, X., Peng, Z.: Convergence rate of maxima of bivariate Gaussian arrays to the Hüsler-Reiss distribution. Statistics and Its Interface **7**(3), 351–362 (2014)

Mourahib, A., Kiriliouk, A., Segers, J.: Multivariate generalized Pareto distributions along extreme directions. Extremes **28**, 239–272 (2025)

Naveau, P., Segers, J.: Multivariate extreme value theory. arXiv preprint arXiv:2412.18477 (2024)

Orey, S., Pruitt, W.E.: Sample functions of the n-parameter Wiener process. The Annals of Probability **1**(1), 138–163 (1973)

Padoan, S.A., Ribatet, M., Sisson, S.A.: Likelihood-based inference for max-stable processes. Journal of the American Statistical Association **105**(489), 263–277 (2010)

Resnick, S.I.: Extreme Values, Regular Variation, and Point Processes. Springer, New York (2008)

Smirnov, N.V.: Limit distributions for the terms of a variational series. Trudy Matematicheskogo Instituta imeni VA Steklova **25**, 3–60 (1949)

Schmidt, R., Stadtmüller, U.: Non-parametric estimation of tail dependence. Scandinavian Journal of Statistics **33**(2), 307–335 (2006)

Shorack, G.R., Wellner, J.A.: Empirical Processes with Applications to Statistics. Wiley, New York (1986)

Vaart, A.W., Wellner, J.A.: Weak Convergence and Empirical Processes. With Applications to Statistics. Springer, New York (1996)

A Lemmas and proofs of theoretical results

In this section, we give the proofs of all the propositions, theorems and lemmas. We first introduce some notation that will be used in the following proofs.

Recall that for the random vector $\boldsymbol{X} = (X_i, i \in V)$ with continuous marginal distribution functions $F_i(x)$, $i \in V$, the random vector $\boldsymbol{U} = (U_i, i \in v)$ is defined by $U_i = 1 - F_i(X_i)$ for $V = \{1, \dots, d\}$. For $i \in V$ and $t = 1, \dots, n$, let

$$\boldsymbol{U}_t = (U_{t1}, \dots, U_{td}),\ U_{ti} = 1 - F_i(X_{ti}).$$

For non-empty $I \subset V$, denote

$$\begin{aligned} T_{n,I}(\boldsymbol{x}_I) &= \frac{1}{k}\sum_{t=1}^{n} \mathbb{1}\left\{U_{ti} \le kx_i/n,\ i \in I\right\}, \\ R_{n,I}(\boldsymbol{x}_I) &= \frac{n}{k} C_I\left(\frac{kx_i}{n},\ i \in I\right). \end{aligned}$$

Define the tail process on $[0,\infty]^{|I|} \smallsetminus \{(\infty,\dots,\infty)\}$ as

$$\nu_{n,I}(\boldsymbol{x}_I) = \sqrt{k}\left(T_{n,I}(\boldsymbol{x}_I) - R_{n,I}(\boldsymbol{x}_I)\right). \tag{A.1}$$

Denote the empirical distribution function of U_{ti}, $t = 1, \dots, n$ by

$$\hat{F}_{U,i}(x) = \sum_{t=1}^{n} \mathbb{1}\{U_{ti} \le x\}/(n+1).$$

For every $s \in \{1, \dots, n\}$, we have

$$\begin{aligned} \hat{F}_{U,i}(U_{si}) &= \frac{1}{n+1}\sum_{t=1}^{n} \mathbb{1}\{1 - F_i(X_{ti}) \le 1 - F_i(X_{si})\} \\ &= 1 - \frac{1}{n+1}\sum_{t=1}^{n} \mathbb{1}\{X_{ti} \le X_{si}\} = 1 - \hat{F}_i(X_{si}) \end{aligned}$$

almost surely. Hence $\hat{U}_{ti} = \hat{F}_{U,i}(U_{ti})$. Let

$$\hat{F}_{U,i}^{\leftarrow}(x) = \inf\{y \in \mathbb{R} : \hat{F}_{U,i}(y) \ge x\} \tag{A.2}$$

denote the generalized inverse function of $\hat{F}_{U,i}$. For $I \subset V$ and $\boldsymbol{x}_I \in [0,\infty]^{|I|} \smallsetminus (\infty,\dots,\infty)$, set

$$\tilde{R}_{n,I}(\boldsymbol{x}_I) = \frac{1}{k}\sum_{t=1}^{n} \mathbb{1}\left\{\hat{U}_{ti} \le kx_i/n, i \in I\right\} \tag{A.3}$$

and

$$\hat{R}_{n,I}(\boldsymbol{x}_I) = \frac{1}{k}\sum_{t=1}^{n} \mathbb{1}\left\{U_{ti} \le \hat{F}_{U,i}^{\leftarrow}(kx_i/n), i \in I\right\}. \tag{A.4}$$

For a subset $I \subset V$, suppose that $1/k \le x_i \le n/k$ for all $i \in I$. Let

$$\hat{\nu}_{n,I}(\boldsymbol{x}_I) = \sqrt{k}\left(\hat{R}_{n,I}(\boldsymbol{x}_I) - R_I(\boldsymbol{x}_I)\right) \tag{A.5}$$

with $\hat{R}_{n,I}(\boldsymbol{x}_I)$ and $R_I(\boldsymbol{x}_I)$ given by (A.4) and (4.1).

A.1 Lemmas

We first present some lemmas that will be used in the proofs of the propositions and the theorems in this section.

Lemma A.1 (Existence of clipped moments for MGPD). *Let $\boldsymbol{Y} = (Y_v, v \in V)$ be a standard MGP distributed random vector, and c be a real number that $0 \le c < \infty$. For any $i, m \in V$, we have $\mathbb{E}\left\{(Y_i^{(m)} + c)_+^2\right\} < \infty$.*

Proof. Assume $\tilde{\boldsymbol{X}} = (\tilde{X}_i, i \in V)$ is a random vector with standard exponential margins and it belongs to the domain of attraction of $\boldsymbol{Y}$. Note that

$$\mathbb{P}(\tilde{X}_i > y + \tilde{X}_m \mid \tilde{X}_m > u) = \frac{\mathbb{P}(\tilde{X}_i > y + \tilde{X}_m, \tilde{X}_m > u)}{\mathbb{P}(\tilde{X}_m > u)} \le \frac{\mathbb{P}(\tilde{X}_i > y + u)}{\mathbb{P}(\tilde{X}_m > u)} \le \exp(-y)$$

by the standard exponential margins. Hence from (2.1) and by letting $u \to \infty$ on the left-hand side of the last inequality, we have

$$\mathbb{P}(Y_i^{(m)} > y) < \exp(-y), \ y > 0.$$

Therefore,

$$\begin{aligned}
&\mathbb{E}\left[\left\{Y_i^{(m)} \vee (-c)\right\}^2\right] \\
&= \mathbb{E}\left[\left\{Y_i^{(m)} \vee (-c)\right\}^2 \mathbb{1}\{Y_i^{(m)} < 0\}\right] + \mathbb{E}\left[\left\{Y_i^{(m)} \vee (-c)\right\}^2 \mathbb{1}\{Y_i^{(m)} \ge 0\}\right] \\
&\le c^2 + \int_0^\infty \mathbb{P}\left\{(Y_i^{(m)})^2 \, \mathbb{1}\{Y_i^{(m)} \ge 0\} > u\right\} \mathrm{d}\, u \\
&= c^2 + 2\int_0^\infty u \, \mathbb{P}(Y_i^{(m)} > u) \, \mathrm{d}\, u \\
&\le c^2 + 2\int_0^\infty u \exp(-u) \, \mathrm{d}\, u = c^2 + 2 < \infty.
\end{aligned}$$

The desired result follows from the decomposition that

$$\begin{aligned}
\mathbb{E}\left\{(Y_i^{(m)} + c)_+^2\right\} &= \mathbb{E}\left[\left\{(Y_i^{(m)} \vee (-c)) + c\right\}^2\right] \\
&= \mathbb{E}\left\{(Y_i^{(m)} \vee (-c))^2\right\} + 2c\,\mathbb{E}\left\{Y_i^{(m)} \vee (-c)\right\} + c^2 < \infty.
\end{aligned}$$

The proof is complete. □

To establish the asymptotic normality of the moment estimators, we need a convergence result for the weighted tail empirical process. A similar result has already been discussed in Proposition 3.1 of Einmahl et al. (2006). The difference from their work is that we use a different weighting function $(x \wedge y)^\eta$, while they weighted the process by $(x \vee y)^\eta$. However, our result can be derived through their proof approach with minor modifications. Another related result is Lemma 1 from Cai et al. (2015). In that lemma, the weight function of the bivariate empirical tail process is x^η. In our case, the weight function $(x \wedge y)^\eta$ can be decomposed into two cases: $x > y$ and $x \le y$. By analyzing these cases separately, our setting can be reduced to the one considered in the cited lemma, and hence, similar conclusions can be derived. However, the result in their work was stated without a formal proof. For completeness, we present the proof in the following lemma. We note that this lemma applies to general MGPDs and is not restricted to the Hüsler–Reiss MGPD.

Lemma A.2. *Let $\boldsymbol{Y}$ be an arbitrary MGP distributed random vector, and let $\boldsymbol{X}_t = (X_{t1}, \dots, X_{td})$, $t = 1, \dots, n$, denote the independent copies of a random vector $\boldsymbol{X}$, which has continuous margins and such that the marginally transformed vector $\tilde{\boldsymbol{X}}$ in* (2.4) *lies in the domain of attraction of $\boldsymbol{Y}$. If Assumption 4.1 holds, then for $T > 0$ and $0 \le \eta < 1/2$, the process*

$$\left(\frac{\nu_{n,ij}(x,y)}{(x \wedge y)^\eta}, (x,y) \in (0,T]^2, \ \frac{\nu_{n,s}(u)}{u^\eta}, u \in (0,T], \ i,j \in V, \ i < j, \ s \in V, \right)$$

converges jointly in distribution to

$$\left(\frac{W_{ij}(x,y)}{(x \wedge y)^\eta}, (x,y) \in (0,T]^2, \ \frac{W_s(u)}{u^\eta}, u \in (0,T], \ i,j \in V, \ i < j, \ s \in V \right)$$

as $n \to \infty$ in $(\ell^\infty((0,T]^2))^{d(d-1)/2} \times (\ell^\infty((0,T]))^d$.

Proof. For nonempty subset $I \subset V$ with $|I| \le 2$ and $\boldsymbol{x}_I = (x_i, i \in I) \in (0,T]^{|I|}$, define

$$f_I(\boldsymbol{x}_I) := f_I(\boldsymbol{x}_I, \cdot) = \frac{\mathbb{1}_{(\boldsymbol{0}, \boldsymbol{x}_I]}(\cdot)}{(\min_{i \in I} x_i)^\eta}$$

as a function on $[0,\infty]^d \setminus \{(\infty, \dots, \infty)\}$, where $(\boldsymbol{0}, \boldsymbol{x}]$ denotes the Cartesian product $\prod_{i \in V}(0, y_i]$ with $y_i = x_i$ if $i \in I$ and $y_i = \infty$ if $i \in V \setminus I$. That is, for any $\boldsymbol{z} \in [0,\infty]^d \setminus \{(\infty, \dots, \infty)\}$,

$$f_I(\boldsymbol{x}_I)(\boldsymbol{z}) = f_I(\boldsymbol{x}_I, \boldsymbol{z}) = \frac{\mathbb{1}\{z_i \le x_i, i \in I\}}{(\min_{i \in I} x_i)^\eta}.$$

Denote the class of $f_I(\boldsymbol{x}_I)$ by $\mathcal{F}$, i.e.,

$$\mathcal{F} = \{ f_I(\boldsymbol{x}_I), \ I \subset V, \ 1 \le |I| \le 2, \ \boldsymbol{x}_I = (x_i, i \in I) \in (0,T]^{|I|} \}.$$

For $I, J \subset V$ and $f_I(\boldsymbol{x}_I), f_J(\boldsymbol{y}_J) \in \mathcal{F}$, define a semi-metric ρ on $\mathcal{F}$ by

$$\rho(f_I(\boldsymbol{x}_I), f_J(\boldsymbol{y}_J)) = \left[\mathbb{E} \left\{ \left(\frac{W_I(\boldsymbol{x}_I)}{(\min_{i \in I} x_i)^\eta} - \frac{W_J(\boldsymbol{y}_J)}{(\min_{i \in J} y_i)^\eta} \right)^2 \right\} \right]^{1/2}.$$

Let

$$Z_{n,t} = \frac{1}{\sqrt{k}} \delta_{\left(\frac{n}{k} U_{ti}, \, i \in V\right)}$$

be the scaled Dirac measure on $[0,\infty]^d \setminus \{(\infty,\dots,\infty)\}$. For any $f \in \mathcal{F}$, let

$$Z_{n,t}(f) = \int f \,\mathrm{d}\, Z_{n,t}.$$

Then we have

$$Z_{n,t}(f_I(\boldsymbol{x}_I)) = \frac{1}{\sqrt{k}} \frac{\mathbb{1}\{U_{ti} < kx_i/n,\ i \in I\}}{(\min_{i\in I} x_i)^\eta}$$

and

$$\frac{\nu_{n,I}(\boldsymbol{x}_I)}{(\min_{i\in I} x_i)^\eta} = \sum_{t=1}^n \{Z_{n,t}(f_I(\boldsymbol{x}_I)) - \mathbb{E}\left(Z_{n,t}(f_I(\boldsymbol{x}_I))\right)\}.$$

Note that for each t, $Z_{n,t}$ can be viewed as a stochastic process indexed by functions $f_I(\boldsymbol{x}_I) \in \mathcal{F}$. For independent stochastic processes $Z_{n,t}$, $t = 1,\dots,n$, indexed by a totally bounded semi-metric space $(\mathcal{F},\rho)$, the weak convergence of $\sum_{t=1}^n (Z_{n,t} - \mathbb{E}\, Z_{n,t})$ is implied by the weak convergence of its finite-dimensional distributions and its asymptotic tightness. Note that the finite-dimensional convergence follows from the Cramér–Wold device together with the univariate Lindeberg–Feller central limit theorem, since convergence of all linear combinations implies convergence of multivariate distributions. Hence it suffices to establish asymptotic tightness. To this end, let $\{\mathcal{F}_{\varepsilon j}\}_{i=1}^{N_\varepsilon}$ be a partition of $\mathcal{F}$ such that $\mathcal{F} = \bigcup_{j=1}^{N_\varepsilon} \mathcal{F}_{\varepsilon j}$ and

$$\sum_{t=1}^n \mathbb{E}^* \left(\sup_{f,g\in\mathcal{F}_{\varepsilon j}} \left|Z_{n,t}(f) - Z_{n,t}(g)\right|^2 \right) \le \varepsilon^2 \tag{A.6}$$

for each $j = 1,,\dots,N_\varepsilon$, where $\mathbb{E}^*$ is the outer integral (for definition see Van der Vaart and Wellner (1996)). For $\varepsilon > 0$, let $N_{[\,]}(\varepsilon,\mathcal{F},\rho)$ be the minimal number of ε-brackets (with respect to ρ) required to cover $\mathcal{F}$. By Theorem 2.11.9 in Van der Vaart and Wellner (1996), it is sufficient to verify:

$$\sum_{t=1}^n \mathbb{E}^* \left[\|Z_{n,t}\|_{\mathcal{F}}\, \mathbb{1}(\|Z_{n,t}\|_{\mathcal{F}} > \lambda)\right] \to 0 \tag{A.7}$$

as $n \to \infty$ for every $\lambda > 0$, where $\|Z_{n,t}\|_{\mathcal{F}} = \sup_{f\in\mathcal{F}} |Z_{n,t}(f)|$, and

$$\int_0^{\delta_n} \sqrt{\log N_{[\,]}(\varepsilon,\mathcal{F},\rho)}\,\mathrm{d}\,\varepsilon \to 0 \tag{A.8}$$

for $\delta_n \to 0$. Together with (A.6), this implies the tightness of $\sum_{t=1}^n (Z_{n,t} - \mathbb{E}(Z_{n,t}))$.

To validate (A.7), we first show that $(\mathcal{F},\rho)$ is a bounded space. Given $I \subset V$, let $\mathcal{F}_I = \{f_I(\boldsymbol{x}_I), \boldsymbol{x}_I \in (0,T]^{|I|}\}$. Recall that f_I are defined for nonempty subsets $I \subset V$ and $|I| \le 2$. Since there are only finitely many such subsets, the totally boundedness of $(\mathcal{F},\rho)$ is equivalent to the totally boundedness of all the subclass $\mathcal{F}_I$, $I \subset V$ with $|I| \le 2$. For brevity, we only show that $\mathcal{F}_I$ is totally bounded for $I = \{i,j\}$ where $|I| = 2$. The remaining cases are similar.

For every $f_{ij}(x,y) \in \mathcal{F}_I$, we show that the mapping: $(x,y) \mapsto f_{ij}(x,y)$ is uniformly continuous from $(0,T]^2$ to $(\mathcal{F}_I,\rho)$. Since $(0,T]^2$ is totally bounded under the Euclidean metric, this will imply total boundedness of $(\mathcal{F}_I, \rho)$. For $x, y, u, v \in (0, T]$, there are four possible orderings of (x, y) and (u, v). Without loss of generality, let $x \geq y$, $u \geq v$, and $x \geq u$, $y \geq v$. For any $\delta > 0$, assuming $|x - u| \leq \delta$ and $|y - v| \leq \delta$. Then we have

$$\begin{aligned}
\rho^2(f_{ij}(x,y), f_{ij}(u,v)) =& \mathbb{E}\left\{\left(\frac{W_{ij}(x,y)}{(x \wedge y)^\eta} - \frac{W_{ij}(u,v)}{(u \wedge v)^\eta}\right)^2\right\} \\
=& \mathbb{E}\left\{\left(\frac{W_{ij}(x,y)}{y^\eta} - \frac{W_{ij}(u,v)}{v^\eta}\right)^2\right\} \\
=& \mathbb{E}\left\{\left(\frac{v^\eta W_{ij}(x,y) - y^\eta W_{ij}(u,v)}{(yv)^\eta}\right)^2\right\} \\
=& \frac{v^{2\eta} R_{ij}(x,y) - 2v^\eta y^\eta R_{ij}(u,v) + y^{2\eta} R_{ij}(u,v)}{(yv)^{2\eta}}.
\end{aligned}$$

Note that $R_{ij}(x,y) \leq x \wedge y$. If $v \leq \delta$, we have

$$\begin{aligned}
\rho^2(f_{ij}(x,y), f_{ij}(u,v)) \leq& \frac{R_{ij}(x,y)}{y^{2\eta}} + \frac{3R_{ij}(u,v)}{v^{2\eta}} \\
\leq& y^{1-2\eta} + 3v^{1-2\eta} \leq (2\delta)^{1-2\eta} + 3\delta^{1-2\eta} \leq 5\delta^{1-2\eta}.
\end{aligned}$$

Otherwise, for $v > \delta$, by noting that

$$|R_{ij}(x,y) - R_{ij}(u,v)| \leq |x - u| + |y - v| \leq 2\delta,$$

we obtain

$$\begin{aligned}
\rho^2(f_{ij}(x,y), f_{ij}(u,v)) &= \frac{R_{ij}(u,v)(y^\eta - v^\eta)^2}{(yv)^{2\eta}} + \frac{v^{2\eta}[R_{ij}(x,y) - R_{ij}(u,v)]}{(yv)^{2\eta}} \\
&\leq \frac{R_{ij}(u,v)(y^\eta - v^\eta)^2}{(yv)^{2\eta}} + \frac{2\delta v^{2\eta}}{(yv)^{2\eta}} \\
&\leq v^{1-4\eta}(y^\eta - v^\eta)^2 + 2\delta^{1-2\eta} \\
&\leq \eta^2 v^{1-4\eta} v^{2\eta-2}(y - v)^2 + 2\delta^{1-2\eta} \\
&\leq v^{-1-2\eta}(y - v)^2 + 2\delta^{1-2\eta} \leq 3\delta^{1-2\eta},
\end{aligned}$$

where the fourth step follows from the mean value theorem. Hence, for every $\varepsilon > 0$, there exists $\delta > 0$ such that for all $|x - u| \leq \delta$ and $|y - v| \leq \delta$, we have $\rho^2(f_{ij}(x,y), f_{ij}(u,v)) < \varepsilon$. This completes the proof for the case $x \geq y$, $u \geq v$, and $x \geq u$, $y \geq v$. The other three cases can be followed by symmetry. Consequently, the map $(x,y) \mapsto f_{ij}(x,y)$ is uniformly continuous from $(0,T]^2$ to $(\mathcal{F}_I, \rho)$. Since there are finitely many $I \subset V$ with $|I| \leq 2$, the whole class $(\mathcal{F}, \rho)$ is totally bounded.

Now we show that (A.7) holds for every $\lambda > 0$. For $I \subset V$ and $\boldsymbol{x}_I \in (0,T]^{|I|}$, assume $i_0 = \{i : x_i = \min_{i \in I} x_i\}$. Since

$$\sup_{f_I(\boldsymbol{x}_I) \in \mathcal{F}} |Z_{n,t}(f_I(\boldsymbol{x}_I))| = \sup_{f_I(\boldsymbol{x}_I) \in \mathcal{F}} \frac{\mathbb{1}\{U_{ti} < kx_i/n,\ i \in I\}}{\sqrt{k}(\min_{i \in I} x_i)^\eta}$$

$$\le \frac{1}{\sqrt{k}} \frac{\mathbb{1}(U_{ti_0} \le \frac{k}{n} x_{i_0})}{x_{i_0}^{\eta}} \le \frac{1}{\sqrt{k}} \frac{1}{\left(\frac{n}{k} U_{ti_0}\right)^{\eta}},$$

we have for each $\lambda > 0$ and $0 \le \eta < 1/2$,

$$\begin{aligned}
&\sum_{t=1}^{n} \mathbb{E}^* \left\{ \|Z_{n,t}\|_{\mathcal{F}} \, \mathbb{1}\left(\|Z_{n,t}\|_{\mathcal{F}} > \lambda\right) \right\} \\
&\quad \le \frac{n}{\sqrt{k}} \mathbb{E}\left[\left(\frac{n}{k} U_{ti_0}\right)^{-\eta} \mathbb{1}\left\{ \frac{1}{\sqrt{k}} \frac{1}{\left(\frac{n}{k} U_{ti_0}\right)^{\eta}} > \lambda \right\} \right] \\
&\quad = \frac{n}{\sqrt{k}} \left(\frac{n}{k}\right)^{-\eta} \int_0^{(k/n)(\sqrt{k}\lambda)^{-1/\eta}} x^{-\eta} \, \mathrm{d}x \\
&\quad = \frac{1}{1-\eta} \lambda^{1-1/\eta} k^{1-1/(2\eta)} \to 0
\end{aligned}$$

as $n \to \infty$. This establishes the validity of (A.7).

It remains to show that (A.8) is satisfied for every sequence $\delta_n \to 0$. Although the original goal is to establish the result on $[0, T]$, we follow Einmahl et al. (2006) and take $T = 1$ for simplicity. The general case follows analogously. Let $\mathcal{F}^j = \{f_I(\boldsymbol{x}_I), \, I \subset V, \, |I| = j\}$. It suffices to verify (A.8) for $j = 1, 2$. We only treat the case $j = 2$, the case $j = 1$ being similar. Fix $\varepsilon > 0$ sufficiently small. Set $\alpha = \varepsilon^{3/(1-2\eta)}$ and $\theta = 1 - \varepsilon^3$. Define

$$\mathcal{F}(\alpha) = \left\{ f_I(x, y) \in \mathcal{F}^2 : x \wedge y \le \alpha \right\},$$

and for intergers $r, s \ge 0$,

$$\mathcal{F}(r, s) = \left\{ f_I(x, y) \in \mathcal{F}^2 : \theta^{r+1} \le x \le \theta^r, \ \theta^{s+1} \le y \le \theta^s \right\}.$$

Then

$$\mathcal{F}^2 = \mathcal{F}(\alpha) \cup \left(\bigcup_{r,s=0}^{[\log\alpha/\log\theta]} \mathcal{F}(r, s) \right).$$

We first verify (A.8) for the class $\mathcal{F}(\alpha)$. By the definition of $Z_{n,t}$,

$$\begin{aligned}
&\sum_{t=1}^{n} \mathbb{E}^* \left\{ \sup_{f,g \in \mathcal{F}(\alpha)} \left(Z_{n,t}(f) - Z_{n,t}(g)\right)^2 \right\} = n \, \mathbb{E}\left\{ \sup_{f,g \in \mathcal{F}(\alpha)} \left(Z_{n,1}(f) - Z_{n,1}(g)\right)^2 \right\} \\
&\quad \le 4n \, \mathbb{E}\left\{ \sup_{f \in \mathcal{F}(\alpha)} Z_{n,1}^2(f) \right\} = \frac{4n}{k} \mathbb{E}\left\{ \sup_{\substack{x,y>0 \\ x\wedge y \le \alpha}} \frac{\mathbb{1}\left(U_{1i} < \frac{kx}{n}, U_{1j} < \frac{ky}{n}\right)}{(x \wedge y)^{2\eta}} \right\} \\
&\quad \le \frac{4n}{k} \mathbb{E}\left\{ \left(\frac{n}{k} U_{1i}\right)^{-2\eta} \mathbb{1}\left(\frac{n}{k} U_{1i} < \alpha\right) \right\} \\
&\quad = \frac{4n}{k} \int_0^{\alpha k/n} \left(\frac{n}{k} x\right)^{-2\eta} \mathrm{d}x = \frac{4\alpha^{1-2\eta}}{1-2\eta} \le \varepsilon^3.
\end{aligned}$$

Thus the contribution of $\mathcal{F}(\alpha)$ is negligible. Now we establish (A.8) for the classes $\{\mathcal{F}(r,s), 0 \le r,s \le [\log\alpha/\log\theta]\}$. For given r,s, without loss of generality, we assume $r = r \wedge s$. Then

$$
\begin{aligned}
&\sum_{t=1}^{n} \mathbb{E}^{\star}\left\{\sup_{f,g\in\mathcal{F}(r,s)} \left(Z_{n,t}(f) - Z_{n,t}(g)\right)^2\right\} \\
&\le n\,\mathbb{E}\left\{\left(\sup_{f\in\mathcal{F}(r,s)} Z_{n,1}(f) - \inf_{f\in\mathcal{F}(r,s)} Z_{n,1}(f)\right)^2\right\} \\
&\le \frac{n}{k}\,\mathbb{E}\left\{\left(\frac{\mathbb{1}\left\{U_{1i} \le \frac{k}{n}\theta^{r}, U_{1j} \le \frac{k}{n}\theta^{s}\right\}}{(\theta^{r+1}\wedge\theta^{s+1})^{\eta}} - \frac{\mathbb{1}\left\{U_{1i} \le \frac{k}{n}\theta^{r+1}, U_{1j} \le \frac{k}{n}\theta^{s+1}\right\}}{(\theta^{r}\wedge\theta^{s})^{\eta}}\right)\right\}^2 \\
&= \frac{n}{k}\,\mathbb{E}\Bigg[\left\{\mathbb{1}\left\{U_{1i} \le \frac{k}{n}\theta^{r}, U_{1j} \le \frac{k}{n}\theta^{s}\right\}\left(\theta^{-\eta(s+1)} - \theta^{-\eta s}\right)\right. \\
&\qquad \left. + \left\{\mathbb{1}\left\{U_{1i} \le \frac{k}{n}\theta^{r}, U_{1j} \le \frac{k}{n}\theta^{s}\right\} - \mathbb{1}\left\{U_{1i} \le \frac{k}{n}\theta^{r+1}, U_{1j} \le \frac{k}{n}\theta^{s+1}\right\}\right\}\theta^{-\eta s}\right\}^2\Bigg] \\
&\le \frac{2n}{k}\Bigg[C_{ij}\left(\frac{k}{n}\theta^{r}, \frac{k}{n}\theta^{s}\right)\frac{(\theta^{-\eta}-1)^2}{\theta^{2\eta s}} \\
&\qquad + \left\{C_{ij}\left(\frac{k}{n}\theta^{r}, \frac{k}{n}\theta^{s}\right) - C_{ij}\left(\frac{k}{n}\theta^{r+1}, \frac{k}{n}\theta^{s+1}\right)\right\}\theta^{-2\eta s}\Bigg] \\
&\le 2\theta^{(1-2\eta)s}\left\{(\theta^{-\eta}-1)^2 + 2(1-\theta)\right\} \le 2\left\{(\theta^{-1/2}-1)^2 + 2(1-\theta)\right\} \le \varepsilon^6 + 4\varepsilon^3 \le 5\varepsilon^3,
\end{aligned}
$$

where $(\theta^{-\eta}-1)^2 \le \varepsilon^6$ holds for sufficiently small ε. Thus, the number of elements of the partition of $\mathcal{F}^2$ is bounded by $C\varepsilon^{-6}\log(1/\varepsilon)^2$ for some constant $C > 0$. Consequently, (A.8) holds as $n \to \infty$. This completes the proof of the lemma. □

Recall that $W(\boldsymbol{x})$ is a mean-zero Gaussian process defined in Section 4 with continuous trajectories. Since the weighted empirical tail process is asymptotically tight, by Theorems 1.5.7 and 1.5.8 of Van der Vaart and Wellner (1996), the processes take values in a separable subspace of $(\ell^{\infty}((0,T]^2))^{d(d-1)/2} \times (\ell^{\infty}((0,T]))^{d}$. Hence, by the Skorokhod's representation theorem (cf. Theorem 1.10.4 in Van der Vaart and Wellner (1996)), there exists a probability space with processes $\nu^{*}_{n,ij}(x,y)$, $\nu^{*}_{n,s}(u)$ and $W^{*}_{ij}(x,y)$, $W^{*}_{s}(u)$ for all $i,j,s \in V$ and $i<j$ such that for $x,y,u \in (0,T]$,

$$
\begin{aligned}
\left(\nu^{*}_{n,ij}(x,y), \nu^{*}_{n,s}(u),\ i,j,s\in V, i<j\right) &\overset{d}{=} \left(\nu_{n,ij}(x,y), \nu_{n,s}(u),\ i,j,s\in V,\ i<j\right), \\
\left(W^{*}_{ij}(x,y), W^{*}_{s}(u),\ i,j,s\in V, i<j\right) &\overset{d}{=} \left(W_{ij}(x,y), W_{s}(u),\ i,j,s\in V,\ i<j\right),
\end{aligned}
$$

and the convergence in Lemma A.2 holds almost surely. Then, for $0 \le \eta < 1/2$, we have

$$
\sup_{0<x,y\le T} \frac{\left|\nu^{*}_{n,ij}(x,y) - W^{*}_{ij}(x,y)\right|}{(x\wedge y)^{\eta}} = o_p(1),
$$

$$\sup_{0<u\le T} \frac{\left|\nu_{n,s}^*(u) - W_s^*(u)\right|}{u^\eta} = o_p(1). \tag{A.9}$$

We shall work within this probability space while keeping the notation unchanged. Next, we establish an inequality analogous to Lemma 3.2 in Einmahl et al. (2006).

Lemma A.3. *Let $0 \le \eta < 1/2$. For any $i, j \in V$ and sufficiently small $\varepsilon \in (0,1)$, we have*

$$\mathbb{P}\left(\sup_{0<x\le\varepsilon} \frac{|W_{ij}(x,1)|}{x^\eta} \ge \lambda\right) \le 4\sum_{l=0}^{\infty} \exp\left(-\frac{\lambda^2}{2^{1+2\eta}} \frac{2^{l(1-2\eta)}}{\varepsilon^{1-2\eta}}\right).$$

Proof. Recall that for any $i, j \in V$, $\{W_{ij}(x,1),\, x \in [0,\infty]\}$ defined in (4.2) is a centered Gaussian process. We firstly show that, for $0 \le a < b$,

$$\mathbb{P}\left\{\sup_{x\in[a,b]} |W_{ij}(x,1)| \ge \lambda\right\} \le 2\,\mathbb{P}\left\{|W_{ij}(b,1)| \ge \lambda\right\}. \tag{A.10}$$

To prove the result, let $\mathcal{F}_x = \sigma\{W_{ij}(s,1) : 0 \le s \le x\}$. For $0 \le s \le x \le y$, we have

$$\begin{aligned}
&\mathrm{Cov}\{W_{ij}(y,1) - W_{ij}(x,1), W_{ij}(s,1)\}\\
&= \mathrm{Cov}\{W_{ij}(y,1), W_{ij}(s,1)\} - \mathrm{Cov}\{W_{ij}(x,1), W_{ij}(s,1)\}\\
&= R(y \wedge s, 1) - R(x \wedge s, 1)\\
&= R(s,1) - R(s,1) = 0.
\end{aligned}$$

Since the process is Gaussian, the increment $W_{ij}(y,1) - W_{ij}(x,1)$ is independent of $\mathcal{F}_x$. Consequently,

$$\mathbb{E}\{W_{ij}(y,1) \mid \mathcal{F}_x\} = W_{ij}(x,1),$$

which shows that $\{W_{ij}(x,1) : x \ge 0\}$ is a continuous martingale. Moreover, by the continuity of the covariance function $R(\cdot,1)$, $W_{ij}(\cdot,1)$ admits a continuous modification. From the Dambis–Dubins–Schwarz representation theorem, there exists a standard Brownian motion B such that

$$W_{ij}(x,1) = B_{R(x,1)}, \qquad x \ge 0.$$

Consequently,

$$\sup_{x\in[0,b]} |W_{ij}(x,1)| = \sup_{t\in[0,R(b,1)]} |B_t|.$$

Since $R(x,1)$ is an increasing function on $x \in [a,b]$, applying Lemma 1.2 of Orey and Pruitt (1973) yields

$$\mathbb{P}\left\{\sup_{x\in[a,b]} |W_{ij}(x,1)| \ge \lambda\right\} \le \mathbb{P}\left\{\sup_{x\in[0,b]} |W_{ij}(x,1)| \ge \lambda\right\} \le 2\,\mathbb{P}\{|W_{ij}(b,1)| \ge \lambda\}.$$

Next, we show the main result. For $l = 1, 2, \ldots$, define

$$\mathcal{A}_l = \left\{x : \frac{\varepsilon}{2^{l+1}} \le x \le \frac{\varepsilon}{2^l}\right\},$$

then using (A.10), we have

$$
\begin{aligned}
\mathbb{P}\left\{\sup_{0<x\le\varepsilon}\frac{|W_{ij}(x,1)|}{x^{\eta}}\ge\lambda\right\} &= \mathbb{P}\left\{\sup_{l\in\{0,1,\dots\}}\sup_{x\in\mathcal{A}_l}\frac{|W_{ij}(x,1)|}{x^{\eta}}\ge\lambda\right\}\\
&\le \sum_{l=0}^{\infty}\mathbb{P}\left\{\sup_{x\in\mathcal{A}_l}\frac{|W_{ij}(x,1)|}{x^{\eta}}\ge\lambda\right\}\\
&\le \sum_{l=0}^{\infty}\mathbb{P}\left\{\sup_{x\in\mathcal{A}_l}|W_{ij}(x,1)|\ge\lambda\left(\frac{\varepsilon}{2^{l+1}}\right)^{\eta}\right\}\\
&\le 2\sum_{l=0}^{\infty}\mathbb{P}\left\{\left|W_{ij}\left(\frac{\varepsilon}{2^{l}},1\right)\right|\ge\lambda\left(\frac{\varepsilon}{2^{l+1}}\right)^{\eta}\right\}\\
&\le 4\sum_{l=0}^{\infty}\exp\left(-\frac{\lambda^2}{2^{1+2\eta}}\frac{2^{l(1-2\eta)}}{\varepsilon^{1-2\eta}}\right), \qquad (A.11)
\end{aligned}
$$

by Mill's inequality and the assumption that $\varepsilon < 1$. Notice that $W_{ij}(\frac{\varepsilon}{2^l},1)$ follows a normal distribution with zero mean and variance $R(\frac{\varepsilon}{2^l},1)$. The last inequality in (A.11) can be justified as follows,

$$
\begin{aligned}
\mathbb{P}\left\{\left|W_{ij}\left(\frac{\varepsilon}{2^{l}},1\right)\right|\ge\lambda\left(\frac{\varepsilon}{2^{l+1}}\right)^{\eta}\right\} &= \mathbb{P}\left\{\frac{\left|W_{ij}\left(\frac{\varepsilon}{2^l},1\right)\right|}{\sqrt{R(\frac{\varepsilon}{2^l},1)}}\ge\frac{\lambda\left(\frac{\varepsilon}{2^{l+1}}\right)^{\eta}}{\sqrt{R(\frac{\varepsilon}{2^l},1)}}\right\}\\
&\le \mathbb{P}\left\{\frac{\left|W_{ij}\left(\frac{\varepsilon}{2^l},1\right)\right|}{\sqrt{R(\frac{\varepsilon}{2^l},1)}}\ge\frac{\lambda\left(\frac{\varepsilon}{2^{l+1}}\right)^{\eta}}{\sqrt{\frac{\varepsilon}{2^l}}}\right\}\\
&\le \frac{2}{\sqrt{2\pi}}\frac{2^{\eta}}{\lambda}\left(\frac{\varepsilon}{2^l}\right)^{1/2-\eta}\exp\left(-\frac{\lambda^2}{2^{1+2\eta}}\frac{2^{l(1-2\eta)}}{\varepsilon^{1-2\eta}}\right)
\end{aligned}
$$

by Mill's inequality. For a fixed $\lambda > 0$, taking ε sufficiently small such that the constant $\frac{1}{\sqrt{2\pi}}\frac{2^{\eta}}{\lambda}\left(\frac{\varepsilon}{2^l}\right)^{1/2-\eta}$ is less than one, this proves the last inequality in (A.11). □

To establish the asymptotic normality of the proposed moment estimators, we require the following lemma, which provides a uniform convergence result for $\hat{\nu}_{n,I}(\boldsymbol{x}_I)$ defined in (A.5). A closely related conclusion is the weak convergence of the weighted tail copula process associated with the stable tail dependence function, as established in Einmahl et al. (2006, Theorem 2.2). Although we draw on some ideas from that theorem, our result differs by establishing weak convergence of the integral of the weighted tail empirical process with different weight functions, which require additional technical conditions in our lemma.

Lemma A.4. *Suppose that Assumption 4.5 holds and that the tail copula $R(\boldsymbol{x})$ of $\boldsymbol{X}$ satisfies, for $i, j \in V$ with $i \neq j$:*

(i) The partial derivative functions $\dot{R}^i_{ij}(x,y)$ and $\dot{R}^j_{ij}(x,y)$ exist and are continuous on $(0,\infty)^2$.

(ii) For $0 < \eta < 1/2$ *and* $x \in (0, \exp(c)]$*, the partial derivative satisfies*

$$\sup_{x\in[1/k,\exp(c)]} x^{-\eta}\left|\dot{R}^j_{ij}(S_{ni}(x),\theta_{nj}) - \dot{R}^j_{ij}(x,1)\right| = o_{\mathbb{P}}(1)$$

with $\theta_{nj} \in (S_{nj}(1) \wedge 1,\ S_{nj}(1) \vee 1)$ *almost surely as* $n \to \infty$*, where* $S_{n,I}(\boldsymbol{x}_I) = (S_{ni}(x_i), i \in I)$ *with*

$$S_{ni}(x) := \frac{n}{k}\hat{F}^{\leftarrow}_{U,i}\left(\frac{kx}{n}\right),$$

and $\hat{F}^{\leftarrow}_{U,i}$ *given by* (A.2).

Then we have

$$\int_{1/k}^{\exp(c)} \frac{\hat{\nu}_{n,ij}(x,1)\ell(-\log x)^{\ell-1}}{x}\,\mathrm{d}\,x \xrightarrow{\mathbb{P}} \int_0^{\exp(c)} \frac{B_{ij}(x,1)\ell(-\log x)^{\ell-1}}{x}\,\mathrm{d}\,x$$

as $n \to \infty$ *for* $\ell = 1, 2$*, where* $B_{ij}(x,1)$ *is given in* (4.4).

Proof. Set $c = -\log a$ with $a \in (0,1]$. Recall that

$$\hat{\nu}_{n,I}(\boldsymbol{x}_I) = \sqrt{k}\left\{\hat{R}_{n,I}(\boldsymbol{x}_I) - R_I(\boldsymbol{x}_I)\right\},\ I \subset V,$$

with $\boldsymbol{x}_I \in [0,\infty]^{|I|} \smallsetminus \{(\infty,\dots,\infty)\}$. By the definition of $\nu_{n,I}$ in (A.1) and the equality that $\hat{R}_{n,I}(\boldsymbol{x}_I) = T_{n,I}(S_{n,I}(\boldsymbol{x}_I))$, we have

$$\begin{aligned}\hat{\nu}_{n,I}(\boldsymbol{x}_I) = \nu_{n,I}(S_{n,I}(\boldsymbol{x}_I)) &+ \sqrt{k}\{R_{n,I}(S_{n,I}(\boldsymbol{x}_I)) - R_I(S_{n,I}(\boldsymbol{x}_I))\}\\ &+ \sqrt{k}\{R_I(S_{n,I}(\boldsymbol{x}_I)) - R_I(\boldsymbol{x}_I)\}.\end{aligned}$$

Hence, for $i, j \in V$ and $i \neq j$,

$$\begin{aligned}&\int_{1/k}^{1/a} \frac{\{\hat{\nu}_{n,ij}(x,1) - B_{ij}(x,1)\}\,\ell(-\log x)^{\ell-1}}{x}\,\mathrm{d}\,x\\ &= \int_{1/k}^{1/a} \frac{\{\nu_{n,ij}(S_{ni}(x),S_{nj}(1)) - W_{ij}(x,1)\}\,\ell(-\log x)^{\ell-1}}{x}\,\mathrm{d}\,x\\ &\quad+ \int_{1/k}^{1/a} \frac{\sqrt{k}\{R_{n,ij}(S_{ni}(x),S_{nj}(1)) - R_{ij}(S_{ni}(x),S_{nj}(1))\}\,\ell(-\log x)^{\ell-1}}{x}\,\mathrm{d}\,x\\ &\quad+ \int_{1/k}^{1/a} \frac{1}{x}\Bigg[\sqrt{k}\{R_{ij}(S_{ni}(x),S_{nj}(1)) - R_{ij}(x,1)\}\\ &\qquad\qquad+ \left\{\dot{R}^i_{ij}(x,1)W_i(x) + \dot{R}^j_{ij}(x,1)W_j(1)\right\}\Bigg]\ell(-\log x)^{\ell-1}\,\mathrm{d}\,x. \qquad (A.12)\end{aligned}$$

We will show, in turn, that each term on the right-hand side of (A.12) converges to zero.

For the first part in (A.12), since for any $i \in V$, $S_{ni}(x)$ is a nondecreasing function of x and $S_{ni}(1/a) \xrightarrow{\mathbb{P}} 1/a$ as $n \to \infty$ (cf. Eq. (3.10) in Einmahl et al. (2006)), there exists a constant T_0 such that, with high probability, $0 < S_{ni}(x) \le T_0$ for all $x \in (0, 1/a]$. Hence for $0 \le \eta < 1/2$,

$$
\begin{aligned}
&\sup_{1/k \le x \le 1/a} \frac{\left|\nu_{n,ij}(S_{ni}(x), S_{nj}(1)) - W_{ij}(x,1)\right|}{x^\eta} \\
&\le \sup_{1/k \le x \le 1/a} \frac{\left|\nu_{n,ij}(S_{ni}(x), S_{nj}(1)) - W_{ij}(S_{ni}(x), S_{nj}(1))\right|}{x^\eta} \\
&\quad + \sup_{1/k \le x \le 1/a} \frac{\left|W_{ij}(S_{ni}(x), S_{nj}(1)) - W_{ij}(x,1)\right|}{x^\eta} \\
&\le \sup_{1/k \le x \le 1/a} \frac{\left|\nu_{n,ij}(S_{ni}(x), S_{nj}(1)) - W_{ij}(S_{ni}(x), S_{nj}(1))\right|}{(S_{ni}(x))^\eta} \times \sup_{1/k \le x \le 1/a} \left(\frac{S_{ni}(x)}{x}\right)^\eta \\
&\quad + \sup_{1/k \le x \le 1/a} \frac{\left|W_{ij}(S_{ni}(x), S_{nj}(1)) - W_{ij}(x,1)\right|}{x^\eta} \\
&\le \sup_{u \in (0, T_0]} \frac{\left|\nu_{n,ij}(u, S_{nj}(1)) - W_{ij}(u, S_{nj}(1))\right|}{u^\eta} \cdot \sup_{s \in (0, k/(na)]} \left(\frac{\hat{F}^{\leftarrow}_{U,i}(s)}{s}\right)^\eta \\
&\quad + \sup_{1/k \le x \le 1/a} \frac{\left|W_{ij}(S_{ni}(x), S_{nj}(1)) - W_{ij}(x,1)\right|}{x^\eta} \\
&=: D_{n11} \cdot D_{n12} + D_{n13},
\end{aligned}
$$

where the last inequality holds with high probability. Since $S_{nj}(1) \xrightarrow{\mathbb{P}} 1$ as $n \to \infty$, T_0 can be set sufficiently large such that with high probability, $S_{nj}(1) \le T_0$, and then

$$
\sup_{u \in (0, T_0]} \frac{\left|\nu_{n,ij}(u, S_{nj}(1)) - W_{ij}(u, S_{nj}(1)\right|}{u^\eta} \le \sup_{u \in (0, T_0]} \frac{\left|\nu_{n,ij}(u, S_{nj}(1)) - W_{ij}(u, S_{nj}(1))\right|}{(u \wedge S_{nj}(1))^\eta},
$$

implying $D_{n11} \xrightarrow{\mathbb{P}} 0$ as $n \to \infty$ by Lemma A.2, (A.9). With the fact that

$$
\sup_{s \ge 1/n} \frac{\hat{F}^{\leftarrow}_{U,i}(s)}{s} = O_{\mathbb{P}}(1), \tag{A.13}
$$

see Eq.(3.10) of Einmahl et al. (2006) or Shorack and Wellner (1986, Page 416), we have $D_{n11} D_{n12} \xrightarrow{\mathbb{P}} 0$ as $n \to \infty$. Furthermore, for $1/k < \varepsilon < 1$, we know that

$$
\begin{aligned}
D_{n13} \le &\sup_{\substack{x \in (0, 1/a] \\ x \ge \varepsilon}} \frac{\left|W_{ij}(S_{ni}(x), S_{nj}(1)) - W_{ij}(x,1)\right|}{x^\eta} \\
&+ \sup_{\substack{x \in (0, 1/a] \\ 1/k \le x < \varepsilon}} \frac{\left|W_{ij}(S_{ni}(x), S_{nj}(1)) - W_{ij}(x,1)\right|}{x^\eta}
\end{aligned}
$$

$$
\begin{aligned}
&\le \sup_{\substack{x\in(0,1/a]\\ x\ge\varepsilon}} \frac{\left|W_{ij}(S_{ni}(x), S_{nj}(1)) - W_{ij}(x,1)\right|}{\varepsilon^{\eta}} + \sup_{\substack{x\in(0,1/a]\\ 1/k\le x\le\varepsilon}} \frac{\left|W_{ij}(x,1)\right|}{x^{\eta}} \\
&\quad + \sup_{\substack{x\in(0,1/a]\\ 1/k\le x\le\varepsilon}} \frac{\left|W_{ij}(S_{ni}(x), S_{nj}(1))\right|}{S_{ni}(x)^{\eta}} \sup_{s\in(0,k/(na)]} \left(\frac{\hat{F}^{\leftarrow}_{U,i}(s)}{s}\right)^{\eta} \\
&=: D_{n14} + D_{n15} + D_{n16}.
\end{aligned}
$$

Since by Smirnov's lemma, see, e.g., Smirnov (1949, Lemma 2.2.3),

$$
\sup_{0<s<k/(na)} \frac{n}{k}\left|\hat{F}^{\leftarrow}_{U,i}(s) - s\right| \overset{a.s}{\to} 0 \tag{A.14}
$$

as $n\to\infty$, it follows from the uniform continuity of W_{ij} that $D_{n14}\to 0$ almost surely for any $\varepsilon > 0$. By Lemma A.3 and (A.13), for any $\delta > 0$, we also have $\mathbb{P}(D_{n15} > \delta) < \delta$ and $\mathbb{P}(D_{n16} > \delta) < \delta$ for sufficiently large n and small ε. Thus

$$
\begin{aligned}
&\left|\int_{1/k}^{1/a} \frac{\{\nu_{n,ij}(S_{ni}(x), S_{nj}(1)) - W_{ij}(x,1)\}\,\ell(-\log x)^{\ell-1}}{x}\,\mathrm{d}\,x\right| \\
&\le \sup_{1/k\le x\le 1/a} \frac{\left|\nu_{n,ij}(S_{ni}(x), S_{nj}(1)) - W_{ij}(x,1)\right|}{x^{\eta}} \times \left|\int_{1/k}^{1/a} x^{\eta-1}\cdot \ell(-\log x)^{\ell-1}\,\mathrm{d}\,x\right| \overset{\mathbb{P}}{\to} 0
\end{aligned} \tag{A.15}
$$

as $n\to\infty$ for $\ell = 1, 2$.

Now we consider the second part in (A.12). Assumption 4.5 implies that

$$
\sup_{\boldsymbol{x}\in(0,1]^{|I|}} \left|\frac{1}{q}C_I(q\boldsymbol{x}) - R_I(\boldsymbol{x})\right| \le K_R q^{\xi},
$$

for details see Eq. (S.18) in Engelke and Volgushev (2022). Let $r = x\wedge y$, then we have, for any $i,j\in V$ and $x,y>0$, that

$$
\begin{aligned}
\frac{1}{q}C_{ij}(qx,qy) &= \frac{r}{qr}C_{ij}\left(qr\frac{x}{r}, qr\frac{y}{r}\right) \\
&= r\left\{R_{ij}\left(x/r, y/r\right) + O((qr)^{\xi})\right\} = R_{ij}\left(x,y\right) + r^{1+\xi}O(q^{\xi})
\end{aligned} \tag{A.16}
$$

as $q\to 0$. Moreover, noting that by (A.16), Assumption 4.5, the relation $k = o\left(n^{\xi/(\xi+\frac{1}{2})}\right)$ and $1+\xi-\eta>0$, we have

$$
\begin{aligned}
&\sup_{u\in(0,T_0]} \frac{\left|\sqrt{k}\left\{R_{n,ij}(u, S_{nj}(1)) - R_{ij}(u, S_{nj}(1))\right\}\right|}{\left[u\wedge S_{nj}(1)\right]^{\eta}} \\
&= \sup_{u\in(0,T_0]} \left\{u\wedge S_{nj}(1)\right\}^{1+\xi-\eta} O_{\mathbb{P}}\left(k^{\xi+1/2}/n^{\xi}\right) = o_{\mathbb{P}}(1).
\end{aligned}
$$

With the fact that $D_{n12} = \sup_{s\in(0,k/(nc)]} \left(\hat{F}^{\leftarrow}_{U,i}(s)/s\right)^{\eta} \xrightarrow{\mathbb{P}} 0$, we obtain that, with arbitrarily high probability,

$$\begin{aligned}
&\sup_{1/k\le x\le 1/a}\left|\frac{\sqrt{k}\left\{R_{n,ij}(S_{ni}(x),S_{nj}(1)) - R_{ij}(S_{ni}(x),S_{nj}(1))\right\}}{x^{\eta}}\right| \\
&\le \sup_{1/k\le x\le 1/a} \frac{\left|\sqrt{k}\left\{R_{n,ij}\left(S_{ni}(x),S_{nj}(1)\right) - R_{ij}\left(S_{ni}(x),S_{nj}(1)\right)\right\}\right|}{S_{ni}(x)^{\eta}} \cdot \sup_{1/k\le x\le 1/a}\left(\frac{S_{ni}(x)}{x}\right)^{\eta} \\
&\le \sup_{u\in(0,T_0]} \frac{\left|\sqrt{k}\left\{R_{n,ij}(u,S_{nj}(1)) - R_{ij}(u,S_{nj}(1))\right\}\right|}{\{u\wedge S_{nj}(1)\}^{\eta}} \cdot \sup_{s\in(0,k/(na)]}\left(\frac{\hat{F}^{\leftarrow}_{U,i}(s)}{s}\right)^{\eta} \\
&\xrightarrow{\mathbb{P}} 0
\end{aligned}$$

as $n\to\infty$. It further follows that for $\ell = 1, 2$,

$$\begin{aligned}
&\left|\int_{1/k}^{1/a} \frac{\sqrt{k}\left\{R_{n,ij}(S_{ni}(x),S_{nj}(1)) - R_{ij}(S_{ni}(x),S_{nj}(1))\right\}\ell(-\log x)^{\ell-1}}{x}\,\mathrm{d}\,x\right| \\
&\quad \le \sup_{\substack{x\in(0,1/a]\\ x\ge 1/k}}\left|\frac{\sqrt{k}\left\{R_{n,ij}(S_{ni}(x),S_{nj}(1)) - R_{ij}(S_{ni}(x),S_{nj}(1))\right\}}{x^{\eta}}\right| \\
&\quad\quad \cdot\left|\int_{1/k}^{1/a} x^{\eta-1}\cdot\ell(-\log x)^{\ell-1}\,\mathrm{d}\,x\right| \\
&\quad \xrightarrow{\mathbb{P}} 0, \qquad n\to\infty.
\end{aligned} \tag{A.17}$$

For the third part on the right-hand side of (A.12), let $(0,T_1)\times(0,T_2)\supset(0,1/a]^2$ be the set such that the partial derivatives of $R_{ij}(x,y)$ exist and are continuous by assumption. Since $S_{ni}(x)$ converges x uniformly on $(0,1/a]$ and $R_{ij}(x,y)$ is continuous on $[0,T_1]\times[0,T_2]$, for large n such that $(S_{ni}(x),S_{nj}(1))\in[0,T_1]\times[0,T_2]$, applying the mean value theorem gives

$$\begin{aligned}
&R_{ij}\left(S_{ni}(x),S_{nj}(1)\right) - R_{ij}(x,1) \\
&\quad = \left\{R_{ij}\left(S_{ni}(x),S_{nj}(1)\right) - R_{ij}\left(S_{ni}(x),1\right)\right\} + \left\{R_{ij}(S_{ni}(x),1) - R_{ij}(x,1)\right\} \\
&\quad = \dot{R}^{i}_{ij}(\theta_{ni},1)\left(S_{ni}(x) - x\right) + \dot{R}^{j}_{ij}(S_{ni}(x),\theta_{nj})\left(S_{nj}(1) - 1\right),
\end{aligned}$$

where $\theta_{ni}\in(S_{ni}(x)\wedge x,\ S_{ni}(x)\vee x)$ and $\theta_{nj}\in(S_{nj}(1)\wedge 1,\ S_{nj}(1)\vee 1)$ almost surely for large n. Thus,

$$\begin{aligned}
&\int_{1/k}^{1/a}\frac{1}{x}\Bigg[\sqrt{k}\left\{R_{ij}(S_{ni}(x),S_{nj}(1)) - R_{ij}(x,1)\right\} \\
&\qquad + \left\{\dot{R}^{i}_{ij}(x,1)W_i(x) + \dot{R}^{j}_{ij}(x,1)W_j(1)\right\}\Bigg]\ell(-\log x)^{\ell-1}\,\mathrm{d}\,x
\end{aligned}$$

$$
\begin{aligned}
&\leq \left| \int_{1/k}^{1/a} \frac{\left[\dot{R}_{ij}^{j}(S_{ni}(x),\theta_{nj})\sqrt{k}\left\{S_{nj}(1)-1\right\} + \dot{R}_{ij}^{j}(x,1)W_j(1)\right]\ell(-\log x)^{\ell-1}}{x}\,\mathrm{d}\,x\right| \\
&\quad + \left| \int_{1/k}^{1/a} \frac{\left[\dot{R}_{ij}^{i}(\theta_{ni},1)\sqrt{k}\left\{S_{ni}(x)-x\right\} + \dot{R}_{ij}^{i}(x,1)W_i(x)\right]\ell(-\log x)^{\ell-1}}{x}\,\mathrm{d}\,x\right| \\
&=: E_{n1} + E_{n2}.
\end{aligned} \tag{A.18}
$$

We begin by analyzing the term E_{n1}. Observe that

$$
\begin{aligned}
|E_{n1}| &\leq \left| \int_{1/k}^{1/a} \frac{\dot{R}_{ij}^{j}(S_{ni}(x),\theta_{nj})\left[\sqrt{k}\{S_{nj}(1)-1\} + W_j(1)\right]\ell(-\log x)^{\ell-1}}{x}\,\mathrm{d}\,x\right| \\
&\quad + \left| \int_{1/k}^{1/a} \frac{\left\{\dot{R}_{ij}^{j}(S_{ni}(x),\theta_{nj}) - \dot{R}_{ij}^{j}(x,1)\right\}W_j(1)\ell(-\log x)^{\ell-1}}{x}\,\mathrm{d}\,x\right| \\
&\leq \left\{\sqrt{k}\{S_{nj}(1)-1\} + W_j(1)\right\}\cdot\left| \int_{1/k}^{1/a} \frac{\dot{R}_{ij}^{j}(S_{ni}(x),\theta_{nj})\ell(-\log x)^{\ell-1}}{x}\,\mathrm{d}\,x\right| \\
&\quad + \sup_{x\in[1/k,1/a]} x^{-\eta}\left|\dot{R}_{ij}^{j}(S_{ni}(x),\theta_{nj}) - \dot{R}_{ij}^{j}(x,1)\right| W_j(1)\cdot\left|\int_{1/k}^{1/a} x^{\eta-1}\cdot\ell(-\log x)^{\ell-1}\,\mathrm{d}\,x\right|,
\end{aligned} \tag{A.19}
$$

where the second term on the right hand side of the above inequality goes to zero in probability by condition (ii) of the assumption. Setting $\eta = 0$ in (A.9), one have

$$
\sqrt{k}\left[S_{nj}(1)-1\right] + W_j(1) \xrightarrow{\mathbb{P}} 0.
$$

Moreover, by the fact that

$$
0 \leq \dot{R}_{ij}^{j}(x,1) \leq R_{ij}(x,1) \leq x, \tag{A.20}
$$

we have

$$
\begin{aligned}
&\left| \int_{1/k}^{1/a} \frac{\dot{R}_{ij}^{j}(S_{ni}(x),\theta_{nj})\ell(-\log x)^{\ell-1}}{x}\,\mathrm{d}\,x\right| \\
&\leq \int_{0}^{1/a} \frac{|\dot{R}_{ij}^{j}(S_{ni}(x),\theta_{nj})|\ell(-\log x)^{\ell-1}}{x}\,\mathrm{d}\,x \\
&\leq \int_{0}^{1/a} \frac{|S_{ni}(x)|}{x}\ell(-\log x)^{\ell-1}\,\mathrm{d}\,x \\
&\leq \sup_{1/k\leq x\leq 1/a} \frac{|S_{ni}(x)|}{x} \times \int_{0}^{1/a} \ell(-\log x)^{\ell-1}\,\mathrm{d}\,x < \infty
\end{aligned}
$$

by (A.14). Hence the first term in the right-hand of (A.19) converges to zero as $n \to \infty$. Note that the tail copula $R(\boldsymbol{x})$ is nondecreasing in each coordinate, its partial derivatives (whenever they exist) are nonnegative. Hence, the inequality (A.20) can be validated by differentiating $R(\lambda\boldsymbol{x}) = \lambda R(\boldsymbol{x})$ with respect to λ and evaluating at $\lambda = 1$, which yields

$$\sum_{j=1}^{d} x_j \dot{R}^j(\boldsymbol{x}) = R(\boldsymbol{x}), \tag{A.21}$$

implying $x_j \dot{R}^j(\boldsymbol{x}) \le R(\boldsymbol{x}) \le \min_{i\in V} x_i$. In particular, this leads to the inequality (A.20). For more details, see Lamma 5 and Lemma A.3 in Einmahl and Segers (2021).

For the term E_{n2}, note that

$$\begin{aligned}
&|E_{n2}| \\
&\le \left| \int_{1/k}^{1/a} \frac{\dot{R}^i_{ij}(\theta_{ni},1)\left[\sqrt{k}\{S_{ni}(x)-x\} + W_i(x)\right]\ell(-\log x)^{\ell-1}}{x}\,\mathrm{d}\,x \right| \\
&\quad + \left| \int_{1/k}^{1/a} \frac{\left\{\dot{R}^i_{ij}(\theta_{ni},1) - \dot{R}^i_{ij}(x,1)\right\} W_i(x)\ell(-\log x)^{\ell-1}}{x}\,\mathrm{d}\,x \right| \\
&\le \sup_{1/k\le x\le 1/a} \left| \dot{R}^i_{ij}(\theta_{ni},1)\frac{\sqrt{k}\{S_{ni}(x)-x\} - W_i(x)}{x^{\eta}} \right| \cdot \left| \int_{1/k}^{1/a} x^{\eta-1}\cdot \ell(-\log x)^{\ell-1}\,\mathrm{d}\,x \right| \\
&\quad + \sup_{x\in[1/k,1/a]} \left| \dot{R}^i_{ij}(\theta_{ni},1) - \dot{R}^i_{ij}(x,1) \right| \cdot \sup_{x\in[1/k,1/a]} \frac{|W_i(x)|}{x^{\eta}} \cdot \left| \int_{1/k}^{1/a} x^{\eta-1}\cdot \ell(-\log x)^{\ell-1}\,\mathrm{d}\,x \right|.
\end{aligned}$$

The first part on the right-hand side goes to zero as $n \to \infty$ from (A.9) and the inequality $0 \le \dot{R}^i_{ij}(\theta_{ni},1) \le 1$, where the latter is implied by $\theta_{ni}\dot{R}^i_{ij}(\theta_{ni},1) \le \dot{R}^i_{ij}(\theta_{ni},1) \le \theta_{ni}$ from (A.21). By assumption $\dot{R}^i_{ij}(x,1)$ is a continuous function, and hence uniformly continuous on the interval $x \in [1/k, 1/a]$. Recall that θ_{ni} lies between x and $S_{ni}(x)$. From (A.14) we know that $S_{ni}(x)$ converges uniformly to x. As a consequence, θ_{ni} also converges to its limiting value. Hence we have

$$\sup_{x\in[1/k,1/a]} \left| \dot{R}^i_{ij}(\theta_{ni},1) - \dot{R}^i_{ij}(x,1) \right| \to 0$$

as $n \to \infty$. Recall that $W_i(x)$ is a normal distributed random variable with zero mean and variance x. For any $M > 0$,

$$\begin{aligned}
\mathbb{P}\left(\left| \sup_{x\in(0,1/a]} \frac{W_i(x)}{x^{\eta}} \right| > M \right) &= \mathbb{P}\left(|W_i(1)| \left| \sup_{x\in(0,1/a]} x^{1/2-\eta} \right| > M \right) \\
&= 2\,\mathbb{P}\left[W_i(1) > M a^{\eta-1/2} \right] \le \frac{2\,\mathbb{E}(W_i(1)^2)}{M^2 a^{2\eta-1}} \to 0
\end{aligned}\tag{A.22}$$

as $M \to \infty$ by Chebyshev inequality. Hence $E_{n2} \xrightarrow{\mathbb{P}} 0$ as $n \to \infty$. Consequently, by (A.12), (A.15), (A.17) and (A.18), we obtain

$$\int_{1/k}^{1/a} \frac{\{\hat{\nu}_{n,ij}(x,1) - B_{ij}(x,1)\}\,\ell(-\log x)^{\ell-1}}{x}\,\mathrm{d}\,x \xrightarrow{\mathbb{P}} 0$$

as $n \to \infty$. Recall the definition of $B_{ij}(x,1)$ in (4.4). By (A.20), we have

$$\int_0^{1/k} \frac{B_{ij}(x,1)\ell(-\log x)^{\ell-1}}{x} \,\mathrm{d}\, x \to 0,$$

in mean square as $n \to \infty$, implying the desired result. □

Recall the definition of $S_{n,I}(\boldsymbol{x}_I)$ in (A.2). The following lemma verifies condition (ii) of Lemma A.4 for the Hüsler–Reiss MGPD.

Lemma A.5. *Let $R(\boldsymbol{x})$ be the tail copula function of a Hüsler–Reiss MGPD with variogram matrix $\boldsymbol{\Gamma} = (\gamma_{ij})_{i,j\in V}$ satisfying Assumption 4.6. For each $j \in V$, assume $\{\theta_{nj}\}_{n\geq 1}$ is a sequence lying between 1 and $S_{nj}(1)$ almost surely for large n. Then, for $i,j \in V$ and $i \neq j$, we have*

$$\sup_{x\in[1/k,\exp(c)]} x^{-\eta} \left| \dot{R}_{ij}^{j}(S_{ni}(x), \theta_{nj}) - \dot{R}_{ij}^{j}(x,1) \right| \to 0$$

in probability as $n \to \infty$.

Proof. For $i,j \in V$, the bivariate tail copula associated with the (i,j) component of the Hüsler–Reiss distribution is a function of γ_{ij} only. For notational simplicity, we write

$$R_{ij}(x,y) = R(x,y;\gamma_{ij}),\ \ i,j \in V$$

with

$$R(x,y;\gamma) = x + y - x\,\Phi\left(\frac{\sqrt{\gamma}}{2} + \frac{\log x - \log y}{\sqrt{\gamma}}\right) - y\,\Phi\left(\frac{\sqrt{\gamma}}{2} + \frac{\log y - \log x}{\sqrt{\gamma}}\right) \tag{A.23}$$

for $(x,y) \in [0,\infty)^2$. Taking derivative with respect to x and y respectively, we get

$$\dot{R}_{ij}^{i}(x,y) = \dot{R}_{ij}^{j}(y,x) = 1 - \Phi\left(\frac{\sqrt{\gamma_{ij}}}{2} + \frac{\log x - \log y}{\sqrt{\gamma_{ij}}}\right). \tag{A.24}$$

Therefore,

$$\begin{aligned}
&\sup_{x\in[1/k,1/a]} x^{-\eta} \left| \dot{R}_{ij}^{j}(S_{ni}(x), S_{nj}(1)) - \dot{R}_{ij}^{j}(x,1) \right| \\
&\quad\leq \sup_{x\in[1/k,1/a]} x^{-\eta} \left| \Phi\left(\sqrt{\gamma_{ij}} - \frac{\log S_{ni}(x) - \log S_{nj}(1)}{2\sqrt{\gamma_{ij}}}\right) - \Phi\left(\sqrt{\gamma_{ij}} - \frac{\log x}{2\sqrt{\gamma_{ij}}}\right) \right| \\
&\quad=: h_n(x).
\end{aligned}$$

Recall that $c = -\log a$. By (A.14), we have

$$\frac{S_{ni}(x)}{S_{nj}(1)} \to x$$

almost surely as $n \to \infty$. Applying Taylor expansion to the standard normal distribution function at $\sqrt{\gamma_{ij}} - \log x/(2\sqrt{\gamma_{ij}})$ gives

$$\begin{aligned} h_n(x) &= \sup_{x\in[1/k,1/a]} x^{-\eta} \left| -\frac{1}{2\sqrt{\gamma_{ij}}} \frac{1}{x} \phi\left(\sqrt{\gamma_{ij}} - \frac{\log x}{2\sqrt{\gamma_{ij}}}\right) \left(\frac{S_{ni}(x)}{S_{nj}(1)} - x\right) \right| (1 + o(1)) \\ &\le \frac{1}{\sqrt{\gamma_{ij}}} \sup_{x\in[1/k,1/a]} \frac{1}{x} \phi\left(\sqrt{\gamma_{ij}} - \frac{\log x}{2\sqrt{\gamma_{ij}}}\right) \cdot \sup_{x\in[1/k,1/a]} x^{-\eta} \left| \frac{S_{ni}(x)}{S_{nj}(1)} - x \right| (1 + o(1)) \end{aligned}$$

for sufficiently large n. Since

$$\lim_{n\to\infty} k\phi\left(\sqrt{\gamma_{ij}} + \frac{\log k}{2\sqrt{\gamma_{ij}}}\right) = 0.$$

and it is continuous on $[1/k, 1/a]$, we have

$$\sup_{x\in[1/k,1/a]} \frac{1}{x} \phi\left(\sqrt{\gamma_{ij}} - \frac{\log x}{2\sqrt{\gamma_{ij}}}\right) < \infty$$

Moreover, note that

$$\begin{aligned} &\sup_{x\in[1/k,1/a]} x^{-\eta} \left| \frac{S_{ni}(x)}{S_{nj}(1)} - x \right| \\ &= \frac{1}{|S_{nj}(1)|} \left[\frac{1}{\sqrt{k}} \sup_{x\in[1/k,1/a]} \left\{ \left| \frac{\sqrt{k}(S_{ni}(x) - x) - W_i(x)}{x^\eta} \right| + \frac{|W_i(x)|}{x^\eta} \right\} \right. \\ &\qquad \left. + \sup_{x\in[1/k,1/a]} x^{1-\eta} |S_{nj}(1) - 1| \right] \\ &\le \frac{1}{|S_{nj}(1)|} \left[\frac{1}{\sqrt{k}} \left\{ \sup_{x\in[1/k,1/a]} \left| \frac{\sqrt{k}(S_{ni}(x) - x) - W_i(x)}{x^\eta} \right| + \sup_{x\in[1/k,1/a]} \frac{|W_i(x)|}{x^\eta} \right\} \right. \\ &\qquad \left. + a^{\eta-1} |S_{nj}(1) - 1| \right] \\ &\to 0 \end{aligned}$$

in probability by (A.9), (A.22) and the fact that $S_{nj}(1) \to 1$ almost surely, as $n \to \infty$. This implies $h_n(x) \to 0$ in probability and the proof is complete. □

A.2 Proofs of propositions and theorems

In this section, we give proofs of the lemma, propositions and theorems in Section 4.

Proof of Lemma 3.1. We first derive the expressions of the moment functions. Recall that for any $i, j \in V$ and $i \neq j$,

$$\left(Y_i^{(j)}, Y_j^{(j)}\right) \overset{d}{=} \left(Y_j^{(i)}, Y_i^{(i)}\right) \overset{d}{=} \left(E + Z_i^{(j)}, E\right),$$

where $Z_i^{(j)}$ is a normally distributed random variable with mean $-\gamma_{ij}/2$ and variance γ_{ij}, independent of the standard exponential random variable E. By the independence of E and $Z_i^{(j)}$, a direct calculation gives

$$\begin{aligned}\mathbb{E}[(E+x)_+] &= \int_0^\infty (y+x)_+ \exp(-y)\,\mathrm{d}\,y = \int_{(-x)\vee 0}^\infty (y+x)\exp(-y)\,\mathrm{d}\,y \\ &= \exp(x)\,\mathbb{1}\{x<0\} + (1+x)\,\mathbb{1}\{x\geq 0\}, \end{aligned} \tag{A.25}$$

and then

$$\begin{aligned} e^{(1)}(\gamma_{ij},c) &= \mathbb{E}\left\{\left(E+Z_i^{(j)}+c\right)_+\right\} \\ &= \int_{-\infty}^\infty \mathbb{E}\{(E+x)_+\}\;\mathrm{d}\,\mathbb{P}[Z_i^{(j)}+c\leq x] \\ &= \int_{-\infty}^0 \mathrm{e}^x\;\mathrm{d}\,\mathbb{P}\left\{Z_i^{(j)}+c\leq x\right\} + \int_0^\infty (1+x)\;\mathrm{d}\,\mathbb{P}\left\{Z_i^{(j)}+c\leq x\right\} \\ &= \exp(c)\Phi\left(-\frac{c+\gamma_{ij}/2}{\sqrt{\gamma_{ij}}}\right) + \sqrt{\gamma_{ij}}\phi\left(\frac{c-\gamma_{ij}/2}{\sqrt{\gamma_{ij}}}\right) + \left(1-\frac{\gamma_{ij}}{2}+c\right)\Phi\left(\frac{c-\gamma_{ij}/2}{\sqrt{\gamma_{ij}}}\right). \end{aligned}$$

Similarly to (A.25), we have

$$\begin{aligned}\mathbb{E}\left[\{(E+x)_+\}^2\right] &= \int_0^\infty \{(y+x)_+\}^2\exp(-y)\,\mathrm{d}\,y = \int_{(-x)\vee 0}^\infty (y+x)^2\exp(-y)\,\mathrm{d}\,y \\ &= 2\exp(x)\,\mathbb{1}(x<0) + (2+2x+x^2)\,\mathbb{1}(x\geq 0). \end{aligned}$$

It follows that

$$\begin{aligned} e^{(2)}(\gamma_{ij},c) &= \mathbb{E}\left[\left\{\left(E+Z_i^{(j)}+c\right)_+\right\}^2\right] \\ &= \int_{-\infty}^\infty \mathbb{E}\left[\{(E+x)_+\}^2\right]\mathrm{d}\,\mathbb{P}\left\{Z_i^{(j)}+c\leq x\right\} \\ &= 2\int_{-\infty}^0 \exp(x)\;\mathrm{d}\,\mathbb{P}\left\{Z_i^{(j)}+c\leq x\right\} + \int_0^\infty (2+2x+x^2)\;\mathrm{d}\,\mathbb{P}\left\{Z_i^{(j)}+c\leq x\right\} \\ &= 2e_{ij}^{(1)} + \int_0^\infty x^2\;\mathrm{d}\,\mathbb{P}\left\{Z_i^{(j)}+c\leq x\right\} \\ &= 2e_{ij}^{(1)} - \sqrt{\gamma_{ij}}\left(\frac{\gamma_{ij}}{2}-c\right)\phi\left(\frac{c-\gamma_{ij}/2}{\sqrt{\gamma_{ij}}}\right) + \left\{\frac{1}{4}\gamma_{ij}^2 + (1-c)\gamma_{ij} + c^2\right\}\Phi\left(\frac{c-\gamma_{ij}/2}{\sqrt{\gamma_{ij}}}\right). \end{aligned}$$

Plugging in the expression of $e^{(1)}(\gamma_{ij},c)$ yields the formula of $e^{(2)}(\gamma_{ij},c)$.

We now show the monotonicity of the moment functions. For given c, the functions $e^{(\ell)}(\gamma) = e^{(\ell)}(\gamma,c)$ with $\ell = 1,2$ that map $[0,\infty)$ to $\mathbb{R}$, are injective and continuously differentiable on $(0,\infty)$, and right-continuous at 0. The derivatives of $e^{(\ell)}(\gamma)$, $\ell = 1,2$, satisfy

$$e^{(1),\prime}(\gamma) = -\frac{1}{2}\Phi\left(\frac{c-\frac{\gamma}{2}}{\sqrt{\gamma}}\right) < 0$$

and

$$e^{(2),\prime}(\gamma) = \left(\frac{\gamma}{2} - c\right)\Phi\left(\frac{c - \frac{\gamma}{2}}{\sqrt{\gamma}}\right) - \sqrt{\gamma}\phi\left(\frac{c - \frac{\gamma}{2}}{\sqrt{\gamma}}\right) < 0 \tag{A.26}$$

for $\gamma \in (0, \infty)$. Here, the inequality in (A.26) follows from Mill's inequality that $(1 - \Phi(x))/\phi(x) < 1/x$ for $x > 0$. Specifically, if $\gamma/2 - c \le 0$, the inequality (A.26) already holds since both terms in the left expression of the inequality are negative for $\gamma > 0$; otherwise, if $\gamma/2 - c > 0$, we have by Mill's inequality that

$$\frac{\left(\frac{\gamma}{2} - c\right)\Phi\left(\frac{c-\frac{\gamma}{2}}{\sqrt{\gamma}}\right)}{\sqrt{\gamma}\phi\left(\frac{c-\frac{\gamma}{2}}{\sqrt{\gamma}}\right)} = \frac{\frac{\gamma}{2} - c}{\sqrt{\gamma}} \cdot \frac{1 - \Phi\left(\frac{\frac{\gamma}{2}-c}{\sqrt{\gamma}}\right)}{\phi\left(\frac{\frac{\gamma}{2}-c}{\sqrt{\gamma}}\right)} < 1,$$

which yields (A.26). Moreover, by right-continuity at $\gamma = 0$, the monotonicity extends to $[0, \infty)$. Consequently, $e^{(\ell)} : [0, \infty) \to (0, \infty)$ $(\ell = 1, 2)$ is strictly decreasing. The proof is complete. □

Proof of Proposition 4.2. Recall that $Y_i^{(j)}$ is $Y_i \mid Y_j > 0$ and that $R(\boldsymbol{x})$ is the tail copula defined in (4.1). By (2.1), (2.4), (4.1) and the assumption on $\boldsymbol{X}$ and $\boldsymbol{Y}$, we know that

$$\mathbb{P}\left(Y_i^{(j)} \ge x, Y_j^{(j)} \ge y\right) = R_{ij}(\exp(-x), \exp(-y)), \qquad (x, y) \in [-\infty, \infty] \times [0, \infty]. \tag{A.27}$$

With the notation $c = -\log a$ for $a \in (0, 1]$, we have

$$\mathbb{E}\left[\left\{Y_i^{(j)} \vee (-c)\right\}^{\ell}\right] = \int_0^{1/a} \frac{R_{ij}(x, 1)\ell(-\log x)^{\ell-1}}{x}\,\mathrm{d}\,x + (\log a)^{\ell} := \tilde{e}^{(\ell)}(\gamma, c); \tag{A.28}$$

for the proof of this expression, see Section A.3.

Note that, with the transformation $u = -\log q$ $(0 < q \le 1)$, the excesses on the exponential scale can be equivalently represented as

$$-\log(1 - F_i(X_i)) - u \mid -\log(1 - F_j(X_j)) > u = \log\left\{\frac{q}{1 - F_i(X_i)}\right\} \mid 1 - F_j(X_j) < q.$$

Therefore, we formulate the pre-asymptotic representation of $\tilde{e}^{(\ell)}(\gamma, c)$ on the uniform scale directly. For $q \in (0, 1]$, let

$$\boldsymbol{U}^{(m)}(q) := \boldsymbol{U} \mid U_m \le q$$

be the random vector with distribution on

$$\mathcal{D}^m(q) := [0, 1]^{m-1} \times [0, q] \times [0, 1]^{d-m}$$

given by

$$\mathbb{P}\left(\boldsymbol{U}^{(m)}(q) \le \boldsymbol{x}\right) = q^{-1}\,\mathbb{P}\left(U_1 \le x_1, \ldots, U_m \le (x_m \wedge q), \ldots, U_d \le x_d\right)$$

$$= q^{-1}C(x_1, \dots, (x_m \wedge q), \dots, x_d), \qquad \boldsymbol{x} \in [0,1]^d. \tag{A.29}$$

Define the clipped pre-asymptotic form moment $e_{q,ij}^{(\ell)}$ for $i, j \in V$ as

$$e_{q,ij}^{(\ell)} := \mathbb{E}\left[\left\{\log\left(q/U_i^{(j)}(q)\right) \vee (-c)\right\}^{\ell}\right] = \mathbb{E}\left[\left\{\log\left((q/U_i^{(j)}(q)) \vee a\right)\right\}^{\ell}\right].$$

Then we have

$$e_{q,ij}^{(\ell)} = \int_0^{1/a} \frac{C_{ij}(qt, q)\ell(-\log t)^{\ell-1}}{qt} \,\mathrm{d}\, t + (\log a)^{\ell}, \tag{A.30}$$

see Section A.3 for the proof. Since $q^{-1}C_{ij}(qx, q) \le x$, it follows from the representation of $\mathbb{E}\left[\left\{Y_i^{(j)} \vee (-c)\right\}^{\ell}\right]$ and $e_{q,ij}^{(\ell)}$, the convergence in (4.1) and the dominated convergence theorem that, for $\ell = 1, 2$,

$$e_{q,ij}^{(\ell)} \to \mathbb{E}\left[\left\{Y_i^{(j)} \vee (-c)\right\}^{\ell}\right], \text{ as } q \to 0. \tag{A.31}$$

Now we consider the empirical moments. Based on the definition of $\hat{Y}_{ti}$ in (2.5), we define

$$\begin{aligned}
\tilde{e}_{n,ij}^{(\ell)}(k, c) &:= \frac{1}{k}\sum_{t=1}^{n}\left\{\hat{Y}_{ti} \vee (-c)\right\}^{\ell} \mathbb{1}\left(\hat{F}_j(X_{tj}) \ge 1 - \frac{k}{n}\right) \\
&= \frac{1}{k}\sum_{t=1}^{n}\left\{\log\left(\frac{k}{n(1 - \hat{F}_i(X_{ti}))} \vee a\right)\right\}^{\ell} \mathbb{1}\left(\hat{F}_j(X_{tj}) \ge 1 - \frac{k}{n}\right).
\end{aligned} \tag{A.32}$$

Using the definition of $\hat{U}_{ti}$, it can be written as

$$\tilde{e}_{n,ij}^{(\ell)}(k, -\log a) = \frac{1}{k}\sum_{t=1}^{n}\left\{-\log\left(\frac{n}{k}\hat{U}_{tj} \wedge a^{-1}\right)\right\}^{\ell} \mathbb{1}\left(\hat{U}_{ti} \le \frac{k}{n}\right).$$

With the definition of $\tilde{R}(\boldsymbol{x})$ in (A.3), we obtain the expansion of $\tilde{e}_{n,ij}^{(\ell)}$ as

$$\tilde{e}_{n,ij}^{(\ell)} = \int_{1/k}^{1/a} \frac{\tilde{R}_{n,ij}(t, 1)\ell(-\log t)^{\ell-1}}{t} \,\mathrm{d}\, t + (\log a)^{\ell}, \tag{A.33}$$

see Section A.3 for the proof.

Note that

$$\sup_{\boldsymbol{x} \in [0, n/k]^{|I|}} \left|\hat{R}_{n,I}(\boldsymbol{x}) - \tilde{R}_{n,I}(\boldsymbol{x})\right| = O(1/k) \tag{A.34}$$

almost surely as $n \to \infty$. For any constant $T \ge 0$, $I \subset V$ and $|I| \le 3$, by (S.47) in Engelke and Volgushev (2022) and the homogeneity property of tail copulas, we know that

$$\sup_{\boldsymbol{x} \in [0, T]^{|I|}} \left|\hat{R}_{n,I}(\boldsymbol{x}) - \frac{n}{k}C_I(k\boldsymbol{x}/n)\right| = O_p(k^{-1/2}). \tag{A.35}$$

Therefore, by setting $q = k/n$ and combining the representations of $\tilde{e}^{(\ell)}_{q,ij}$ and $\tilde{e}^{(\ell)}_{n,ij}$ with (A.34)–(A.35), we have

$$\begin{aligned}
&\left|\tilde{e}^{(\ell)}_{n,ij} - \tilde{e}^{(\ell)}_{\frac{k}{n},ij}\right| \\
&= \Bigg| \int_{1/k}^{1/a} \frac{\{\hat{R}_{n,ij}(x,1) - \frac{n}{k}C_{ij}\left(\frac{kx}{n},\frac{k}{n}\right)\}\ell(-\log x)^{\ell-1}}{x}\,\mathrm{d}\,x \\
&\qquad\qquad - \int_0^{1/k} \frac{\frac{n}{k}C_{ij}\left(\frac{kx}{n},\frac{k}{n}\right)\ell(-\log x)^{\ell-1}}{x}\,\mathrm{d}\,x + O_p(k^{-1}(\log k)^{\ell}) \Bigg| \\
&\le \sup_{x,y\in[0,1/a]^2} \left| \hat{R}_{n,ij}(x,y) - \frac{n}{k}C_{ij}\left(\frac{kx}{n},\frac{ky}{n}\right)\right| \cdot (\log k - \log a)^{\ell} + O_p\left(k^{-1}(\log k)^{\ell}\right) \\
&= O_p\left(k^{-1}(\log k)^{\ell}\right) = o_p(1)
\end{aligned}$$

as $n \to \infty$. Setting $q = k/n$ in (A.31) and combining with the above equation yields that

$$\tilde{e}^{(\ell)}_{n,ij} \xrightarrow{\mathbb{P}} \mathbb{E}\left[\left\{Y_i^{(j)} \vee (-c)\right\}^{\ell}\right], \text{ as } n \to \infty. \tag{A.36}$$

To establish the main result, note that the clipped moment function $\hat{e}^{(\ell)}_{n,ij}$ in (3.2) can be expressed as continuous functions of the empirical moment functions studied above. Specifically, by the definition of $\tilde{e}^{(\ell)}_{n,ij}$ in (A.32) and $\hat{e}^{(\ell)}_{n,ij}$, for $i,j \in V$ $(i \ne j)$ and $\ell = 1,2$, we have

$$\begin{aligned}
\hat{e}^{(1)}_{n,ij} &= \frac{1}{k}\sum_{t=1}^{n}\left\{\left(\log\frac{k}{n(1-\hat{F}_i(X_{ti}))} - \log a\right)_+\right\}\mathbb{1}\left(\hat{F}_j(X_{tj}) \ge 1 - \frac{k}{n}\right) \\
&= \frac{1}{k}\sum_{t=1}^{n}\left\{\log\left(\frac{k}{n(1-\hat{F}_i(X_{ti}))} \vee a\right)\right\}\mathbb{1}\left(\hat{F}_j(X_{tj}) \ge 1 - \frac{k}{n}\right) - \log a \\
&= \tilde{e}^{(1)}_{n,ij} - \log a
\end{aligned} \tag{A.37}$$

and

$$\begin{aligned}
\hat{e}^{(2)}_{n,ij} &= \frac{1}{k}\sum_{t=1}^{n}\left\{\left(\log\frac{k}{n(1-\hat{F}_i(X_{ti}))} - \log a\right)_+\right\}^2\mathbb{1}\left(\hat{F}_j(X_{tj}) \ge 1 - \frac{k}{n}\right) \\
&= \frac{1}{k}\sum_{t=1}^{n}\left\{\log\left(\frac{k}{n(1-\hat{F}_i(X_{ti}))} \vee a\right) - \log a\right\}^2\mathbb{1}\left(\hat{F}_j(X_{tj}) \ge 1 - \frac{k}{n}\right) \\
&= \tilde{e}^{(2)}_{n,ij} - 2(\log a)\tilde{e}^{(1)}_{n,ij} + (\log a)^2.
\end{aligned} \tag{A.38}$$

By the continuous mapping theorem, (A.36) and the representations above, the convergence in Proposition 4.2 holds. □

Proof of Theorem 4.4. From Proposition 4.2 we know that, as $n \to \infty$,

$$\hat{e}^{(\ell)}_{n,ij} \xrightarrow{\mathbb{P}} e^{(\ell)}(\gamma_{ij}).$$

By the symmetry of the variogram matrix, we have $\gamma_{ij} = \gamma_{ji}$. Therefore, $e^{(\ell)}(\gamma_{ij}) = e^{(\ell)}(\gamma_{ji})$. Consequently,

$$\frac{1}{2}\left(\hat{e}^{(\ell)}_{n,ij} + \hat{e}^{(\ell)}_{n,ji}\right) \xrightarrow{\mathbb{P}} e^{(\ell)}(\gamma_{ij}) \tag{A.39}$$

as $n \to \infty$ for $\ell = 1, 2$.

Furthermore, by Lemma 3.1, the function $e^{(\ell)}(\gamma)$ is strictly decreasing on $(0, \infty)$, with boundary limits

$$\lim_{\gamma\to 0} e^{(1)}(\gamma, c) = 1 + c, \quad \lim_{\gamma\to 0} e^{(2)}(\gamma, c) = (1 + c)^2 + 1$$

and

$$\lim_{\gamma\to\infty} e^{(\ell)}(\gamma, c) = 0.$$

Hence, $e^{(\ell)}(\gamma) = e^{(\ell)}(\cdot, c) : (0, \infty) \to \mathcal{R}^{(\ell)}_c$ $(\ell = 1, 2)$ is a bijection with range $\mathcal{R}^{(1)}_c = (0, 1+c)$ or $\mathcal{R}^{(2)}_c = (0, (1 + c)^2 + 1)$. Therefore, its inverse function $e^{(\ell),\leftarrow}$ is well-defined and continuous on $\mathcal{R}^{(\ell)}_c$. Consequently, the moment estimator $\hat{\gamma}^{\mathrm{M},(\ell)}_{n,ij}$ $(\ell = 1, 2)$ in (3.6) is well-defined whenever

$$\frac{1}{2}\left(\hat{e}^{(\ell)}_{n,ij} + \hat{e}^{(\ell)}_{n,ji}\right) \in \mathcal{R}^{(\ell)}_c,$$

which holds with probability tending to one by (A.39). This establishes the existence of $\hat{\gamma}^{\mathrm{M},(\ell)}_{n,ij}$ with probability tending to one.

Consider the case where $\gamma_{ij} = 0$. Recall that we interpret $e^{(\ell)}(0, c)$ by continuous extension

$$e^{(1)}(0, c) = 1 + c, \quad e^{(2)}(0, c) = (1 + c)^2 + 1.$$

Since $e^{(\ell)}(\gamma, c)$ is continuous and strictly monotone on $(0, \infty)$, its inverse admits a right-continuous extension at the upper boundary points of $\mathcal{R}^{(\ell)}_c$, in the sense that

$$e^{(1),\leftarrow}(y) \to 0 \quad \text{as } y \uparrow 1 + c$$

and

$$e^{(2),\leftarrow}(y) \to 0 \quad \text{as } y \uparrow (1 + c)^2 + 1.$$

Therefore, the inverse mapping can be continuously extended to $e^{(\ell)}(0, c)$ via right-continuity by setting

$$e^{(\ell),\leftarrow}(e^{(\ell)}(0, c)) = 0.$$

Hence, the moment estimators remain well-defined in this case.

Finally, the weak consistency of $\hat{\gamma}^{\mathrm{M},(\ell)}_{n,ij}$ follows from the continuous mapping theorem applied to extended inverse mapping $e^{(\ell),\leftarrow} : \tilde{\mathcal{R}}^{(\ell)}_c \to [0, \infty)$ and (A.39), where $\tilde{\mathcal{R}}^{(1)}_c = (0, 1 + c]$ and $\tilde{\mathcal{R}}^{(2)}_c = (0, (1 + c)^2 + 1]$. □

Proof of Proposition 4.7. Recall that

$$\hat{\nu}_{n,ij}(x, 1) = \sqrt{k}\left\{\hat{R}_{n,ij}(x, 1) - R_{ij}(x, 1)\right\}.$$

For $\ell = 1, 2$, using the expansion of $\tilde{e}^{(\ell)}_{n,ij} = \tilde{e}^{(\ell)}_{n,ij}(k, c)$ in (A.33) and that of $\tilde{e}^{(\ell)}(\gamma) = \tilde{e}^{(\ell)}(\gamma, c)$ in (A.28), together with (A.34) we have

$$\begin{aligned}
&\sqrt{k}\left(\tilde{e}^{(\ell)}_{n,ij} - \tilde{e}^{(\ell)}(\gamma_{ij})\right) \\
&= \sqrt{k}\left\{\int_{1/k}^{1/a} \frac{\tilde{R}_{n,ij}(x,1)\ell(-\log x)^{\ell-1}}{x}\,\mathrm{d}\,x - \int_0^{1/a} \frac{R_{ij}(x,1)\ell(-\log x)^{\ell-1}}{x}\,\mathrm{d}\,x\right\} \\
&= \sqrt{k}\Bigg[\int_{1/k}^{1/a} \frac{\left\{\tilde{R}_{n,ij}(x,1) - R_{ij}(x,1)\right\}\ell(-\log x)^{\ell-1}}{x}\,\mathrm{d}\,x \\
&\qquad - \int_0^{1/k} \frac{R_{ij}(x,1)\ell(-\log x)^{\ell-1}}{x}\,\mathrm{d}\,x\Bigg] \\
&= \int_{1/k}^{1/a} \frac{\hat{\nu}_{n,ij}(x,1)\ell(-\log x)^{\ell-1}}{x}\,\mathrm{d}\,x - \sqrt{k}\int_0^{1/k} \frac{R_{ij}(x,1)\ell(-\log x)^{\ell-1}}{x}\,\mathrm{d}\,x \\
&\quad + O\left(k^{-1/2}(\log k)^{\ell}\right)
\end{aligned}$$

almost surely. Note that

$$\sqrt{k}\int_0^{1/k} \frac{R_{ij}(x,1)\ell(-\log x)^{\ell-1}}{x}\,\mathrm{d}\,x \leq \sqrt{k}\int_0^{1/k} \ell(-\log x)^{\ell-1}\,\mathrm{d}\,x \leq \frac{2}{\sqrt{k}}(\log k + 1)$$

for $\ell = 1, 2$ by the inequality $R_{ij}(x,1) \leq x$. Under the assumption that $\gamma_{ij} > 0$, the partial derivatives of the bivariate tail copula function of the Hüsler–Reiss distribution are continuous. Combining Lemma A.4, Lemma A.5 with the Slutsky's theorem (see Theorem 2.7 in Van der Vaart and Wellner (1996)), we obtain

$$\begin{aligned}
&\sqrt{k}\left(\tilde{e}^{(1)}_{n,ij} - \tilde{e}^{(1)}(\gamma_{ij}),\, \tilde{e}^{(2)}_{n,ij} - \tilde{e}^{(2)}(\gamma_{ij}),\ i, j \in V,\ i \neq j\right) \\
&\xrightarrow{d} \left(\int_0^{1/a} \frac{B_{ij}(x,1)}{x}\,\mathrm{d}\,x,\ -2\int_0^{1/a} \frac{B_{ij}(x,1)\log x}{x}\,\mathrm{d}\,x,\ i, j \in V,\ i \neq j\right).
\end{aligned} \tag{A.40}$$

To obtain the asymptotic distribution of the original moment functions, we next relate $\hat{e}^{(1)}_{n,ij}(k, c)$ to $\tilde{e}^{(1)}_{n,ij}(k, c)$. By (A.37) and (A.38), we have

$$\sqrt{k}\left(\hat{e}^{(1)}_{n,ij}(k,c) - e^{(1)}(\gamma_{ij})\right) = \sqrt{k}\left(\tilde{e}^{(1)}_{n,ij}(k,c) - \tilde{e}^{(1)}(\gamma_{ij})\right)$$

and

$$\begin{aligned}
&\sqrt{k}\left(\hat{e}^{(2)}_{n,ij}(k,c) - e^{(2)}(\gamma_{ij})\right) \\
&\qquad = \sqrt{k}\left(\tilde{e}^{(2)}_{n,ij}(k,c) - \tilde{e}^{(2)}(\gamma_{ij})\right) - 2\log a \cdot \sqrt{k}\left(\tilde{e}^{(1)}_{n,ij}(k,c) - \tilde{e}^{(1)}(\gamma_{ij})\right).
\end{aligned}$$

Combining these identities with (A.40) yields the desired weak convergence result. □

Proof of Theorem 4.8. By Proposition 4.7 and the relation that $c = -\log a$, we have

$$\left(\frac{1}{2}\left\{\sqrt{k}\left(\hat{e}^{(\ell)}_{n,ij}(k,c) - e^{(\ell)}(\gamma_{ij})\right) + \sqrt{k}\left(\hat{e}^{(\ell)}_{n,ji}(k,c) - e^{(\ell)}(\gamma_{ji})\right)\right\}, \, i,j \in V, \, i \neq j\right)$$
$$\xrightarrow{d} \left(\frac{1}{2}\int_0^{\exp(c)} \frac{\{B_{ij}(x,1) + B_{ji}(x,1)\}\,\ell(-\log x + c)^{\ell-1}}{x}\,\mathrm{d}\,x, \, i,j \in V, \, i \neq j\right) \tag{A.41}$$

for $\ell = 1, 2$ as $n \to \infty$.

Recall that $e^{(\ell),\leftarrow}$ denotes the inverse function of $e^{(\ell)}(\gamma)$ for $\ell = 1, 2$. From the definition of $\hat{\gamma}^{\mathrm{M},(\ell)}_{n,ij}$ in (3.6) and the fact that $\gamma_{ij} = \gamma_{ji}$, we can write

$$\sqrt{k}\left(\hat{\gamma}^{\mathrm{M},(\ell)}_{n,ij} - \gamma_{ij}\right) = e^{(\ell),\leftarrow}\left(\frac{\hat{e}^{(\ell)}_{n,ij} + \hat{e}^{(\ell)}_{n,ji}}{2}\right) - e^{(\ell),\leftarrow}\left(\frac{e^{(\ell)}(\gamma_{ij}) + e^{(\ell)}(\gamma_{ji})}{2}\right)$$

for $i, j \in V$ and $i \neq j$. Here, the inverse at $(\hat{e}^{(\ell)}_{n,ij} + \hat{e}^{(\ell)}_{n,ji})/2$ is well-defined whenever $(\hat{e}^{(\ell)}_{n,ij} + \hat{e}^{(\ell)}_{n,ji})/2 \in \mathcal{R}^{(\ell)}_c$ (see the proof of Theorem 4.4). Note that $e^{(\ell)}(\gamma)$ is differentiable on $(0, \infty)$, with derivative $e^{(\ell),\prime}(x)$ given in Theorem 4.8. Hence, by the delta method and (A.41), we have for $\ell = 1, 2$ that

$$\left(\sqrt{k}\left(\hat{\gamma}^{\mathrm{M},(\ell)}_{n,ij} - \gamma_{ij}\right); \, i,j \in V, \, i \neq j\right)$$
$$\xrightarrow{d} \left(\frac{1}{2e^{(\ell),\prime}(\gamma_{ij})}\int_0^{\exp(c)} \frac{(B_{ij}(x,1) + B_{ji}(x,1))\cdot \ell \cdot (-\log x + c)^{\ell-1}}{x}\,\mathrm{d}\,x; \, i,j \in V, \, i \neq j\right).$$

The proof is complete. □

A.3 Proofs of the representations of moments

A.3.1 Proof of Eq. (A.28)

Proof. For $x \in \mathbb{R}$ and $\ell = 1, 2$, we have

$$x^\ell = \int_0^\infty \mathbb{1}\{t \leq x\}\ell t^{\ell-1}\,\mathrm{d}\,t + (-1)^\ell \int_0^\infty \mathbb{1}\{t < -x\}\ell t^{\ell-1}\,\mathrm{d}\,t. \tag{A.42}$$

Hence,

$$\begin{aligned}\left\{Y_i^{(j)} \vee (-c)\right\}^\ell &= \int_0^\infty \mathbb{1}\{t \leq Y_i^{(j)} \vee (-c)\}\cdot \ell t^{\ell-1}\,\mathrm{d}\,t \\ &\quad + (-1)^\ell \int_0^\infty \mathbb{1}\{t < -(Y_i^{(j)} \vee (-c))\}\cdot \ell t^{\ell-1}\,\mathrm{d}\,t \\ &= \int_0^\infty \mathbb{1}\{t \leq Y_i^{(j)} \vee (-c)\}\cdot \ell t^{\ell-1}\,\mathrm{d}\,t + (-1)^\ell \int_0^c \mathbb{1}\{-c \leq Y_i^{(j)} < -t\}\cdot \ell t^{\ell-1}\,\mathrm{d}\,t \\ &\quad + (-1)^\ell \int_0^c \mathbb{1}\{Y_i^{(j)} < -c\}\cdot \ell t^{\ell-1}\,\mathrm{d}\,t \\ &= \int_0^\infty \mathbb{1}\left\{t \leq Y_i^{(j)}\right\}\cdot \ell t^{\ell-1}\,\mathrm{d}\,t + (-1)^\ell \int_0^c \mathbb{1}\{-c \leq Y_i^{(j)} < -t\}\cdot \ell t^{\ell-1}\,\mathrm{d}\,t\end{aligned}$$

$$+(-1)^{\ell}c^{\ell}\,\mathbb{1}\left\{Y_i^{(j)}<-c\right\}. \tag{A.43}$$

By Lemma A.1, Fubini's theorem and (A.27), we have

$$\begin{aligned}
\mathbb{E}\left[\left\{Y_i^{(j)}\vee(-c)\right\}^{\ell}\right] &= (-c)^{\ell}\,\mathbb{P}\left(Y_i^{(j)}<-c\right)+(-1)^{\ell}\int_0^c \mathbb{P}(-c\le Y_i^{(j)}<-t)\cdot \ell t^{\ell-1}\,\mathrm{d}\,t\\
&\quad+\int_0^{\infty}\mathbb{P}\left(t\le Y_i^{(j)}\right)\cdot \ell t^{\ell-1}\,\mathrm{d}\,t\\
&= (-c)^{\ell}\,\mathbb{P}\left(Y_i^{(j)}<-c\right)+(-1)^{\ell}\int_0^c \mathbb{P}(-c\le Y_i^{(j)}<-t)\cdot \ell t^{\ell-1}\,\mathrm{d}\,t\\
&\quad+\int_0^{\infty}\mathbb{P}\left(t\le Y_i^{(j)}\right)\cdot \ell t^{\ell-1}\,\mathrm{d}\,t\\
&= (-c)^{\ell}\left\{1-R_{ij}(\exp(c),1)\right\}\\
&\quad+(-1)^{\ell}\int_0^c\left\{R_{ij}(\exp(c),1)-R_{ij}(\exp(t),1)\right\}\cdot \ell t^{\ell-1}\,\mathrm{d}\,t\\
&\quad+\int_0^{\infty}R_{ij}(\exp(-t),1)\cdot \ell t^{\ell-1}\,\mathrm{d}\,t,
\end{aligned}$$

where Fubini's theorem is used in the first step together with the fact in Lemma A.1 and the second equality follows from (A.27). Since we set $c=-\log a$. By the changes of variables $t=\log u$ and $t=-\log u$ in the first and second integrals, respectively, the expectation can be rewritten as

$$\begin{aligned}
\mathbb{E}\left[\left\{Y_i^{(j)}\vee(-c)\right\}^{\ell}\right] &= (-\log a)^{\ell}\left\{1-R_{ij}(1/a,1)\right\}\\
&\quad+(-1)^{\ell}\int_1^{1/a}\frac{\left\{R_{ij}(1/a,1)-R_{ij}(u,1)\right\}\cdot \ell(\log u)^{\ell-1}}{u}\,\mathrm{d}\,u\\
&\quad+\int_0^1\frac{R_{ij}(u,1)\cdot \ell(-\log u)^{\ell-1}}{u}\,\mathrm{d}\,u\\
&= \int_0^{1/a}\frac{R_{ij}(u,1)\cdot \ell(-\log u)^{\ell-1}}{u}\,\mathrm{d}\,u+(\log a)^{\ell}.
\end{aligned}$$

□

A.3.2 Proof of Eq. (A.30)

Proof. Similarly to (A.42), for $x>0$ and $\ell=1,2$, we have

$$(\log x)^{\ell}=\int_1^{\infty}\frac{\mathbb{1}\{t\le x\}\ell(\log t)^{\ell-1}}{t}\,\mathrm{d}\,t-\int_0^1\frac{\mathbb{1}\{t>x\}\ell(\log t)^{\ell-1}}{t}\,\mathrm{d}\,t. \tag{A.44}$$

Therefore,

$$\begin{aligned}
&\left\{\log\left(q/U_i^{(j)}(q)\vee a\right)\right\}^{\ell}\\
&\quad=\int_1^{\infty}\frac{\mathbb{1}\{t\le[q/U_i^{(j)}(q)\vee a]\}\ell(\log t)^{\ell-1}}{t}\,\mathrm{d}\,t-\int_0^1\frac{\mathbb{1}\{t>[q/U_i^{(j)}(q)\vee a]\}\ell(\log t)^{\ell-1}}{t}\,\mathrm{d}\,t
\end{aligned}$$

$$
\begin{aligned}
&= \int_1^\infty \frac{\mathbb{1}\{t \le q/U_i^{(j)}(q)\}\ell(\log t)^{\ell-1}}{t}\,\mathrm{d}t - \int_a^1 \frac{\mathbb{1}\{a \le q/U_i^{(j)}(q) < t\}\ell(\log t)^{\ell-1}}{t}\,\mathrm{d}t \\
&\quad - \int_a^1 \frac{\mathbb{1}\left\{q/U_i^{(j)}(q) < a\right\}\ell(\log t)^{\ell-1}}{t}\,\mathrm{d}t \\
&= \int_1^\infty \frac{\mathbb{1}\left\{t \le q/U_i^{(j)}(q)\right\}\ell(\log t)^{\ell-1}}{t}\,\mathrm{d}t - \int_a^1 \frac{\mathbb{1}\{a \le q/U_i^{(j)}(q) < t\}\ell(\log t)^{\ell-1}}{t}\,\mathrm{d}t \\
&\quad + (\log a)^\ell\, \mathbb{1}\left\{q/U_i^{(j)}(q) < a\right\}.
\end{aligned}
$$

Therefore, by Fubini's theorem and (A.29), we have

$$
\begin{aligned}
&\mathbb{E}\left[\left\{\log\left(q/U_i^{(j)}(q) \vee a\right)\right\}^\ell\right] \\
&\quad = (\log a)^\ell\, \mathbb{P}\left(q/U_i^{(j)}(q) < a\right) - \int_a^1 \frac{\mathbb{P}(a \le q/U_i^{(j)}(q) < t)\ell(\log t)^{\ell-1}}{t}\,\mathrm{d}t \\
&\qquad + \int_1^\infty \frac{\mathbb{P}\left(t \le q/U_i^{(j)}(q)\right)\ell(\log t)^{\ell-1}}{t}\,\mathrm{d}t \\
&\quad = (\log a)^\ell \left\{1 - q^{-1}C_{ij}(qa^{-1}, q)\right\} - \int_1^{1/a} \frac{[C_{ij}(qa^{-1}, q) - C_{ij}(qt, q)]\ell(-\log t)^{\ell-1}}{qt}\,\mathrm{d}t \\
&\qquad + \int_0^1 \frac{C_{ij}(qt, q)\ell(-\log t)^{\ell-1}}{qt}\,\mathrm{d}t \\
&\quad = \int_0^{1/a} \frac{C_{ij}(qt, q)\ell(-\log t)^{\ell-1}}{qt}\,\mathrm{d}t + (\log a)^\ell,
\end{aligned}
$$

where Fubini's theorem is used in the first step and the second equality follows from (A.29). □

A.3.3 Proof of Eq. (A.33)

Proof. Recall that $0 < a \le 1$. By (A.44) (with the equal sign in the indicator function moved from the first integrand to the second), we have for $i, j \in V$, $i \ne j$ and $\ell = 1, 2$ that

$$
\begin{aligned}
&\left\{-\log\left(\frac{n}{k}\hat{U}_{ti} \wedge a^{-1}\right)\right\}^\ell \\
&\quad = -\int_1^\infty \frac{\mathbb{1}\{t < \left(\frac{n}{k}\hat{U}_{ti} \wedge a^{-1}\right)\}\ell(-\log t)^{\ell-1}}{t}\,\mathrm{d}t + \int_0^1 \frac{\mathbb{1}\{t \ge \left(\frac{n}{k}\hat{U}_{ti} \wedge a^{-1}\right)\}\ell(-\log t)^{\ell-1}}{t}\,\mathrm{d}t \\
&\quad = -\int_1^{1/a} \frac{\mathbb{1}\{a^{-1} < \frac{n}{k}\hat{U}_{ti}\}\ell(-\log t)^{\ell-1}}{t}\,\mathrm{d}t - \int_1^{1/a} \frac{\mathbb{1}\{t < \frac{n}{k}\hat{U}_{ti} \le a^{-1}\}\ell(-\log t)^{\ell-1}}{t}\,\mathrm{d}t \\
&\qquad + \int_{1/k}^1 \frac{\mathbb{1}\left\{t \ge \frac{n}{k}\hat{U}_{ti}\right\}\ell(-\log t)^{\ell-1}}{t}\,\mathrm{d}t \\
&\quad = (\log a)^\ell\, \mathbb{1}\left\{\hat{U}_{ti} > \frac{k}{n}a^{-1}\right\} - \int_1^{1/a} \frac{\mathbb{1}\left\{\frac{k}{n}t < \hat{U}_{ti} \le \frac{k}{n}a^{-1}\right\}\ell(-\log t)^{\ell-1}}{t}\,\mathrm{d}t
\end{aligned}
$$

$$+\int_{1/k}^{1}\frac{\mathbb{1}\{\hat U_{ti}\le \frac{k}{n}t\}\ell(-\log t)^{\ell-1}}{t}\,\mathrm{d}t.$$

Hence, using the equations

$$\mathbb{1}\left\{\frac{k}{n}t<\hat U_{ti}\le\frac{k}{n}a^{-1},\hat U_{tj}\le k/n\right\}=\mathbb{1}\left\{\hat U_{ti}\le\frac{k}{n}a^{-1},\hat U_{tj}\le k/n\right\}-\mathbb{1}\left\{\hat U_{ti}\le\frac{k}{n}t,\hat U_{tj}\le k/n\right\}$$

and

$$\mathbb{1}\left\{\hat U_{ti}>\frac{k}{n}a^{-1},\hat U_{tj}\le k/n\right\}=\mathbb{1}\left\{\hat U_{tj}\le k/n\right\}-\mathbb{1}\left\{\hat U_{ti}\le\frac{k}{n}a^{-1},\hat U_{tj}\le k/n\right\},$$

we get

$$\begin{aligned}
\hat e^{(\ell)}_{n,ij}&=\frac{1}{k}\sum_{t=1}^{n}\left\{-\log\left(\frac{n}{k}\hat U_{ti}\wedge a^{-1}\right)\right\}^{\ell}\mathbb{1}\left\{\hat U_{tj}\le k/n\right\}\\
&=\frac{1}{k}\sum_{t=1}^{n}\left\{(\log a)^{\ell}\,\mathbb{1}\left\{\hat U_{ti}>\frac{k}{n}a^{-1},\hat U_{tj}\le k/n\right\}\right.\\
&\quad-\int_{1}^{1/a}\frac{\mathbb{1}\left\{\frac{k}{n}t<\hat U_{ti}\le\frac{k}{n}a^{-1},\hat U_{tj}\le k/n\right\}\ell(-\log t)^{\ell-1}}{t}\,\mathrm{d}t\\
&\quad\left.+\int_{1/k}^{1}\frac{\mathbb{1}\{\hat U_{ti}\le\frac{k}{n}t,\hat U_{tj}\le k/n\}\ell(-\log t)^{\ell-1}}{t}\,\mathrm{d}t\right\}\\
&=\int_{1/k}^{1/a}\frac{\tilde R_{n,ij}(t,1)\ell(-\log t)^{\ell-1}}{t}\,\mathrm{d}t+(\log a)^{\ell}.
\end{aligned}$$

The proof is complete. □

A.4 Explicit expression of the asymptotic variance

In this section, we derive the explicit expression of the asymptotic variance of the limiting distribution in Theorem 4.8.

For a Hüsler–Reiss MGP distributed random vector $\boldsymbol{Y}$ with variogram matrix $\boldsymbol{\Gamma}=(\gamma_{ij})_{i,j\in V}$, the bivariate tail copula is given in (A.23), while its partial derivatives with respect to x and y are specified by (A.24). We redefine the notations as

$$\dot R_1(x,y;\gamma_{ij})=\dot R^{i}_{ij}(x,y),\qquad\qquad \dot R_2(x,y;\gamma_{ij})=\dot R^{j}_{ij}(x,y).$$

For each pair $(i,j)\in V^2$ with $i\ne j$, the limiting random variable in Theorem 4.8 is a centered Gaussian random variable with variance $\sigma^2_\ell(\gamma_{ij},c)$, which has expression

$$\sigma^2_\ell(\gamma,c)=\int_0^{\exp(c)}\int_0^{\exp(c)}V(x,y;\gamma)\cdot f_\ell(x;\gamma,c)\cdot f_\ell(y;\gamma,c)\,\mathrm{d}x\,\mathrm{d}y,$$

with

$$f_\ell(x;\gamma,c)=\frac{\ell\cdot(-\log x+c)^{\ell-1}}{e^{(\ell),\prime}(\gamma)\cdot x},\qquad x>0,$$

and

$$V(x,y;\gamma_{ij}) = \mathbb{E}\left(\frac{B_{ij}(x,1)+B_{ji}(x,1)}{2}\cdot\frac{B_{ij}(y,1)+B_{ji}(y,1)}{2}\right).$$

Note that by the definition of the process $B_{ij}(x,1)$, we have

$$B_{ij}(x,1) \overset{d}{=} B_{ji}(x,1)$$

for each pair of $(i,j) \in V \times V$ and $i \neq j$. By Eq. (4.2), Eq. (4.4) and the relation that $\dot{R}_1(x,y;\gamma) = \dot{R}_2(y,x;\gamma)$, a direct calculation gives that

$$\begin{aligned}
V(x,y;\gamma) = \frac{1}{2}\Bigg\{ & \left(1-\dot{R}_1(x,1;\gamma)-\dot{R}_1(y,1;\gamma)\right) R(x\wedge y,1;\gamma) \\
& - \left(1-\dot{R}_1(y,1;\gamma)\right)\dot{R}_1(1,x;\gamma)R(y,1;\gamma) \\
& - \left(1-\dot{R}_1(x,1;\gamma)\right)\dot{R}_1(1,y;\gamma)R(x,1;\gamma) + (x\wedge y)\dot{R}_1(x,1;\gamma)\dot{R}_1(y,1;\gamma) \\
& + \dot{R}_1(1,x;\gamma)\dot{R}_1(1,y;\gamma)\Bigg\} \\
& + \frac{1}{2}\Bigg\{ R(x\wedge 1, y\wedge 1;\gamma) - (R_1(x,1;\gamma)+R_1(1,y;\gamma))R(x\wedge 1,y;\gamma) \\
& - (R_1(1,x;\gamma)+R_1(y,1;\gamma))R(x,y\wedge 1;\gamma) + R_1(x,1;\gamma)R_1(y,1;\gamma)R(x,y;\gamma) \\
& + R_1(1,x;\gamma)R_1(y,1;\gamma)(y\wedge 1) + R_1(x,1;\gamma)R_1(1,y;\gamma)(x\wedge 1) \\
& + R_1(1,x;\gamma)R_1(1,y;\gamma)R(1,1;\gamma)\Bigg\}.
\end{aligned}$$

B Additional simulation results

The randomly generated 10-dimensional variogram matrix $\mathbf{\Gamma}_1$ used in the simulation study in Section 5 is

$$\mathbf{\Gamma}_1 = \begin{pmatrix}
0.00 & 0.89 & 0.82 & 0.61 & 0.98 & 1.05 & 0.25 & 0.69 & 1.08 & 0.75 \\
0.89 & 0.00 & 1.91 & 1.40 & 2.08 & 2.16 & 1.07 & 1.50 & 2.05 & 1.64 \\
0.82 & 1.91 & 0.00 & 1.51 & 2.07 & 2.12 & 1.07 & 1.71 & 2.00 & 1.58 \\
0.61 & 1.40 & 1.51 & 0.00 & 1.65 & 1.73 & 0.85 & 1.32 & 1.72 & 1.37 \\
0.98 & 2.08 & 2.07 & 1.65 & 0.00 & 2.22 & 1.23 & 1.84 & 2.13 & 1.74 \\
1.05 & 2.16 & 2.12 & 1.73 & 2.22 & 0.00 & 1.31 & 1.90 & 2.20 & 1.81 \\
0.25 & 1.07 & 1.07 & 0.85 & 1.23 & 1.31 & 0.00 & 0.93 & 1.33 & 1.00 \\
0.69 & 1.50 & 1.71 & 1.32 & 1.84 & 1.90 & 0.93 & 0.00 & 1.83 & 1.45 \\
1.08 & 2.05 & 2.00 & 1.72 & 2.13 & 2.20 & 1.33 & 1.83 & 0.00 & 1.83 \\
0.75 & 1.64 & 1.58 & 1.37 & 1.74 & 1.81 & 1.00 & 1.45 & 1.83 & 0.00
\end{pmatrix}.$$

In Figures 11–13, we report the mean bias, variance, and MSE of the first 10 entries of $\mathbf{\Gamma}_1$ based on the empirical variogram estimator $\hat{\gamma}_{n,ij}^{\mathrm{EMP}}(k)$ and the moment estimators $\hat{\gamma}_{n,ij}^{\mathrm{M},(\ell)}(k,-\log a)$ with $\ell = 1,2$ and $a = 0.25$. The samples are generated from

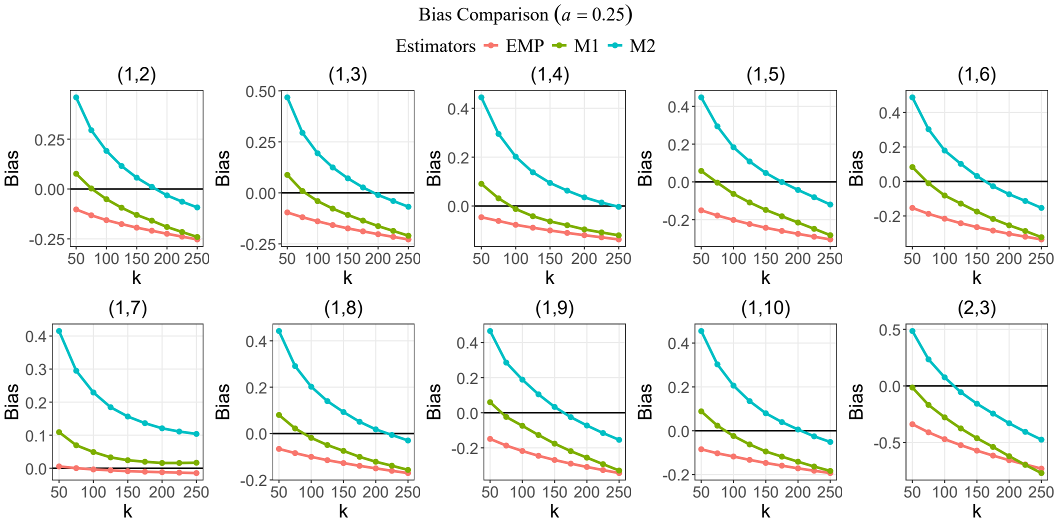


Fig. 11 The mean estimation bias for the first ten elements γ_{ij} of $\mathbf{\Gamma}_1$, based on the empirical variogram estimator $\hat{\gamma}_{n,ij}^{\mathrm{EMP}}$ (EMP), first-order moment estimator $\hat{\gamma}_{n,ij}^{\mathrm{M},(1)}(k, -\log a)$ (M1) and second-order moment estimator $\hat{\gamma}_{n,ij}^{\mathrm{M},(2)}(k, -\log a)$ (M2) with $a = 0.25$ in 300 replications. The random samples with size $n = 1000$ are drawn from the 10−dimensional Hüsler–Reiss distribution with variogram matrix $\mathbf{\Gamma}_1$.

the 10-dimensional Hüsler–Reiss distribution with variogram matrix $\mathbf{\Gamma}_1$. These results are provided here for completeness and to facilitate a more detailed inspection of the component-wise estimation performance.

In addition, the related analysis for samples generated from a 5-dimensional Hüsler–Reiss distribution with parameter matrix $\mathbf{\Gamma}_2$ is presented in Figures 14–16.

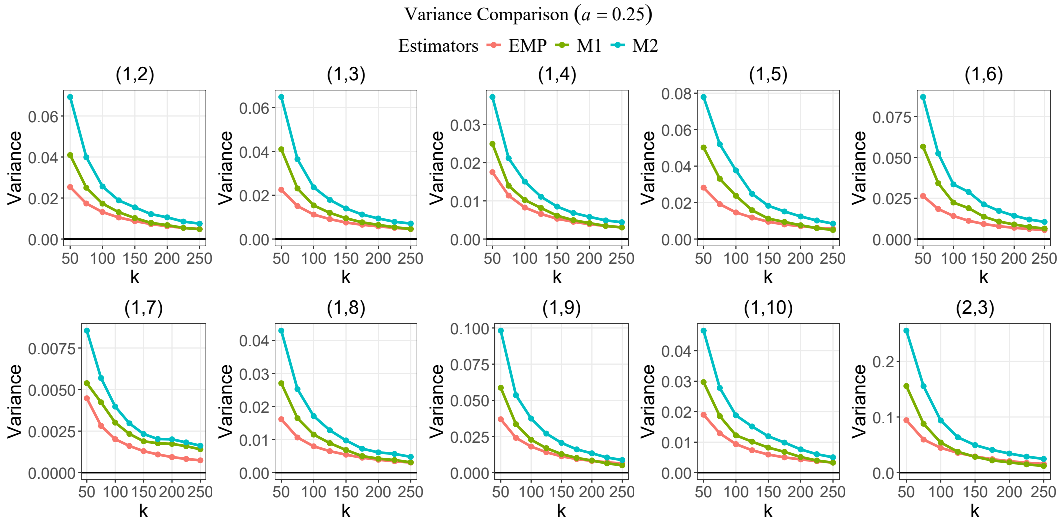


Fig. 12 The mean estimation variance for the first ten elements γ_{ij} of $\mathbf{\Gamma}_1$, based on the empirical variogram estimator $\hat{\gamma}_{n,ij}^{\mathrm{EMP}}$ (EMP), first-order moment estimator $\hat{\gamma}_{n,ij}^{\mathrm{M},(1)}(k, -\log a)$ (M1) and second-order moment estimator $\hat{\gamma}_{n,ij}^{\mathrm{M},(2)}(k, -\log a)$ (M2) with $a = 0.25$ in 300 replications. The random samples with size $n = 1000$ are drawn from the 10−dimensional Hüsler–Reiss distribution with variogram matrix $\mathbf{\Gamma}_1$.

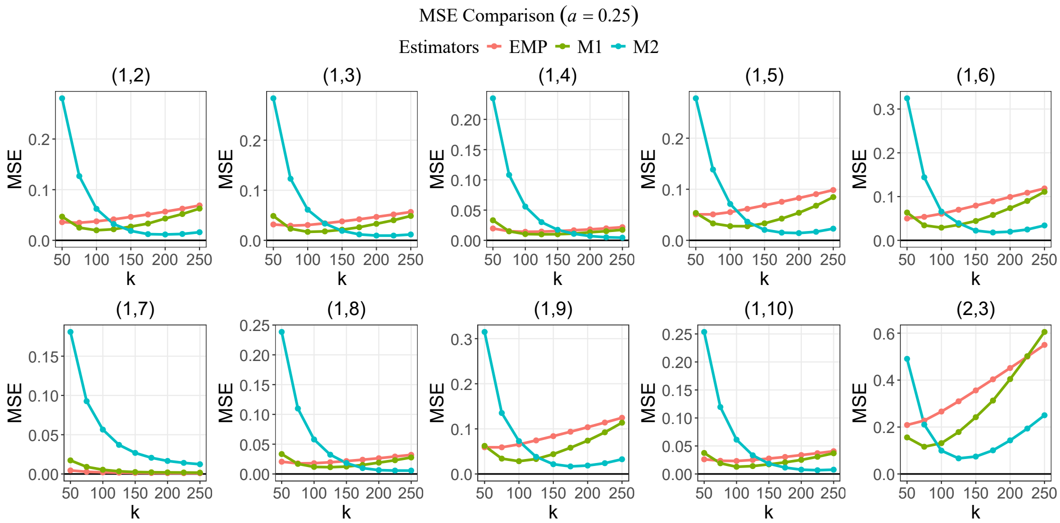


Fig. 13 The MSE for the first ten elements γ_{ij} of $\mathbf{\Gamma}_1$, based on the empirical variogram estimator $\hat{\gamma}_{n,ij}^{\mathrm{EMP}}$ (EMP), first-order moment estimator $\hat{\gamma}_{n,ij}^{\mathrm{M},(1)}(k, -\log a)$ (M1) and second-order moment estimator $\hat{\gamma}_{n,ij}^{\mathrm{M},(2)}(k, -\log a)$ (M2) with $a = 0.25$ in 300 replications. The random samples with size $n = 1000$ are drawn from the 10−dimensional Hüsler–Reiss distribution with variogram matrix $\mathbf{\Gamma}_1$.

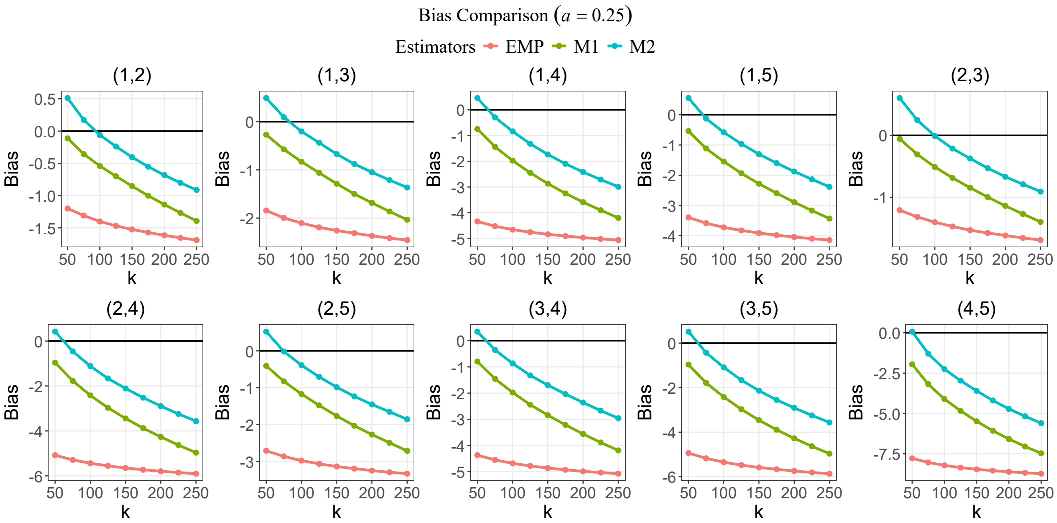


Fig. 14 The mean bias for γ_{ij} of $\mathbf{\Gamma}_2$ with $i, j \in V$ and $i \neq j$, based on the empirical variogram estimator $\hat{\gamma}_{n,ij}^{\mathrm{EMP}}$ (EMP), first-order moment estimator $\hat{\gamma}_{n,ij}^{\mathrm{M},(1)}(k, -\log a)$ (M1) and second-order moment estimator $\hat{\gamma}_{n,ij}^{\mathrm{M},(2)}(k, -\log a)$ (M2) with $a = 0.25$ in 300 replications. The random samples with size $n = 1000$ are drawn from the 5–dimensional Hüsler–Reiss distribution with variogram matrix $\mathbf{\Gamma}_2$.

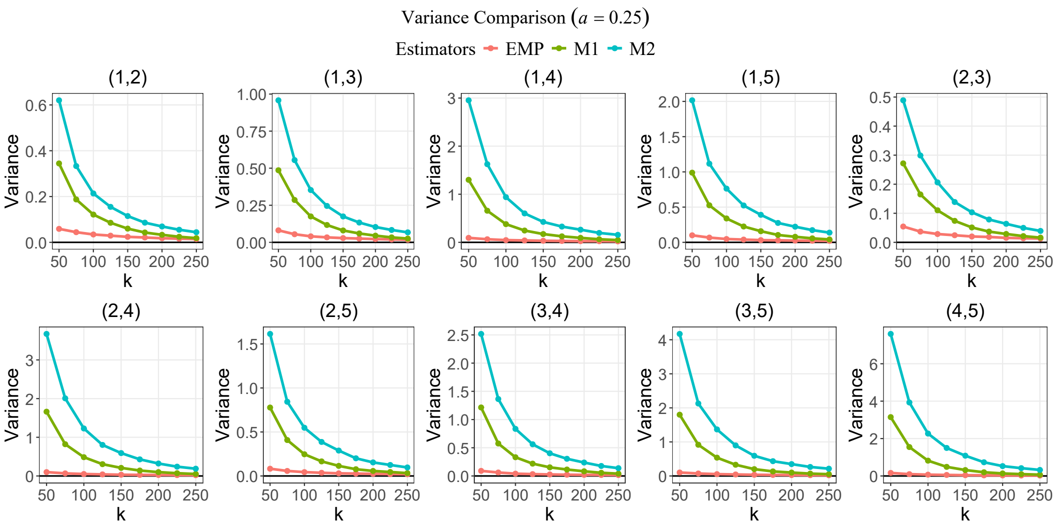


Fig. 15 The mean estimation variance for γ_{ij} of $\mathbf{\Gamma}_2$ with $i, j \in V$ and $i \neq j$, based on the empirical variogram estimator $\hat{\gamma}_{n,ij}^{\mathrm{EMP}}$ (EMP), first-order moment estimator $\hat{\gamma}_{n,ij}^{\mathrm{M},(1)}(k, -\log a)$ (M1) and second-order moment estimator $\hat{\gamma}_{n,ij}^{\mathrm{M},(2)}(k, -\log a)$ (M2) with $a = 0.25$ in 300 replications. The random samples with size $n = 1000$ are drawn from the 5–dimensional Hüsler–Reiss distribution with variogram matrix $\mathbf{\Gamma}_2$.

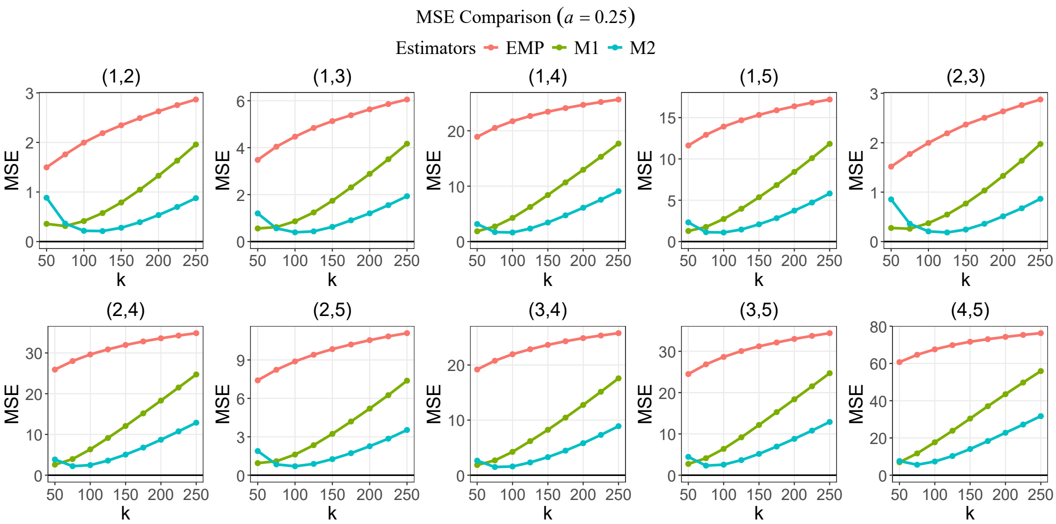


Fig. 16 The MSE for γ_{ij} of $\mathbf{\Gamma}_2$ with $i, j \in V$ and $i \neq j$, based on the empirical variogram estimator $\hat{\gamma}_{n,ij}^{\mathrm{EMP}}$ (EMP), first-order moment estimator $\hat{\gamma}_{n,ij}^{\mathrm{M},(1)}(k, -\log a)$ (M1) and second-order moment estimator $\hat{\gamma}_{n,ij}^{\mathrm{M},(2)}(k, -\log a)$ (M2) with $a = 0.25$ in 300 replications. The random samples with size $n = 1000$ are drawn from the 5–dimensional Hüsler–Reiss distribution with variogram matrix $\mathbf{\Gamma}_2$.