Extremal Dependence of Moving Average Processes Driven by Exponential-Tailed Lévy Noise

Zhongwei Zhang[*], David Bolin[†], Sebastian Engelke[*], Raphaël Huser[†]

Abstract

Moving average processes driven by exponential-tailed Lévy noise are important extensions of their Gaussian counterparts in order to capture deviations from Gaussianity, more flexible dependence structures, and sample paths with jumps. Popular examples include non-Gaussian Ornstein–Uhlenbeck processes and type G Matérn stochastic partial differential equation random fields. This paper is concerned with the open problem of determining their extremal dependence structure. We leverage the fact that such processes admit approximations on grids or triangulations that are used in practice for efficient simulations and inference. These approximations can be expressed as special cases of a class of linear transformations of independent, exponential-tailed random variables, that bridge asymptotic dependence and independence in a novel, tractable way. This result is of independent interest since models that can capture both extremal dependence regimes are scarce and the construction of such flexible models is an active area of research. This new fundamental result allows us to show that the integral approximation of general moving average processes with exponential-tailed Lévy noise is asymptotically independent when the mesh is fine enough. Under mild assumptions on the kernel function we also derive the limiting residual tail dependence function. For the popular exponential-tailed Ornstein–Uhlenbeck process we prove that it is asymptotically independent, but with a different residual tail dependence function than its Gaussian counterpart. Our results are illustrated through simulation studies.

Keywords: exponential tail; extremal dependence; moving average process; non-Gaussian OU process; spatial extremes; type G Matérn SPDE field

[*]Research Center for Statistics, University of Geneva, 1205 Geneva, Switzerland

[†]Statistics Program, CEMSE Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia

1 Introduction

Moving average processes, also referred to as process convolutions, are popular and natural constructions for non-stationary and non-Gaussian processes that are widely applied in spatial statistics (Higdon, 2002; Cressie & Pavlicová, 2002; Rodrigues & Diggle, 2010; Ver Hoef & Peterson, 2010). They are defined as

$$u(\boldsymbol{s}) = \int_{\mathcal{D}} G(\|\boldsymbol{s} - \boldsymbol{t}\|)\mathcal{M}(\mathrm{d}\boldsymbol{t}), \tag{1}$$

where $\mathcal{D}$ is a Borel subset of $\mathbb{R}^d$ that possibly can depend on $\boldsymbol{s}$, $G : \mathbb{R}_+ \to \mathbb{R}_+$ is a measurable $\mathcal{M}$-integrable function, and $\mathcal{M}$ is a random measure (Kallenberg, 2017). A classical example in the temporal domain $d = 1$ is the non-Gaussian Ornstein–Uhlenbeck (OU) process developed in Barndorff-Nielsen & Shephard (2001) to model stochastic volatility in financial econometrics; see also an application of this model to longitudinal data in Asar et al. (2020).

Another related approach for constructing non-Gaussian processes is via a stochastic partial differential equation (SPDE) (Barndorff-Nielsen & Shephard, 2001; Bolin, 2014; Bolin & Wallin, 2020). The stationary solution (if it exists) to such an SPDE admits an integral representation of the form (1). In the spatial domain $\mathbb{R}^2$, an important instance of these constructions are the type G Matérn SPDE random fields (Bolin, 2014; Bolin & Wallin, 2020), a non-Gaussian extension of the popular Gaussian Matérn SPDE random fields (Lindgren et al., 2011).

The key advantage of the SPDE-based process formulation is that one can approximate its solution by using the finite element method to obtain sparsity in the resulting precision matrix (or inverse of the dispersion matrix in the non-Gaussian case), thus achieving computationally efficient simulation and inference (Lindgren et al., 2011; Bolin, 2014; Bolin & Wallin, 2020). This finite element approximation has the form of a linear transformation $u_n(\boldsymbol{s}) = \sum_{i=1}^{n} a_i(\boldsymbol{s})Y_i$, where the coefficients $a_i(\boldsymbol{s}) \geq 0$ are determined by the constructed basis functions on the triangulation with n mesh nodes, and Y_i are independent Gaussian or non-Gaussian random variables, depending on the random measure $\mathcal{M}$. Non-Gaussian SPDE models are widely used in applications whenever the assumption of Gaussianity is too stringent. As an illustration, the left panel of Figure 1 shows the triangulation constructed in Bolin & Wallin (2020) for Argo float data, a data set that consists of measurements of seawater temperature and salinity in the global ocean and that has motivated the use of non-Gaussian processes due to its non-Gaussian features such as skewness and heavier
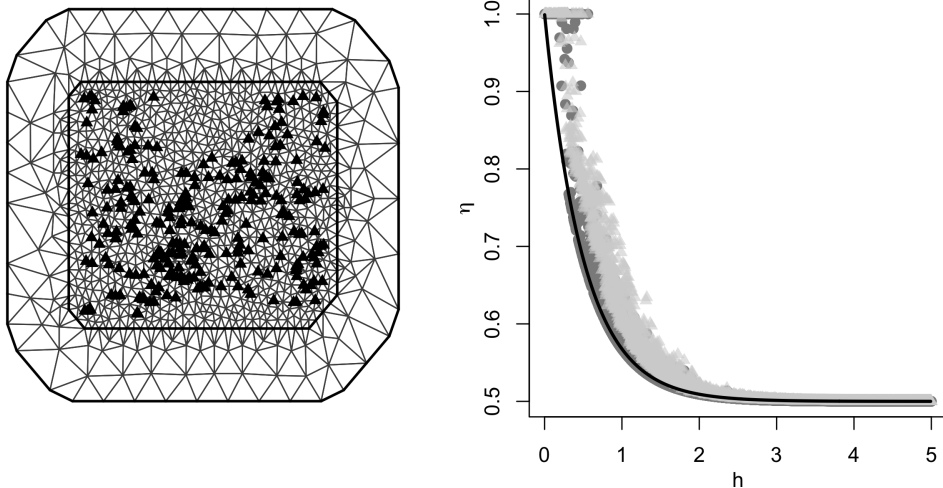
Figure 1: Observation locations and triangulation mesh used for Argo float data in Bolin & Wallin (2020) (left panel), and its induced residual tail dependence coefficients η for all pairs of locations for the finite element (light-grey triangle points) and integral approximations (dark-grey round points) of a type G Matérn SPDE model, and their limiting function (black line).

tails than Gaussian distributions (Kuusela & Stein, 2018).

The dependence structure of SPDE and integral-based process constructions has mostly been studied in terms of their Pearson's correlation. This linear dependence measure does, however, not fully describe the process for non-Gaussian distributions. In fact, the correlation function of these processes is the same regardless of whether Gaussian or non-Gaussian noise is used. Extremal (or tail) dependence describes the strength of dependence in the joint upper or lower tail of a multivariate distribution. It is crucial for risk assessment as it quantifies whether the largest realizations at different locations occur simultaneously. This paper is concerned with the open problem of determining the extremal dependence structure of popular non-Gaussian processes and their linear approximations that are used in practice.

Let (X_1, X_2) be a random vector with marginal distribution functions F_{X_1} and F_{X_2}, respectively. A commonly used measure of extremal dependence is the (upper) tail dependence coefficient (Coles et al., 1999)

$$\chi = \lim_{q \uparrow 1} \chi(q) = \lim_{q \uparrow 1} \Pr(F_{X_1}(X_1) > q \mid F_{X_2}(X_2) > q), \tag{2}$$

which satisfies $\chi \in [0, 1]$ provided that this limit exists. Since the extremal dependence in the lower tail can be obtained by negating the random vector, here we only focus on the upper tail. We call X_1 and X_2 asymptotically dependent if $\chi > 0$, and asymptotically independent otherwise.

3

In the latter case, the residual tail dependence coefficient η (Ledford & Tawn, 1996) measures the second-order extremal dependence behavior and is defined as

$$\eta = \lim_{q\uparrow 1} \frac{\log(1-q)}{\log \Pr(F_{X_1}(X_1) > q, F_{X_2}(X_2) > q)}, \tag{3}$$

provided that this limit exists. For $\chi = 0$, the coefficient $1/\eta$ describes the rate of convergence of $\chi(q) \to 0$ as $q \to 1$.

The extremal dependence structure of continuously indexed processes of the form (1) is challenging to derive due to the lack of analytical expressions of the induced multivariate distribution or density functions. We therefore first consider a discretely indexed model of linear transformations

$$\begin{cases} X_1 = a_{11}Y_1 + \cdots + a_{1n}Y_n, \\ X_2 = a_{21}Y_1 + \cdots + a_{2n}Y_n, \end{cases} \tag{4}$$

where $Y_i, i = 1, \ldots, n$, are independent and have exponential tails with the same index, and with coefficients $a_{ji} \geq 0, j = 1, 2, i = 1, \ldots, n$. This model is motivated by the finite element approximation of the Matérn SPDE fields and integral approximation of general moving average processes. For example, if we approximate (1) by $u_n(\boldsymbol{s}) = \sum_{i=1}^n G(\|\boldsymbol{s} - \boldsymbol{t}_i\|)\mathcal{M}(D_i)$ for a partition $\{D_i\}$ of $\mathcal{D}$ and $\boldsymbol{t}_i \in D_i$, and let $X_j = u_n(\boldsymbol{s}_j)$, then $a_{ji} = G(\|\boldsymbol{s}_j - \boldsymbol{t}_i)$ and $Y_i = \mathcal{M}(D_i)$. The exponential tail of the noise variables appears naturally in the commonly used non-Gaussian processes. Under mild assumptions we show that X_1 and X_2 are asymptotically dependent if we have $\arg\max_{i\in\{1,\ldots,n\}} a_{1i} = \arg\max_{i\in\{1,\ldots,n\}} a_{2i}$, and they are asymptotically independent if $\arg\max_{i\in\{1,\ldots,n\}} a_{1i} \cap \arg\max_{i\in\{1,\ldots,n\}} a_{2i} = \emptyset$. This novel result is of independent interest since it provides a tractable way to bridge asymptotic dependence and independence by simply varying the coefficients in (4). It thus gives a partial answer to the second open problem raised in Nolde & Zhou (2021) who ask how to build flexible parametric models that bridge both extremal dependence regimes.

With this fundamental building block, we then study the extremal dependence structure of general moving average processes with exponential-tailed noise. We show that under mild assumptions on the kernel function and the noise, the integral approximation of such processes is asymptotically independent when the mesh is fine enough, and we derive the limiting residual tail dependence func-

tion. Moreover, we prove that the exponential-tailed non-Gaussian OU process (Barndorff-Nielsen & Shephard, 2001) is asymptotically independent, but with a different residual tail dependence function than its Gaussian counterpart. As for the Argo float measurements data application, we consider the normal inverse Gaussian (NIG) Matérn SPDE fields, which are a specific class of type G Matérn SPDE fields. The right panel of Figure 1 shows the residual tail dependence coefficients of the finite element and integral approximations, and its theoretical limiting function as the triangulation mesh size goes to zero. We conduct additional empirical studies in Section 4.3 and Appendix D to illustrate our results.

For moving average processes of the form (1), Opitz (2018) studied the extremal dependence structure when the kernel function G is an indicator function, while Krupskii & Huser (2022) studied the extremal dependence structure when $\mathcal{M}$ is a Cauchy random measure; for other cases, the extremal dependence structure remains unknown. Our results are thus interesting since they motivate new constructions of non-Gaussian moving average processes and SPDE-based processes for extremes.

2 Preliminaries on Exponential-Tailed Functions

A function $h : \mathbb{R} \to \mathbb{R}_+$ with $\lim_{x \to \infty} h(x) = 0$ is said to have an exponential tail with index $\beta \geq 0$, denoted as $h \in \mathscr{L}_\beta$, if

$$\lim_{x \to \infty} h(x-t)/h(x) = \exp(t\beta), \quad t \in \mathbb{R}.$$

A univariate distribution function F is said to have an exponential tail if its survival function $\bar{F}$ satisfies $\bar{F} \in \mathscr{L}_\beta$ for some $\beta \geq 0$. This is different from the conventional notation $F \in \mathscr{L}_\beta$, where the set $\mathscr{L}_\beta$ is restricted to be the family of all exponential-tailed distribution functions. Our definition of $\mathscr{L}_\beta$ for general, nonnegative functions allows us to also cover exponential-tailed density functions. If $\bar{F} \in \mathscr{L}_\beta$ is such that we further have $\lim_{x \to \infty} \overline{F * F}(x)/\bar{F}(x) = M < \infty$, where $*$ denotes the convolution operator, then F is called convolution tail equivalent, denoted by $\bar{F} \in \mathscr{S}_\beta$. Clearly, $\mathscr{S}_\beta \subset \mathscr{L}_\beta$.

A function $g : \mathbb{R}_+ \to \mathbb{R}_+$ is called regularly varying at ∞ with index $\rho \in \mathbb{R}$ if for any $t > 0$, $\lim_{x \to \infty} g(tx)/g(x) = t^\rho$, in which case we write $g \in \mathrm{RV}_\rho$. If $\rho = 0$, g is called slowly varying. Exponential-tailed and regularly varying functions are intimately related to each other, since for

$\beta \geq 0$, $h \in \mathscr{L}_\beta$ if and only if $h \circ \log \in \mathrm{R}V_{-\beta}$. Using this relationship, we obtain a similar representation of exponential-tailed distribution functions to Karamata's representation of regularly varying functions (see Corollary 2.1 in Resnick (2007)), namely for $\bar{F} \in \mathscr{L}_\beta$, we have

$$\bar{F}(x) = a(x) \exp \Big\{ -\int_0^x \beta(v)\mathrm{d}v \Big\}, \quad x \in \mathbb{R}, \tag{5}$$

where $a(x) \to a \in (0, \infty)$ and $\beta(x) \to \beta$ as $x \to \infty$. This will be repeatedly used in our proofs.

Two prominent examples of exponential-tailed distributions are the generalized inverse Gaussian (GIG) distribution and the generalized hyperbolic (GH) distribution (Barndorff-Nielsen, 1977; Prause, 1999; McNeil et al., 2005). A random variable R is said to have a GIG distribution, denoted as $R \sim \mathrm{GIG}(\lambda, \tau, \psi)$, if its Lebesgue density is

$$f_{\mathrm{GIG}}(x) = \Big(\frac{\psi}{\tau}\Big)^{\lambda/2} \frac{x^{\lambda-1}}{2K_\lambda(\sqrt{\tau\psi})} \exp \Big\{ -\frac{1}{2}\Big(\frac{\tau}{x} + \psi x\Big) \Big\}, \quad x > 0,$$

where K_λ is the modified Bessel function of the second kind with index λ. If a random variable Z has the stochastic representation as a normal mean-variance mixture, i.e.,

$$Z = \mu + \gamma R + \sqrt{R}W, \quad \mu, \gamma \in \mathbb{R},$$

where W is a standard normal random variable and $R \sim \mathrm{GIG}(\lambda, \tau, \psi)$, then Z is said to have the GH distribution, denoted as $Z \sim \mathrm{GH}(\lambda, \tau, \psi, \mu, \gamma)$, and its Lebesgue density is given in Appendix E. The admissible parameter values in both the GIG and GH distributions are $\lambda < 0, \tau > 0, \psi \geq 0$, or $\lambda = 0, \tau > 0, \psi > 0$, or $\lambda > 0, \tau \geq 0, \psi > 0$. The special cases $\psi = 0$ and $\tau = 0$ should be understood as limiting cases. Another limiting case, when R is degenerate and thus Z follows a Gaussian distribution, is not considered in this paper as Z does not have an exponential tail in this case.

It is easy to see that the GIG density function (thus, also its distribution function) has exponential tails. The GH density and distribution functions also have exponential tails and the details are given in Appendix E. Furthermore, the GIG and GH distributions are convolution tail equivalent if and only if $\lambda < 0$; see Embrechts & Goldie (1982) for more details. The GIG and GH distributions are infinitely divisible, and their associated Lévy processes are widely used in finance; see

Eberlein (2001) and McNeil et al. (2005) for an overview. The main examples in Barndorff-Nielsen & Shephard (2001) are the GIG Lévy processes, while the non-Gaussian noise considered in Bolin (2014), Wallin & Bolin (2015), and Bolin & Wallin (2020) is based on certain subclasses of the GH distribution.

3   Linear Transformations

3.1   General Framework and Outline

In this section, we focus on the extremal dependence of the random vector $\boldsymbol{X} = (X_1, X_2)$ defined in (4) with $a_{ji} \geq 0$, $\max a_{1i} > 0$, $\max a_{2i} > 0$, and $a_{1i} + a_{2i} > 0$, $j = 1, 2, i = 1, \ldots, n$. We work under the following assumptions throughout the paper unless otherwise stated:

A.1 $Y_i, i = 1, \ldots, n$, are mutually independent with identical distribution function F_Y such that $\bar{F}_Y \in \mathscr{L}_\beta$ for some $\beta > 0$;

A.2 F_Y is absolutely continuous with density f_Y.

The assumption of identical distribution functions in condition A.1 can be relaxed to different exponential-tailed distribution functions with the same index β, i.e., $Y_i \sim F_{Y_i}$ such that $\bar{F}_{Y_i} \in \mathscr{L}_\beta, \beta > 0, i = 1, \ldots, n$, but for the sake of simplicity we assume here that they have a common distribution F_Y.

Although the dependence structure in model (4) is defined via simple linear transformations, surprisingly, nontrivial extremal dependence structures can be obtained. Specifically, it turns out that the extremal dependence of $\boldsymbol{X} = (X_1, X_2)$ mainly depends on the largest coefficients among $a_{1i}, i = 1, \ldots, n$, and $a_{2i}, i = 1, \ldots, n$. We will show that the components of the vector $\boldsymbol{X}$ are asymptotically dependent if there is equality of the maximizing sets

$$\underset{i \in \{1, \ldots, n\}}{\arg \max} \, a_{1i} = \underset{i \in \{1, \ldots, n\}}{\arg \max} \, a_{2i}. \tag{6}$$

We discuss this case in Section 3.2. On the other hand, there is asymptotic independence if these

sets have an empty intersection

$$\underset{i\in\{1,\ldots,n\}}{\arg\max}\, a_{1i} \cap \underset{i\in\{1,\ldots,n\}}{\arg\max}\, a_{2i} = \emptyset, \tag{7}$$

which is discussed in Section 3.3. The delicate boundary case where the two maximizing sets $\arg\max_{i\in\{1,\ldots,n\}} a_{1i}$ and $\arg\max_{i\in\{1,\ldots,n\}} a_{2i}$ are not equal but have a non-empty intersection, is discussed in Section 3.4.

It is interesting to note that this behaviour is in sharp contrast to the case of heavy-tailed linear (e.g., Gnecco et al., 2021) and max-linear (e.g., Wang & Stoev, 2011) models, where the random variables Y_i have common survival function $\bar{F}_Y \in \mathrm{R}V_{-\beta}$. In this more classical case, only asymptotic dependence (or complete independence) may arise and all coefficients a_{ji} contribute to the corresponding tail dependence coefficient χ.

3.2 Asymptotic Dependence

Here we show that $\boldsymbol{X} = (X_1, X_2)$ defined via (4) is asymptotically dependent when (6) holds, and we give the explicit expression of its tail dependence coefficient. All proofs are given in the Appendix.

Proposition 3.1. *Let $Y_1, \ldots, Y_n$ be a sequence of random variables satisfying the condition A.1 with $\bar{F}_Y \in \mathscr{L}_\beta$ for some $\beta > 0$, and let $\boldsymbol{X} = (X_1, X_2)$ be constructed as in (4). If the set equality in (6) holds, then X_1 and X_2 are asymptotically dependent and the tail dependence coefficient of $\boldsymbol{X}$ can be expressed as*

$$\chi = \mathrm{E}\left[\min\left\{\frac{\exp(\beta Z_1)}{M_{Z_1}(\beta)}, \frac{\exp(\beta Z_2)}{M_{Z_2}(\beta)}\right\}\right]$$

where $Z_1 = \sum_{i\notin I_{max}} a_{1i} Y_i / a_{1max}$, $Z_2 = \sum_{i\notin I_{max}} a_{2i} Y_i / a_{2max}$, and M_{Z_j} is the moment generating function of $Z_j, j = 1, 2$. Here the index set I_{max} is the common set of maximizers in (6) and $a_{jmax} = \max_{i=1,\ldots,n} a_{ji}, j = 1, 2$.

In this asymptotically dependent case where (6) holds, there is a link between our discrete model (4) and the random scale constructions considered in Huser & Wadsworth (2019) and Engelke et al. (2019). As shown in the proof of Proposition 3.1, we can assume without loss of generality that $a_{1max} = a_{2max} = 1$. We can then rewrite model (4) as $X_1 = Z_c + Z_1$ and $X_2 = Z_c + Z_2$, where

8

$Z_c = \sum_{i \in I_{\max}} Y_i$, and Z_1 and Z_2 are the same as in Proposition 3.1. Further notice that Z_c has an exponential tail if and only if $\exp(Z_c)$ has a regularly varying tail. Hence, if one is interested in the extremal dependence structure of the random vector $(\exp(X_1), \exp(X_2))$, which is the same as that of (X_1, X_2) since extremal dependence is invariant to monotonically increasing marginal transformations, then this problem falls into the setting of random scale constructions. More precisely, we have $(\exp(X_1), \exp(X_2)) = \exp(Z_c)(\exp(Z_1), \exp(Z_2))$ with $\exp(Z_c)$ being the shared random component. An application of Proposition 1 in Engelke et al. (2019) yields the asymptotic dependence of $(\exp(X_1), \exp(X_2))$ and thus also (X_1, X_2), and gives the same tail dependence coefficient as in Proposition 3.1. This alternative proof transforms the sum of exponential-tailed random variables into a product of regularly varying random variables and exploits the properties of regularly varying functions to derive the extremal dependence structure. In comparison, our proof in Proposition 3.1 treats linear transformations of a vector of independent exponential-tailed random variables in a direct manner and reveals many asymptotic properties of exponential-tailed distribution functions. This might be of independent interest and these results could be useful in other contexts.

To further investigate the expression of the tail dependence coefficient χ in Proposition 3.1, we now consider the specific example where the Y_i are GH distributed. Suppose $Y_1, Y_2 \sim \mathrm{GH}(\lambda, \tau, \psi, \mu, \gamma)$ with $\psi > 0, \gamma = 0$, then Appendix E shows that their densities and survival functions have exponential tails with the same index, i.e., $f_{Y_i}, \bar{F}_{Y_i} \in \mathscr{L}_{\sqrt{\psi}}$. Proposition 3.1 yields the following result.

Example 1. *Let Y_1, Y_2 be independent and have a common distribution $GH(\lambda, \tau, \psi, \mu, \gamma)$ with $\psi > 0, \gamma = 0$. Let $X_1 = Y_1 + a_{12}Y_2$, $X_2 = Y_1 + a_{22}Y_2$ with $0 \le a_{12}, a_{22} < 1$. Then (X_1, X_2) is asymptotically dependent with tail dependence coefficient*

$$\chi = \frac{\int_{(a_{22}-a_{12})y \le c} \exp(a_{22}\sqrt{\psi}y)F_{Y_1}(\mathrm{d}y)}{M_{Y_1}(a_{22}\sqrt{\psi})} + \frac{\int_{(a_{22}-a_{12})y > c} \exp(a_{12}\sqrt{\psi}y)F_{Y_1}(\mathrm{d}y)}{M_{Y_1}(a_{12}\sqrt{\psi})},$$

where $c = \{\log M_{Y_1}(a_{22}\sqrt{\psi}) - \log M_{Y_1}(a_{12}\sqrt{\psi})\}/\sqrt{\psi}$, and M_{Y_1} is the moment generating function of Y_1.

Clearly, the tail dependence coefficient χ in Example 1 is larger than zero. Also note that when $a_{12} = a_{22}$, $X_1 = X_2$ and thus $\chi = 1$. In fact, one can show that when $a_{12} \ne a_{22}$, χ is always strictly less than 1; see Proposition E.1 in Appendix E. To further investigate the properties of
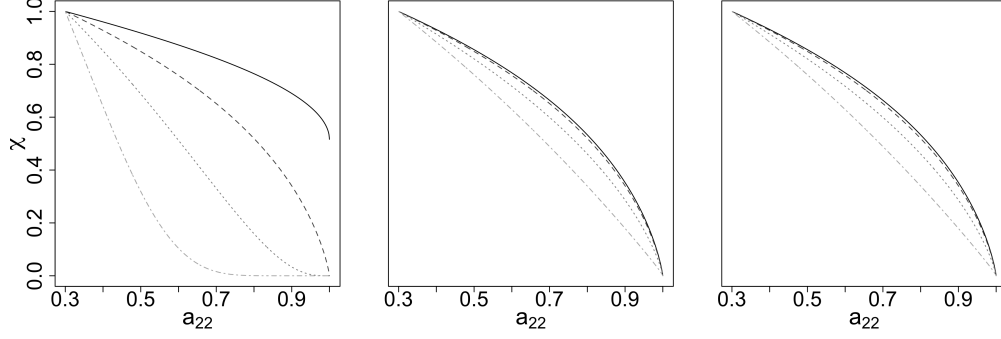
Figure 2: Values of χ as a function of a_{22} for fixed $a_{12} = 0.3$ and parameters $\tau = 1, \psi = 1, \lambda = -0.5, 1, 5, 30$ (left), $\lambda = 1, \psi = 1, \tau = 0.5, 1, 5, 30$ (center), and $\lambda = 1, \tau = 1, \psi = 0.5, 1, 5, 30$ (right) in Example 1, where lines in each plot are ordered from highest to lowest for increasing values of λ, τ and ψ, respectively.

the tail dependence coefficient χ expressed in Example 1, we now assume $0 \le a_{12} < a_{22} < 1$ and examine its limit as $a_{22} \uparrow 1$. This is interesting because in the following subsection we see that when $a_{12} < 1 < a_{22}$, which implies $\arg\max_{i\in\{1,2\}} a_{1i} = 1$ and $\arg\max_{i\in\{1,2\}} a_{2i} = 2$, then X_1 and X_2 are asymptotically independent and necessarily $\chi = 0$. Hence, the investigation of $\lim_{a_{22}\uparrow 1} \chi$ answers the question of whether χ smoothly transitions from asymptotic dependence to independence. It turns out that this is true only in some cases.

Proposition 3.2. *Let χ be as in Example 1 with $0 \le a_{12} < a_{22} < 1$. Then*

$$\lim_{a_{22}\uparrow 1} \chi = \begin{cases} \dfrac{\int_{-\infty}^{c^*} \exp(\sqrt{\psi}y)F_{Y_1}(\mathrm{d}y)}{M_{Y_1}(\sqrt{\psi})} + \dfrac{\int_{c^*}^{\infty} \exp(a_{12}\sqrt{\psi}y)F_{Y_1}(\mathrm{d}y)}{M_{Y_1}(a_{12}\sqrt{\psi})}, & \text{if } \lambda < 0, \\[4mm] 0, & \text{if } \lambda \ge 0, \end{cases}$$

where $c^* = \lim_{a_{22}\uparrow 1} c/(a_{22} - a_{12}) = \dfrac{\log\{M_{Y_1}(\sqrt{\psi})\} - \log\{M_{Y_1}(a_{12}\sqrt{\psi})\}}{(1-a_{12})\sqrt{\psi}} < \infty$ when $\lambda < 0$.

The result in Proposition 3.2 indicates that when $\lambda < 0$, which implies that the distributions of Y_1, Y_2 are convolution tail equivalent (Pakes, 2004), there is a discontinuity in χ when a_{22} tends to 1 from below. To illustrate how fast the χ coefficient in Example 1 tends to its limit, we set $a_{12} = 0.3$, and plot χ against a_{22} for various values of λ, τ, ψ in Figure 2. The results show that λ is the most important parameter that regulates the decay rate of χ as a_{22} increases, and the larger λ the faster χ decays. This indicates that the normal inverse Gaussian distribution, which is a subclass of the GH distribution when λ is fixed to -0.5, might not be a good option to use for modeling extreme events in the presence of both asymptotic dependence and independence, since there is no smooth

10

transition between the two. On the other hand, the variance gamma distribution, another subclass of the GH distribution with $\lambda > 0$, might be a better candidate to consider as there is a smooth transition.

3.3   Asymptotic Independence

In this section we show that $\boldsymbol{X}$ defined in (4) is asymptotically independent when (7) holds. We further give the explicit expression of the residual tail dependence coefficient η; recall its definition (3) in the introduction.

The idea of this proof is to first expand the bivariate random vector $\boldsymbol{X}$ to n dimensions, then use the geometric approach (Nolde, 2014; Nolde & Wadsworth, 2022) to cast the computation of the residual tail dependence coefficient η as a convex optimization problem, and finally solve this optimization problem. We refer to the Supplementary Material (Section 1) for the proof and additional related results.

Proposition 3.3. *Let* $Y_1, \ldots, Y_n$ *be a sequence of random variables satisfying A.1 and A.2. Let* $\boldsymbol{X} = (X_1, X_2)$ *be constructed as in (4). Then the residual tail dependence coefficient is*

$$\eta = \Big[\min_{\substack{i,j=1,\ldots,n \\ i \neq j}} \Big\{ \frac{|\tilde{a}_{2i} - \tilde{a}_{1i}| + |\tilde{a}_{2j} - \tilde{a}_{1j}|}{|\tilde{a}_{2i}\tilde{a}_{1j} - \tilde{a}_{1i}\tilde{a}_{2j}|}, \max(1/\tilde{a}_{1i}, 1/\tilde{a}_{2i}), \max(1/\tilde{a}_{1j}, 1/\tilde{a}_{2j}) \Big\} \Big]^{-1},$$

where $\tilde{a}_{ji} = a_{ji}/(\max_{r \in \{1,\ldots,n\}} a_{jr}), j = 1, 2, i = 1, \ldots, n$, *and the first term in the minimum is set to* $+\infty$ *whenever* $\tilde{a}_{2i}\tilde{a}_{1j} - \tilde{a}_{1i}\tilde{a}_{2j} = 0$. *Moreover, if (7) holds then* $\eta < 1$, *and necessarily* X_1 *and* X_2 *are asymptotically independent. If (7) does not hold, then* $\eta = 1$.

We remark that when (7) does not hold, more assumptions are needed to determine the extremal dependence regime of $\boldsymbol{X}$, i.e., whether the tail dependence coefficient $\chi > 0$ or $\chi = 0$; we refer to Section 3.2 for the asymptotic dependence case and Section 3.4 for the boundary case.

We now consider the specific case that is similar to Example 1 and where a direct application of Proposition 3.3 yields a simple form of η.

Example 2. *Let* Y_1, Y_2 *be independent and have a common distribution* $GH(\lambda, \tau, \psi, \mu, \gamma)$ *with* $\psi > 0, \gamma = 0$. *Let* $X_1 = Y_1 + a_{12}Y_2, X_2 = a_{21}Y_1 + Y_2$ *with* $0 \leq a_{12}, a_{21} \leq 1$. *Then, the resid-*

ual tail dependence coefficient of $\boldsymbol{X} = (X_1, X_2)$ is

$$\eta = \frac{1 - a_{12}a_{21}}{2 - a_{12} - a_{21}}.$$

Furthermore, if a_{12}, a_{21} are both strictly less than 1, then $\boldsymbol{X}$ is asymptotically independent.

An interesting observation in Proposition 3.3 and Example 2 is that the expression of η only depends on the coefficients a_{ji}. In other words, this means that as long as we do not change the coefficients a_{ji} in the model (4) and $Y_1, \ldots, Y_n$ satisfy conditions A.1 and A.2, then we always obtain the same residual tail dependence coefficient η, regardless of coefficient β or the exact distribution of Y_i. Another observation of Example 2 is that if we fix a_{12} (or a_{21}), then η is monotonically increasing with respect to a_{21} (or a_{12}). Furthermore, the range of η is $[1/2, 1]$, where $1/2$ is achieved when $a_{12} = a_{21} = 0$, and $\eta = 1$ when $a_{12} = 1$ or $a_{21} = 1$.

We now consider another example that clearly distinguishes the extremal dependence measures from dependence measures for the bulk of the distribution.

Example 3. *Let $Y_1, \ldots, Y_n$ be independent and identically distributed with common distribution function F_Y and $\bar{F}_Y \in \mathscr{L}_\beta, \beta > 0$. Let $X_1 = Y_1$, $X_2 = a_{21}Y_1 + Y_2 + a_{23}Y_3 + \cdots + a_{2n}Y_n$ with $0 \leq a_{21}, a_{23}, \ldots, a_{2n} < 1$. Then by Proposition 3.3, $\boldsymbol{X} = (X_1, X_2)$ is asymptotically independent with residual tail dependence coefficient $\eta = 1/(2 - a_{21})$.*

We observe that the expression of η does not depend on n, i.e., the number of independent terms in the construction of X_2. Loosely speaking, when more independent terms $a_{2i}Y_i$ with $0 < a_{2i} < 1$ are added to X_2, these added terms do not contribute much to the extremal dependence between X_1 and X_2 and the residual tail dependence coefficient remains the same. This is in clear contrast with the classical Pearson's correlation dependence measure

$$\text{Corr}(X_1, X_2) = a_{21}\Big(a_{21}^2 + 1 + \sum_{i=3}^{n} a_{2i}^2\Big)^{-1/2},$$

which decreases monotonically as n increases.

3.4 Boundary Case and Further Remarks

In Sections 3.2 and 3.3 we have shown that the discrete model $\boldsymbol{X}$ in (4) is asymptotically dependent or independent when (6) or (7) hold, respectively. The intuition for this phenomenon is that the extremal dependence of (X_1, X_2) is determined by whether the main contribution for the tails of X_1 and X_2 comes from the same components Y_i. If this is the case, then the stronger form of extremal dependence, namely asymptotic dependence, is achieved; otherwise, asymptotic independence is obtained. In the sequel, let $I_j = \arg\max_{i \in \{1,\ldots,n\}} a_{ji}$, $j = 1, 2$.

Clearly, there is a boundary case where neither (6) nor (7) hold, that is, we have $I_1 \neq I_2$ and

$$I_{\max} := I_1 \cap I_2 \neq \emptyset.$$

To avoid complications, we here only consider the case where $I_1 \not\subset I_2$ and $I_2 \not\subset I_1$. With the notation of Proposition 3.1 and the paragraph thereafter, without losing generality, we assume that $a_{1\max} = a_{2\max} = 1$, and recall that we can write $(X_1, X_2) = Z_c + (Z_1, Z_2)$, or equivalently, $(\exp(X_1), \exp(X_2)) = \exp(Z_c)(\exp(Z_1), \exp(Z_2))$. Theorem 3 in Embrechts & Goldie (1980) now implies that the survival functions of $\exp(Z_c)$, $\exp(Z_1)$, and $\exp(Z_2)$ are all regularly varying with the same index $-\beta$. Hence, Proposition 6 in Engelke et al. (2019) can be applied to determine the extremal dependence structure of (X_1, X_2). More precisely, in this case, the residual tail dependence coefficient of (X_1, X_2) is $\eta = 1$ (this can also be obtained using Proposition 3.3), but more assumptions are needed to determine whether $\chi > 0$ or $\chi = 0$. We leave the other boundary case $I_{\max} \neq \emptyset$, $I_1 \subset I_2$, or *vice versa*, for future research.

Here we have studied the extremal dependence structure of linear transformations, or sums of exponential-tailed random vectors. The results are directly applicable to products of regularly varying random vectors. Indeed, assume that $\bar{Y}_1, \ldots, \bar{Y}_n$ are independent copies of a positive random variable Y with absolutely continuous distribution function $F_{\bar{Y}}$, and $\bar{F}_{\bar{Y}} \in \mathrm{RV}_{-\beta}$ with $\beta > 0$. Let $a_{ji} \geq 0$ for $j = 1, 2, i = 1, \ldots, n$. Then the product model

$$\begin{cases} \bar{X}_1 = \bar{Y}_1^{a_{11}} \cdots \bar{Y}_n^{a_{1n}}, \\ \bar{X}_2 = \bar{Y}_1^{a_{21}} \cdots \bar{Y}_n^{a_{2n}}, \end{cases}$$

has the same extremal dependence structure as the sum model (4) with $Y_i = \log(\bar{Y}_i)$, and Propositions 3.1 and 3.3 give the extremal dependence coefficients of $(\bar{X}_1, \bar{X}_2)$.

We here restrict our attention in (4) to the case with non-negative coefficients a_{ji}, because both the integral approximation of moving average processes and the finite element approximation of the type G SPDE Matérn fields have non-negative coefficients; see Section 4 for more details. Consequently, the residual tail dependence coefficient η in Proposition 3.3 can be shown to have the lower bound $1/2$, which implies that only positive extremal association can be achieved. However, if the interest lies in capturing negative extremal association, i.e., $\eta < 1/2$, then one can achieve this by considering negative coefficients a_{ji} and assuming that both the left and right tails of the distribution of Y_i have exponential decays. We leave this for future research.

4 Moving Average Processes

4.1 Main Results

We now focus on the extremal dependence structure of moving average processes (1). Note that here $u(\boldsymbol{s})$ refers to the usual definition of stochastic integrals, where in the first step one defines integrals of simple functions, and then a non-random integrand function G is called $\mathcal{M}$-integrable if there exists a sequence of simple functions that converges pointwise to G, such that the limit in probability of the resulting integrals of simple functions exists, and this limit is defined to be the stochastic integral with respect to G (Rajput & Rosinski, 1989). A useful characterization of $\mathcal{M}$-integrable functions is given in Rajput & Rosinski (1989, Theorem 2.7). Throughout the paper we assume that $u(\boldsymbol{s})$ is well defined, i.e., G is $\mathcal{M}$-integrable.

We first consider the more commonly used case where the domain of integration $\mathcal{D}$ is fixed and does not depend on $\boldsymbol{s}$. To consider a framework that is applicable to general moving average processes which are not necessarily SPDEs, we only assume that:

B.1 the function $G(h)$ is non-negative, continuous and strictly decreasing as $h \to \infty$;

B.2 for any bounded Borel set $B \subset \mathcal{D}$, the random variable $\mathcal{M}(B)$ has an absolutely continuous distribution function $F_{\mathcal{M}(B)}$ with exponential tail $\bar{F}_{\mathcal{M}(B)} \in \mathscr{L}_\beta$, $\beta > 0$.

We note that for the important class of type G Matérn SPDE random fields on $\mathbb{R}^d$, the function

14

G is non-negative, absolutely continuous, and monotonically decreasing (see Proposition E.2 in Appendix E). Assumption B.2 implies that $\mathcal{M}(B)$ has an exponential-tailed distribution with the same index for every bounded Borel set B. This seemingly restrictive assumption is, in fact, a natural result of the convolution-closure property of exponential tails (Embrechts & Goldie, 1980, Theorem 3), and Assumption B.2 is satisfied for instance for the NIG Matérn SPDE fields and the variance Gamma Matérn SPDE fields considered in Bolin (2014), Wallin & Bolin (2015) and Bolin & Wallin (2020).

We now assume that $d = 2$, which is the most important case for spatial applications; the cases $d = 1$ and $d \geq 3$ can be treated analogously. Let $\{D_n : D_n \subset \mathcal{D}\}$ be an increasing sequence of bounded Borel sets in $\mathbb{R}^2$, and $D_n^1, D_n^2, \ldots, D_n^J \subset D_n$ be a partition of D_n obtained by triangulating D_n with m_n mesh nodes $M_n = \{\boldsymbol{a}_n^i, i = 1, \ldots, m_n\}$. Then we obtain an approximation of $u(\boldsymbol{s})$ as

$$u_n(\boldsymbol{s}) = \int_{\mathcal{D}} \sum_{i=1}^J G(\|\boldsymbol{s} - \boldsymbol{d}_i\|) 1_{D_n^i}(\boldsymbol{t}) \mathcal{M}(\mathrm{d}\boldsymbol{t}) = \sum_{i=1}^J G(\|\boldsymbol{s} - \boldsymbol{d}_i\|) \mathcal{M}(D_n^i), \tag{8}$$

where $\boldsymbol{d}_i \in D_n^i$, and $1_{D_n^i}$ is the indicator function. We say that the sequence of points M_n is dense in $\mathcal{D}$ if for any point $\boldsymbol{s} \in \mathcal{D}$, there is a sequence $\{\bar{\boldsymbol{a}}_n\}, \bar{\boldsymbol{a}}_n \in M_n$ such that $\lim_{n\to\infty} \|\bar{\boldsymbol{a}}_n - \boldsymbol{s}\| = 0$. Assume that M_n is dense. Clearly, for any fixed $\boldsymbol{s} \in \mathcal{D}$, function $G_n^J(\|\boldsymbol{s} - \boldsymbol{t}\|) = \sum_{i=1}^J G(\|\boldsymbol{s} - \boldsymbol{d}_i\|) 1_{D_n^i}(\boldsymbol{t})$ converges to the function $G(\|\boldsymbol{s} - \boldsymbol{t}\|)$ pointwise as $m_n \to \infty$. Hence, by the definition of stochastic integrals, we know that $u_n(\boldsymbol{s})$ converges to $u(\boldsymbol{s})$ in probability. It follows that $(u_n(\boldsymbol{s}_1), u_n(\boldsymbol{s}_2))$, $\boldsymbol{s}_1 \neq \boldsymbol{s}_2$, converges in probability to $(u(\boldsymbol{s}_1), u(\boldsymbol{s}_2))$.

Furthermore, by Assumption B.2, $\bar{F}_{\mathcal{M}(D_n^i)} \in \mathscr{L}_\beta$. Note also that $\mathcal{M}(D_n^i), i = 1, \ldots, J$, are independent since $\{D_n^i, i = 1, \ldots, J\}$ forms a partition of D_n. Hence, to understand the extremal dependence structure of $(u_n(\boldsymbol{s}_1), u_n(\boldsymbol{s}_2))$, it is sufficient to focus on the coefficients $G(\|\boldsymbol{s}_1 - \boldsymbol{d}_i\|)$ and $G(\|\boldsymbol{s}_2 - \boldsymbol{d}_i\|)$. Since $\boldsymbol{s}_1 \neq \boldsymbol{s}_2$, we know that if the mesh is fine enough, $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ will fall into different triangles $D_n^{i_1}$ and $D_n^{i_2}$. Consequently, we have

$$\arg\max_i G(\|\boldsymbol{s}_1 - \boldsymbol{d}_i\|) = i_1 \neq \arg\max_i G(\|\boldsymbol{s}_2 - \boldsymbol{d}_i\|) = i_2.$$

Proposition 3.3 then yields the asymptotic independence of $(u_n(\boldsymbol{s}_1), u_n(\boldsymbol{s}_2))$ and allows the computation of its residual tail dependence coefficient. We focus on the limit of the residual tail

dependence coefficient of the approximation model $(u_n(\boldsymbol{s}_1), u_n(\boldsymbol{s}_2))$ as $m_n \to \infty$.

Theorem 4.1. *Let $u_n(\boldsymbol{s})$ be the integral approximation of the moving average process (1), and let $\eta_n(h)$ be the residual tail dependence coefficient of $(u_n(\boldsymbol{s}_1), u_n(\boldsymbol{s}_2))$ for $h = \|\boldsymbol{s}_1 - \boldsymbol{s}_2\|$. If Assumptions B.1 and B.2 are satisfied and the function G is convex, then if $m_n \to \infty$ and the sequence of mesh nodes M_n is dense in $\mathcal{D}$, the limit of η_n as $n \to \infty$ is*

$$\eta(h) = \frac{1}{2} + \frac{G(h)}{2G(0)}.$$

The idea of our proof is to use Proposition 3.3 to obtain an explicit formula for $\eta_n(h)$ as a minimum (or maximum) over a finite number of terms, and with this, to show that for convex G, the optimum can be achieved as the mesh size tends to zero. We remark that when $G(0) = \infty$, $\eta(h)$ reduces to a constant $1/2$. When the function G is non-convex, we conjecture that the limiting residual tail dependence function is

$$\eta(h) = \max\left\{\frac{1}{2} + \frac{G(h)}{2G(0)}, \frac{G(h/2)}{G(0)}\right\}.$$

Our simulation studies in Section 4.3 seem to support our conjecture, but a rigorous proof would have to use a different technique than the proof of Theorem 4.1, which relies heavily on the convexity of G.

We now consider another important case where the integration domain $\mathcal{D}$ depends on s, namely when $d = 1$ and

$$u(s) = \int_{-\infty}^{s} G(s - t)\mathcal{M}(\mathrm{d}t), \quad s \in (-\infty, T]. \tag{9}$$

This is an interesting case because it is used in practice (Barndorff-Nielsen & Shephard, 2001; Ver Hoef & Peterson, 2010), and this one-sided integral turns out to yield different residual tail dependence functions.

Let $-n < s_1 < s_2 \leq T$ and $-n = t_0 < \cdots < t_{n_1} \leq s_1 < \cdots \leq t_{n_2} \leq s_2 < \cdots < t_{m_n} = T$ be an arbitrary partition of $[-n, T]$. Then the approximation (8) becomes

$$u_n(s_1) = \sum_{i=0}^{n_1-1} G(s_1 - t_i)\mathcal{M}([t_i, t_{i+1})), \quad u_n(s_2) = \sum_{i=0}^{n_2-1} G(s_2 - t_i)\mathcal{M}([t_i, t_{i+1})). \tag{10}$$

Clearly when $m_n \to \infty$ such that the sequence of partition points $M_n = \{t_i, i = 0, \ldots, m_n\}$ is dense in $(-\infty, T]$, $(u_n(s_1), u_n(s_2))$ converges in probability to $(u(s_1), u(s_2))$. Note that when the mesh is coarse and there is no partition point in (s_1, s_2), i.e., $t_{n_1} = t_{n_2}$, then the largest coefficients in the expressions of $u_n(s_1)$ and $u_n(s_2)$ are $G(s_1 - t_{n_1-1})$ and $G(s_2 - t_{n_1-1})$ respectively, which correspond to the same variable $\mathcal{M}([t_{n_1-1}, t_{n_1}))$. Hence, Proposition 3.1 gives the asymptotic dependence of $(u_n(t_1), u_n(t_2))$. Otherwise, when there is at least one partition point in (s_1, s_2), $(u_n(t_1), u_n(t_2))$ is asymptotically independent, and in the following, we derive its limiting residual tail dependence function as the mesh size tends to zero.

Theorem 4.2. *Let $u_n(s)$ be the integral approximation* (10) *of the one-side integral* (9), *and let $\eta_n(h)$ be the residual tail dependence coefficient of $(u_n(s_1), u_n(s_2))$ for $h = |s_1 - s_2|$. If Assumptions B.1 and B.2 are satisfied, then if $m_n \to \infty$ and the sequence of mesh nodes M_n is dense in $\mathcal{D}$, the limit of η_n as $n \to \infty$ is*

$$\eta(h) = \frac{1}{2 - G(h)/G(0)}.$$

4.2 Non-Gaussian Ornstein-Uhlenbeck (OU) Process

As the first application of the general results of the previous section, here we study the extremal dependence structure of non-Gaussian OU processes $u(t)$ (Barndorff-Nielsen & Shephard, 2001), defined as the stationary solution to the stochastic differential equation (SDE)

$$\mathrm{d}u(t) = -au(t)\mathrm{d}t + \mathrm{d}z(at), \quad a > 0, t \in \mathbb{R}, \tag{11}$$

where $z = \{z(t) : t \in \mathbb{R}\}$ is a Lévy process satisfying $\mathrm{E}[\log\{1 + |z(1)|\}] < \infty$ to guarantee the existence of such a stationary solution. Although the background driving Lévy process z can be chosen arbitrarily, examples considered in Barndorff-Nielsen & Shephard (2001) all have exponential tails for the density of the Lévy measure of $z(1)$.

Non-Gaussian OU processes are generalizations of classical OU processes by replacing the Brownian motion $z(t)$ in the SDE (11) by general Lévy processes. The existence of such processes is established based on the notion of self-decomposability and the stochastic integral representation of self-decomposable random variables (Jurek & Vervaat, 1983). More precisely, let V be a self-decomposable random variable (namely for every $\alpha \in (0, 1)$, we have the decomposition

$V = \alpha V + \tilde{V}_\alpha$, where $\tilde{V}_\alpha$ is a random variable independent of V), then there exists a Lévy process $z(t)$ and a stationary stochastic process $u(s)$ such that (11) holds for all $a > 0$ and $u(s)$ has the same distribution as V for all $s \geq 0$. Conversely, if z is a Lévy process and $u(t)$ is a stationary stochastic process such that $u(s)$ satisfying (11) for all $a > 0$, then the marginal distribution of $u(s)$ is self-decomposable. We refer to Jurek & Vervaat (1983) and Barndorff-Nielsen & Shephard (2001) for more details.

Importantly, the solution to the SDE (11) can be represented as

$$u(s) = \int_{-\infty}^{s} e^{-a(s-t)}\mathrm{d}z(at) = e^{-as}u(0) + \int_{0}^{s} e^{-a(s-t)}\mathrm{d}z(at), \quad s \geq 0, \tag{12}$$

where $u(0)$ is independent of $\int_{0}^{s} e^{-a(s-t)}\mathrm{d}z(at)$. Using this representation, we can show that if the stationary solution $u(s)$ has an absolutely continuous and exponential-tailed distribution F with $\bar{F} \in \mathscr{L}_\beta, \beta > 0$, then the process $u(s)$ is asymptotically independent.

Theorem 4.3. *Let $u(s)$ be a non-Gaussian OU process defined as a stationary solution to (11) and V has the stationary self-decomposable distribution. If the distribution function of V is absolutely continuous with exponential tail $\bar{F}_V \in \mathscr{L}_\beta$, $\beta > 0$, then $(u(s_1), u(s_2))$ is asymptotically independent for $s_1 \neq s_2$. The corresponding residual tail dependence coefficient is*

$$\eta = \frac{1}{2 - e^{-a|s_2-s_1|}}.$$

Note that when the Lévy process z in (11) is a Brownian motion, we obtain the classical Gaussian OU process with the correlation between $u(s_1)$ and $u(s_2)$ as $e^{-a|s_2-s_1|}$ and residual tail dependence coefficient as $(1 + e^{-a|s_2-s_1|})/2$. Theorem 4.3 shows that when the marginal distribution of the non-Gaussian OU process $u(s)$ has an exponential tail, its induced extremal dependence structure is indeed different from the classical Gaussian OU process, although its correlation function remains the same. More precisely, using a first-order Taylor expansion, one can observe that as $|s_2 - s_1|$ tends to zero, the residual tail dependence coefficient η of the Gaussian OU process increases to 1 at a linear rate $a/2$, whilst η of the non-Gaussian OU process increases to 1 at a linear rate a.

Instead of specifying the self-decomposable distribution function of $u(s)$, one can alternatively define the non-Gaussian OU process by specifying the background driving Lévy process z, which

reduces to specifying the distribution of $z(1)$. Although the relationship between the tail of $z(1)$ and that of $u(t)$ is unclear, $u(t)$ and $z(1)$ are closely linked by their Lévy measures U_u and $U_{z(1)}$ through the equation $U_u([x, \infty)) = \int_1^\infty s^{-1} U_{z(1)}([sx, \infty)) \mathrm{d}s$ (Barndorff-Nielsen, 1998, Theorem 2.2). Using this link, we show in the following that under mild assumptions, a convolution equivalent tail of $z(1)$ implies a convolution equivalent tail for $u(t)$.

Proposition 4.4. *Denote the distribution function of $z(1)$ and $u(t)$ by $F_{z(1)}$ and F_u, respectively. Suppose that the function $U_{z(1)}([x, \infty))$ is continuous in x on $(0, \infty)$. If $\bar{F}_{z(1)} \in \mathscr{S}_\beta$, $\beta \geq 0$, then $\bar{F}_u \in \mathscr{S}_\beta$. Moreover, $\bar{F}_{z(t)} \in \mathscr{S}_\beta$, where $F_{z(t)}$ denotes the distribution function of $z(t)$. Hence, all increments of the Lévy process $z(t)$ have exponential-tailed distribution functions with the same index β.*

One example is given by $U_{z(1)}([x, \infty)) = cx^{-\epsilon} \exp(-\beta x)$, where $\epsilon > 1$ and $c, \beta > 0$; see Example 2.4.1 in Barndorff-Nielsen & Shephard (2001). The restriction $\epsilon > 1$ implies that the normalized Lévy measure of $z(1)$ is convolution tail equivalent and hence $\bar{F}_{z(1)} \in \mathscr{S}_\beta$ (Shimura & Watanabe, 2005, Theorem B).

Now we assume that $z(1)$ has a convolution tail equivalent distribution $F_{z(1)}$ with $\bar{F}_{z(1)} \in \mathscr{S}_\beta$, $\beta > 0$, the function $U_{z(1)}([x, \infty))$ is continuous in x on $(0, \infty)$, and the distribution function of $z(t)$ is absolutely continuous for all t. That is, Assumption B.2 is satisfied. Then we are ready to link our results in Theorem 4.2 and Theorem 4.3.

Clearly, for the OU processes, the integrand function in (12) satisfies Assumption B.1. Hence, Theorem 4.2 implies that the limiting residual tail dependence function of the approximation model of the form (10) is $\eta(h) = 1/(2 - e^{-ah})$. On the other hand, the convolution equivalent tail of $z(1)$ implies a convolution equivalent tail for $u(s)$. If we further assume that the stationary distribution of $u(s)$ is absolutely continuous, then Theorem 4.3 gives its residual tail dependence function as $\eta(h) = 1/(2 - e^{-ah})$, which coincides with the limit of its approximation model. In Figure 3, we illustrate this convergence of the residual tail dependence coefficient η of the approximating model to the true non-Gaussian OU process for $a = 0.2$, $s_1 = 0$, $T = 4$, and three equidistant partitions of the interval $[0, T]$ with mesh length $\Delta = 0.4, 0.2, 0.05$.

The preceding analysis seemingly indicates that the extremal dependence structure is preserved in the limit for convergent (in probability) random vectors. It is however important to note that in
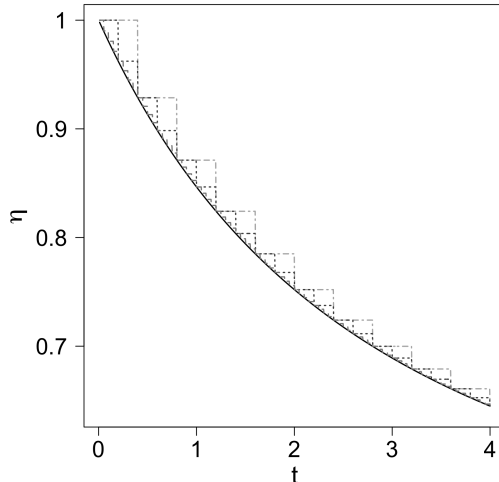
Figure 3: The residual tail dependence coefficient η of the true non-Gaussian OU process (black line), and its integral approximations for three different, equidistant partitions of $[0,T]$ with mesh length $\Delta \in \{0.4, 0.2, 0.05\}$, where the lines are ordered from highest to lowest for decreasing Δ.

general, convergence in probability or even almost surely does not (!) necessarily imply convergence of the corresponding tail dependence coefficients, as shown in the following counterexample.

Example 4. *Let* $X_{1,n} = X/n + \epsilon_1, X_{2,n} = X/n + \epsilon_2$, *where* $X, \epsilon_1, \epsilon_2$ *are independent,* X *has a regularly varying survival function* $\bar{F}_X \in RV_\rho$, *and* $\epsilon_i, i = 1, 2$ *are standard normal random variables. It is straightforward to see that* $(X_{1,n}, X_{2,n})$ *converges almost surely to the limiting random vector* (ϵ_1, ϵ_2). *However, the extremal dependence is clearly not preserved in the limit, since by Proposition 4 in* Engelke et al. (2019) *we know that for any finite n,* $(X_{1,n}, X_{2,n})$ *is asymptotically dependent with tail dependence coefficient* $\chi = 1$, *whilst* (ϵ_1, ϵ_2) *is asymptotically independent with residual tail dependence coefficient* $\eta = 1/2$.

4.3 Type G Matérn SPDE Random Fields

The second application of our general results is the popular class of type G Matérn SPDE random fields defined as the stationary solution to the SPDE

$$(\kappa^2 - \Delta)^{\alpha/2} u(\boldsymbol{s}) = \dot{\mathcal{M}}(\boldsymbol{s}), \quad \boldsymbol{s} \in \mathbb{R}^d, \tag{13}$$

where $\alpha = \nu + d/2$, $d = 1, 2, \ldots$ is the dimension, $\nu > 0$ is the smoothness parameter, $\kappa > 0$ is the range parameter, Δ is the Laplacian and $\dot{\mathcal{M}}$ is the so-called type G Lévy noise (Rosinski, 1991).

The solution $u(\boldsymbol{s})$ can be expressed as process convolutions

$$u(\boldsymbol{s}) = \int_{\mathbb{R}^d} G(\|\boldsymbol{s} - \boldsymbol{t}\|)\mathcal{M}(\mathrm{d}\boldsymbol{t}), \quad \boldsymbol{s} \in \mathbb{R}^d, \tag{14}$$

where $\mathcal{M}$ is the random measure associated with the noise $\dot{\mathcal{M}}$, and G is the Green's function of the differential operator in (13) of the form

$$G(\|\boldsymbol{s} - \boldsymbol{t}\|) = \frac{2^{1-\frac{\alpha-d}{2}}}{(4\pi)^{d/2}\Gamma(\alpha/2)\kappa^{\alpha-d}}(\kappa\|\boldsymbol{s} - \boldsymbol{t}\|)^{\frac{\alpha-d}{2}} K_{\frac{\alpha-d}{2}}(\kappa\|\boldsymbol{s} - \boldsymbol{t}\|), \tag{15}$$

with Γ being the gamma function and K being the modified Bessel function of the second kind.

Notably, for the important class of type G Matérn SPDE random fields, the function G is absolutely continuous and monotonically decreasing (see Proposition E.2 in Appendix E). This implies that Assumption B.1 is satisfied. Moreover, Assumption B.2 is satisfied for the NIG Matérn SPDE fields and variance Gamma Matérn SPDE fields considered in Bolin (2014), Wallin & Bolin (2015), and Bolin & Wallin (2020). Hence, if we consider its integral approximation of the form (8), Theorem 4.1 gives the limiting residual tail dependence function for convex G.

We remark that for $d = 2$, the specific Green's function G in (15) is convex only when $\alpha \leq 3$ and $G(0)$ is bounded when $\alpha > 2$. This implies that when $\alpha = 2$, which is the case commonly used in practice, if the constructed mesh is very fine, then the resulting discrete approximation model would have $\eta \approx 1/2$ (near independence) between all pairs of locations, regardless of the distance between them. From a practical point of view, this indicates that the case $\alpha = 2$ might not be suitable for modeling extremal dependence, whilst $\alpha = 3$ might be more useful. We also remark that the frequently used finite element approximation is in the form of linear transformations as well, and more details are given in Appendix C.

We conduct simulations to illustrate our results. We consider the NIG Matérn SPDE model with range parameter $\kappa = 2$, smoothness parameter $\alpha = 2$, and NIG noise location parameter $\mu = -1$, skewness parameter $\gamma = 1$, and shape parameters $\psi = \tau = 1$. We first consider its finite element approximation and examine the convergence of the pre-asymptotic tail dependence coefficient $\chi(q)$ as $q \to 1$. We randomly select 100 sites in the unit square and consider a fine mesh constructed based on a 1600-node lattice with outer extensions; see the left panel of Figure 4.
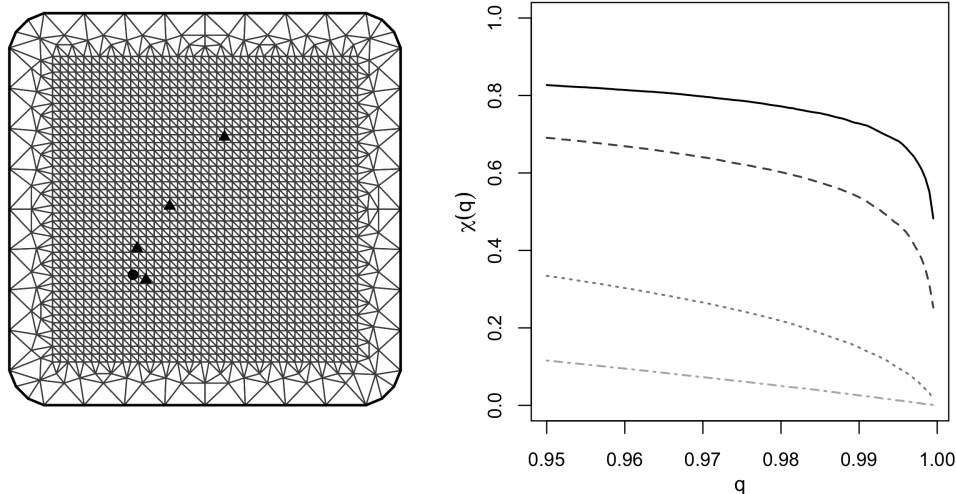
Figure 4: Left panel: constructed mesh and selected pairs of sites with different distances (triangle points to the round point) in the first simulation study in Section 4.3; right panel: empirical tail dependence coefficient $\chi(q)$ for different probability levels q, of the finite element approximation of the NIG Matérn SPDE model. Four pairs of locations with different distances are plotted, with the lines ordered from top to bottom for increasing distances.

Then we simulate $5 * 10^7$ observations at each site, and compute the empirical tail dependence coefficient $\chi(q)$ with respect to different probability levels u. Figure 4 (right panel) shows $\chi(q)$ for four different pairs of locations with different distances. The plot indicates that the two pairs with longer distances are asymptotically independent and also depicts the decay rate of $\chi(q)$ as $q \to 1$. The other two pairs with shorter distances are also asymptotically independent by Proposition 3.3, but their corresponding $\chi(q)$ decays at a much slower rate and much more simulations are needed to show that its limit is zero.

We now compare the integral approximation with the finite element approximation and examine the effect of the smoothness parameter α on the extremal dependence structure of the approximation models. We choose the same SPDE range parameter, NIG noise parameters and the fine mesh constructed based on 1600 lattice nodes as in the first simulation, and consider more sites, namely 225 randomly selected sites in the unit square, and smoothness parameter $\alpha = 2, \ldots, 5$. We numerically compute the residual tail dependence coefficient η of all pairs of sites for both approximations using the formula from Proposition 3.3.

Figure 5 depicts η against the distance between all pairs of sites. A first observation is that with such a fine mesh the difference between the integral approximation and the finite element
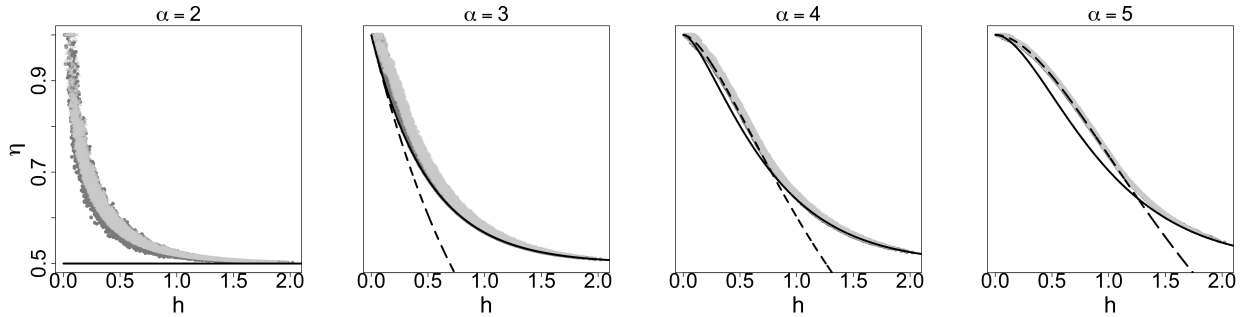
Figure 5: Coefficients η for the finite element approximation (light-grey triangle points) and integral approximation (dark-grey round points) of the type G Matérn SPDE model based on a fine mesh, and the functions $1/2 + G(h)/\{2G(0)\}$ (black solid line) and $G(h/2)/G(0)$ (black dashed line).

approximation is negligible. When $\alpha = 2$, i.e., the function G in (15) is convex but unbounded at 0, the residual tail dependence function $\eta(h)$ is still fairly far from its limiting function (as the mesh size tends to 0), which is $\eta(h) \equiv 1/2$. On the other hand, when $\alpha = 3$, i.e., G is convex and bounded at 0, then the function $\eta(h)$ of both approximation models has almost converged to its limiting function. While the cases $\alpha = 4, 5$ are not covered by our theory since G is nonconvex, the η values of the approximation models seem to converge to our conjectured limiting function, namely $\eta(h) = \max\{1/2 + G(h)/\{2G(0)\}, G(h/2)/G(0)\}$. This provides numerical evidence for our conjecture.

5 Conclusion

Linear transformations of a random vector with independent components are classical constructions to capture complex correlation dependence structures due to their simplicity and analytical tractability. For instance, the approximation models used in practice in the well-known SPDE approach (Lindgren et al., 2011; Bolin, 2014) are of this form. In this paper, we derived the first results on the extremal dependence structure induced by such constructions when the independent components have exponential tails. These general results are leveraged to study the extremal dependence structure of moving average processes driven by exponential-tailed Lévy noise. In particular, the classical exponential-tailed non-Gaussian OU processes are shown to be asymptotically independent, but with a different residual tail dependence function than their Gaussian counterpart. As for the type G Matérn random fields, or more general moving average processes, we have

shown that under certain assumptions on the kernel function and the noise process, the integral approximation is asymptotically independent when the mesh is fine enough and the limiting residual tail dependence function is derived.

In terms of statistical modeling, linear transformations of exponential-tailed random vectors have the potential to bridge asymptotic dependence and independence in a tractable way, and models with such features are in pursuit in the extremes community (Nolde & Zhou, 2021). In fact, this desirable property distinguishes them from other marginal distributions. For instance, linear combinations of heavy-tailed random variables can only result in asymptotic dependence or complete independence. On the other hand, linear transformations of Gaussian random vectors exclusively yield asymptotic independence or complete dependence. Therefore an interesting and natural question is whether exponential tails are the only ones that can exhibit both extremal dependence classes under linear transformations.

Acknowledgement

Appendix A   Lemmas and Proofs for Section 3

Before proving Proposition 3.1, we first present three useful lemmas as follows. The first one concerns the asymptotic expansion of the quantile of an exponential-tailed distribution function, the second states how much the quantile changes in the limit when an exponential-tailed random variable convolves an independent lighter-tailed random variable, and the third one can be thought of as a more generalized definition of exponential-tailed functions.

Lemma 1. *Let Y be a random variable with distribution function F_Y such that $\bar{F}_Y \in \mathscr{L}_\beta, \beta > 0$, then we have the asymptotic expansion of its quantile function*

$$F_Y^{-1}(u) \sim -\frac{\log(1-u)}{\beta}, \quad u \uparrow 1.$$

Proof. Using the representation (5), we have

$$\bar{F}_Y(x) = a(x)\exp\left\{-\int_0^x \beta(v)\mathrm{d}v\right\},$$

where $a(x) \to a \in (0, \infty)$ and $\beta(x) \to \beta$ as $x \to \infty$. Now consider the function $y(u) = -\log(1-u)/\beta$, $0 < u < 1$. We have

$$y(F_Y(x)) = -\frac{\log\{\bar{F}_Y(x)\}}{\beta} = \frac{-\log\{a(x)\} + \int_0^x \beta(v)\mathrm{d}v}{\beta} \sim x, \quad x \to \infty.$$

Furthermore,

$$y(F_Y(F_Y^{-1}(u))) = -\frac{\log\{\bar{F}_Y(F_Y^{-1}(u))\}}{\beta} \sim -\frac{\log(1-u)}{\beta}, \quad u \uparrow 1,$$

where the last step is due to the fact that $\bar{F}(F^{-1}(u)) \sim 1 - u$ and that the logarithm function is slowly varying and thus preserves asymptotic equivalence (see Proposition 2.6 (iii) of Resnick (2007) or Section 3.4 of Buldygin et al. (2018) for a more detailed treatment). Therefore, combining these two above results gives $F_Y^{-1}(u) \sim -\frac{\log(1-u)}{\beta}$ as $u \uparrow 1$, and the proof is complete. $\qquad\square$

Lemma 2. *Let Y be a random variable with distribution function F_Y such that $\bar{F}_Y \in \mathscr{L}_\beta, \beta > 0$. Let Z_1, Z_2 be random variables independent of Y and have moment generating functions M_{Z_i} such that $M_{Z_i}(\gamma) < \infty$ for some $\gamma > \beta$. Then for $X_1 = Y + Z_1$, $X_2 = Y + Z_2$, we have*

$$c_1 := \lim_{u\uparrow 1}\{F_{X_1}^{-1}(u) - F_Y^{-1}(u)\} = \frac{1}{\beta}\log\{M_{Z_1}(\beta)\},$$

$$c_2 := \lim_{u\uparrow 1}\{F_{X_2}^{-1}(u) - F_Y^{-1}(u)\} = \frac{1}{\beta}\log\{M_{Z_2}(\beta)\},$$

$$c_3 := \lim_{u\uparrow 1}\{F_{X_2}^{-1}(u) - F_{X_1}^{-1}(u)\} = c_2 - c_1,$$

where F_{X_1}, F_{X_2} are the distribution functions of X_1 and X_2, respectively.

Proof. By Lemma 1 in Cline (1986) or Proposition 3 in Breiman (1965), we have that $\bar{F}_{X_1}, \bar{F}_{X_2} \in \mathscr{L}_\beta$ and

$$\lim_{x\to\infty}\frac{\bar{F}_{X_1}(x)}{\bar{F}_Y(x)} = M_{Z_1}(\beta), \quad \lim_{x\to\infty}\frac{\bar{F}_{X_2}(x)}{\bar{F}_Y(x)} = M_{Z_1}(\beta).$$

Using the representation (5), we have

$$\bar{F}_{X_1}(x) \sim M_{Z_1}(\beta)\bar{F}_Y(x) \sim M_{Z_1}(\beta)a(x)\exp\left\{-\int_0^x \beta(v)\mathrm{d}v\right\}, \quad x \to \infty.$$

Since $\bar{F}_Y(F_Y^{-1}(u)) \sim 1 - u \sim \bar{F}_{X_1}(F_{X_1}^{-1}(u))$ as u tends to 1, we have that as $u \uparrow 1$,

$$a(F_Y^{-1}(u))\exp\left\{-\int_0^{F_Y^{-1}(u)} \beta(v)\mathrm{d}v\right\} \sim M_{Z_1}(\beta)a(F_{X_1}^{-1}(u))\exp\left\{-\int_0^{F_{X_1}^{-1}(u)} \beta(v)\mathrm{d}v\right\}.$$

As $\bar{F}_Y, \bar{F}_{X_1} \in \mathscr{L}_\beta$, using Lemma 1 we know that

$$F_Y^{-1}(u) \sim F_{X_1}^{-1}(u) \sim -\frac{\log(1-u)}{\beta}, \quad u \uparrow 1.$$

The fact $a(x) \to a \in (0, \infty)$ implies that the function $a(\cdot)$ is slowly varying and thus it preserves asymptotic equivalence. That is, $a(F_Y^{-1}(u)) \sim a(F_{X_1}^{-1}(u))$, as $u \uparrow 1$. Hence,

$$M_{Z_1}(\beta)\exp\left\{-\int_{F_Y^{-1}(u)}^{F_{X_1}^{-1}(u)} \beta(v)\mathrm{d}v\right\} \to 1, \quad u \uparrow 1.$$

By dominated convergence, we have

$$M_{Z_1}(\beta)\exp[-\beta \lim_{u\uparrow 1}\{F_{X_1}^{-1}(u) - F_Y^{-1}(u)\}] = 1.$$

Therefore, rearranging the terms gives

$$c_1 = \lim_{u\uparrow 1}\{F_{X_1}^{-1}(u) - F_Y^{-1}(u)\} = \frac{1}{\beta}\log\{M_{Z_1}(\beta)\}.$$

Similarly, we get $c_2 = \log\{M_{Z_2}(\beta)\}/\beta$, and

$$c_3 = \lim_{u\uparrow 1}\{F_{X_2}^{-1}(u) - F_{X_1}^{-1}(u)\} = \lim_{u\uparrow 1}[\{F_{X_2}^{-1}(u) - F_Y^{-1}(u)\} - \{F_{X_1}^{-1}(u) - F_Y^{-1}(u)\}] = c_2 - c_1.$$

$\square$

Lemma 3. *Let F be a distribution function such that $\bar{F} \in \mathscr{L}_\beta, \beta \geq 0$. Suppose g_1, g_2 are two real-valued functions on $\mathbb{R}$ that satisfy $g_1(u) \to \infty$, $g_2(u) \to \infty$, $g_1(u)/g_2(u) \to 1$, and $g_1(u) - g_2(u) \to g$*

with $g \in (-\infty, \infty)$ as $u \to u_0$, then

$$\lim_{u \to u_0} \frac{\bar{F}(g_1(u))}{\bar{F}(g_2(u))} = \exp(-g\beta).$$

Proof. Since $\bar{F} \in \mathscr{L}_\beta$, we know that $\bar{F}(t)$ has the representation (5). Using the same argument as in Lemma 2, we know that $a(\cdot)$ preserves asymptotic equivalence. Thus, $a(g_1(u)) \sim a(g_2(u)), u \to u_0$ and we further have

$$\begin{aligned}
\lim_{u \to u_0} \frac{\bar{F}(g_1(u))}{\bar{F}(g_2(u))} &= \lim_{u \to u_0} \frac{a(g_1(u))}{a(g_2(u))} \exp\left\{ -\int_{g_2(u)}^{g_1(u)} \beta(v)\mathrm{d}v \right\} \\
&= \lim_{u \to u_0} \exp\left\{ -\int_{g_2(u)}^{g_1(u)} \beta(v)\mathrm{d}v \right\} \\
&= \exp(-g\beta),
\end{aligned}$$

where the last equality holds by dominated convergence theorem. This completes the proof. $\square$

Now we are ready to prove Proposition 3.1.

Proof of Proposition 3.1. Note that extremal dependence is a copula property, i.e., it is invariant to strictly increasing marginal transformations. This implies that, $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \tilde{X}_2)^\top$ with $\tilde{X}_1 = X_1/a_{1\max}$, $\tilde{X}_2 = X_2/a_{2\max}$, will have the same extremal dependence structure as $\boldsymbol{X}$. Hence, without losing generality we assume that $a_{1\max} = a_{2\max} = 1$ and $0 \le a_{ji} \le 1, j = 1, 2, i = 1, \ldots, n$. We thus have $Z_1 = \sum_{i \notin I_{\max}} a_{1i}Y_i$, $Z_2 = \sum_{i \notin I_{\max}} a_{2i}Y_i$. Denote $Z_c = \sum_{i \in I_{\max}} Y_i$, then $X_1 = Z_c + Z_1$ and $X_2 = Z_c + Z_2$.

In order to derive the tail dependence coefficient χ, we need to compare the following probability with $1 - u$ as $u \uparrow 1$

$$\begin{aligned}
I := &\Pr\left(X_1 > F_{X_1}^{-1}(u), X_2 > F_{X_2}^{-1}(u)\right) \\
= &\Pr\left(Z_c + Z_1 > F_{X_1}^{-1}(u), Z_c + Z_2 > F_{X_2}^{-1}(u)\right) \\
= &\Pr\left(Z_c + Z_1 > F_{X_1}^{-1}(u), Z_c + Z_2 > F_{X_2}^{-1}(u), Z_2 - Z_1 \le F_{X_2}^{-1}(u) - F_{X_1}^{-1}(u)\right) + \\
&\Pr\left(Z_c + Z_1 > F_{X_1}^{-1}(u), Z_c + Z_2 > F_{X_2}^{-1}(u), Z_2 - Z_1 > F_{X_2}^{-1}(u) - F_{X_1}^{-1}(u)\right).
\end{aligned}$$

For $Z_j = \sum_{i \notin I_{\max}} a_{ji}Y_i, j = 1, 2$, if $a_{ji} = 0$ for all $i \notin I_{\max}$, then $Z_j = 0$ and clearly its moment

generating function satisfies $M_{Z_j}(t) = 1 < \infty$. If there is at least one $a_{ji} > 0$ for some $i \notin I_{\max}$, then by Theorem 3 in Embrechts & Goldie (1980), we know that Z_c has exponential-tailed distribution function with index β, and Z_j has the distribution function F_{Z_j} with $\bar{F}_{Z_j} \in \mathscr{L}_{\beta/\max_{i \notin I_{\max}} a_{ji}}$. Hence, Z_j has a lighter exponential tail than Z_c and there exists $\gamma > \beta$ such that $M_{Z_i}(\gamma) < \infty$. Therefore, the conditions in Lemma 2 are satisfied.

Since $Z_c + Z_2 > F_{X_2}^{-1}(u)$ and $Z_2 - Z_1 \leq F_{X_2}^{-1}(u) - F_{X_1}^{-1}(u)$ imply $Z_c + Z_1 > F_{X_1}^{-1}(u)$, we have that

$$\{Z_c + Z_2 > F_{X_2}^{-1}(u), Z_2 - Z_1 \leq F_{X_2}^{-1}(u) - F_{X_1}^{-1}(u)\} \subseteq \{Z_c + Z_1 > F_{X_1}^{-1}(u)\}.$$

Hence,

$$
\begin{aligned}
&\lim_{u \uparrow 1} \frac{\Pr\left(Z_c + Z_1 > F_{X_1}^{-1}(u), Z_c + Z_2 > F_{X_2}^{-1}(u), Z_2 - Z_1 \leq F_{X_2}^{-1}(u) - F_{X_1}^{-1}(u)\right)}{\Pr\left(Z_c > F_{Z_c}^{-1}(u)\right)} \\
&= \lim_{u \uparrow 1} \frac{\Pr\left(Z_c + Z_2 > F_{X_2}^{-1}(u), Z_2 - Z_1 \leq F_{X_2}^{-1}(u) - F_{X_1}^{-1}(u)\right)}{\Pr\left(Z_c > F_{Z_c}^{-1}(u)\right)} \\
&= \lim_{u \uparrow 1} \int \int_{z_2 - z_1 \leq F_{X_2}^{-1}(u) - F_{X_1}^{-1}(u)} \frac{\Pr\left(Z_c > F_{X_2}^{-1}(u) - z_2\right)}{\Pr\left(Z_c > F_{Z_c}^{-1}(u)\right)} F_{\mathbf{Z}}(\mathrm{d}z_1, \mathrm{d}z_2) \\
&= \int \int_{z_2 - z_1 \leq c_3} \exp(\beta z_2 - c_2) F_{\mathbf{Z}}(\mathrm{d}z_1, \mathrm{d}z_2),
\end{aligned}
$$

where the last equality holds due to Lemma 2, Lemma 3, and dominated convergence theorem. Similarly, we have

$$
\begin{aligned}
&\lim_{u \uparrow 1} \frac{\Pr\left(Z_c + Z_1 > F_{X_1}^{-1}(u), Z_c + Z_2 > F_{X_2}^{-1}(u), Z_2 - Z_1 > F_{X_2}^{-1}(u) - F_{X_1}^{-1}(u)\right)}{\Pr\left(Z_c > F_{Z_c}^{-1}(u)\right)} \\
&= \int \int_{z_2 - z_1 > c_3} \exp(\beta z_1 - c_1) F_{\mathbf{Z}}(\mathrm{d}z_1, \mathrm{d}z_2).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\chi &= \lim_{u \uparrow 1} \frac{I}{1 - u} \\
&= \lim_{u \uparrow 1} \frac{\Pr\left(Z_c + Z_1 > F_{X_1}^{-1}(u), Z_c + Z_2 > F_{X_2}^{-1}(u), Z_2 - Z_1 \le F_{X_2}^{-1}(u) - F_{X_1}^{-1}(u)\right)}{\Pr\left(Z_c > F_{Z_c}^{-1}(u)\right)} + \\
&\qquad \lim_{u \uparrow 1} \frac{\Pr\left(Z_c + Z_1 > F_{X_1}^{-1}(u), Z_c + Z_2 > F_{X_2}^{-1}(u), Z_2 - Z_1 > F_{X_2}^{-1}(u) - F_{X_1}^{-1}(u)\right)}{\Pr\left(Z_c > F_{Z_c}^{-1}(u)\right)} \\
&= \int\int_{z_2 - z_1 \le c_3} \exp(\beta z_2 - c_2) F_{\mathbf{Z}}(\mathrm{d}, z_1 \mathrm{d}z_2) + \int\int_{z_2 - z_1 > c_3} \exp(\beta z_1 - c_1) F_{\mathbf{Z}}(\mathrm{d}z_1, \mathrm{d}z_2) \\
&= \frac{\int\int_{z_2 - z_1 > c_3} \exp(\beta z_1) F_{\mathbf{Z}}(\mathrm{d}z_1, \mathrm{d}z_2)}{M_{Z_1}(\beta)} + \frac{\int\int_{z_2 - z_1 \le c_3} \exp(\beta z_2) F_{\mathbf{Z}}(\mathrm{d}z_1, \mathrm{d}z_2)}{M_{Z_2}(\beta)} \\
&= \mathrm{E}[\min\{\exp(\beta Z_1)/M_{Z_1}(\beta), \exp(\beta Z_2)/M_{Z_2}(\beta)\}].
\end{aligned}
$$

$\square$

We now prove Proposition 3.2.

Proof of Proposition 3.2. For $Y_1 \sim \mathrm{GH}(\lambda, \tau, \psi, \mu = 0, \gamma = 0)$, from Section 2 we know that its density function $f_{Y_1}(y)$ has asymptotic expansion

$$
f_{Y_1}(y) \sim c_0 y^{\lambda - 1} e^{-\sqrt{\psi} y}, \quad y \to \infty,
$$

where $c_0 > 0$ is a constant. It can be easily seen that its moment generating function $M_{Y_1}(t)$ evaluated at point $t = \sqrt{\psi}$ is finite if $\lambda < 0$, but is infinite if $\lambda \ge 0$. Hence, if $\lambda < 0$, we have

$$
\lim_{a_{22} \uparrow 1} \chi = \frac{\int_{-\infty}^{c^*} \exp(\sqrt{\psi} y) F_{Y_1}(\mathrm{d}y)}{M_{Y_1}(\sqrt{\psi})} + \frac{\int_{c^*}^{\infty} \exp(a_{12}\sqrt{\psi} y) F_{Y_1}(\mathrm{d}y)}{M_{Y_1}(a_{12}\sqrt{\psi})}.
$$

If $\lambda \ge 0$, we know that $\lim_{a_{22} \uparrow 1} M_{Y_1}(a_{22}\sqrt{\psi}) = \infty$ and thus $\lim_{a_{22} \uparrow 1} c = \infty$. Hence, the second term in the expression of χ in Example 1 vanishes. We now show the first term tends to zero as $a_{22} \uparrow 1$. Since $f_{Y_1}(y) \sim c_0 y^{\lambda - 1} e^{-\sqrt{\psi} y}, y \to \infty$, we have for any $\varepsilon > 0$, there exists $N > 0$ such that

$f_{Y_1}(y) < (1+\varepsilon)c_0 y^{\lambda-1} e^{-\sqrt{\psi}y}$ for every $y > N$. Hence, for $\lambda > 0$

$$
\frac{\int_{-\infty}^{c/(a_{22}-a_{12})} \exp(a_{22}\sqrt{\psi}y) F_{Y_1}(\mathrm{d}y)}{M_{Y_1}(a_{22}\sqrt{\psi})}
$$

$$
= \frac{\int_{-\infty}^{c/(a_{22}-a_{12})} \exp(a_{22}\sqrt{\psi}y) f_{Y_1}(y)\mathrm{d}y}{M_{Y_1}(a_{22}\sqrt{\psi})}
$$

$$
< \frac{\int_{-\infty}^{N} \exp(a_{22}\sqrt{\psi}y) f_{Y_1}(y)\mathrm{d}y}{M_{Y_1}(a_{22}\sqrt{\psi})} + \frac{\int_{N}^{c/(a_{22}-a_{12})} (1+\varepsilon)cy^{\lambda-1}\exp\{-(1-a_{22})\sqrt{\psi}y\}\mathrm{d}y}{M_{Y_1}(a_{22}\sqrt{\psi})}
$$

$$
< \frac{\exp(a_{22}\sqrt{\psi}N)\int_{-\infty}^{N} f_{Y_1}(y)\mathrm{d}y}{M_{Y_1}(a_{22}\sqrt{\psi})} + \frac{\int_{N}^{c/(a_{22}-a_{12})} (1+\varepsilon)cy^{\lambda-1}\mathrm{d}y}{M_{Y_1}(a_{22}\sqrt{\psi})}
$$

$$
= \frac{\exp(a_{22}\sqrt{\psi}N) F_{Y_1}(N)}{M_{Y_1}(a_{22}\sqrt{\psi})} + \frac{(1+\varepsilon)c[\{c/(a_{22}-a_{12})\}^{\lambda} - N^{\lambda}]/\lambda}{M_{Y_1}(a_{22}\sqrt{\psi})}. \tag{16}
$$

Substituting c by $c = \frac{\log\{M_{Y_1}(a_{22}\sqrt{\psi})\} - \log\{M_{Y_1}(a_{12}\sqrt{\psi})\}}{\sqrt{\psi}}$, one can easily see that (16) tends to zero as $a_{22} \uparrow 1$. Hence, we have

$$
0 \leq \liminf_{a_{22}\uparrow 1} \frac{\int_{-\infty}^{c/(a_{22}-a_{12})} \exp(a_{22}\sqrt{\psi}y) F_{Y_1}(\mathrm{d}y)}{M_{Y_1}(a_{22}\sqrt{\psi})} \leq \limsup_{a_{22}\uparrow 1} \frac{\int_{-\infty}^{c/(a_{22}-a_{12})} \exp(a_{22}\sqrt{\psi}y) F_{Y_1}(\mathrm{d}y)}{M_{Y_1}(a_{22}\sqrt{\psi})} = 0,
$$

and thus $\lim_{a_{22}\uparrow 1} \frac{\int_{-\infty}^{c/(a_{22}-a_{12})} \exp(a_{22}\sqrt{\psi}y) F_{Y_1}(\mathrm{d}y)}{M_{Y_1}(a_{22}\sqrt{\psi})} = 0$. Note that for $\lambda = 0$, we only need to change the term $(c^{\lambda} - N^{\lambda})/\lambda$ in (16) to $(\log c - \log N)$ and the same conclusion holds. Therefore, for $\lambda \geq 0$, we have $\lim_{a_{22}\uparrow 1} \chi = 0$. This completes the proof. $\qquad\square$

We now present four useful lemmas which are needed for proving Proposition 3.3. The former three concern the analytical solution to a certain convex optimization problem, while the last one presents some properties of the density function of an exponential-tailed distribution function.

Lemma 4. *For $n \geq 3$, let $g_n(x_1, x_2)$ be the minimizer of the following unconstrained optimization problem*

$$
g_n(x_1, x_2) = \min_{x_3,\ldots,x_n} \left| \frac{a_{22}x_1 - a_{12}x_2 + (a_{12}a_{23} - a_{13}a_{22})x_3 + \cdots + (a_{12}a_{2n} - a_{1n}a_{22})x_n}{a_{11}a_{22} - a_{12}a_{21}} \right| +
$$

$$
\left| \frac{-a_{21}x_1 + a_{11}x_2 + (a_{21}a_{13} - a_{23}a_{11})x_3 + \cdots + (a_{21}a_{1n} - a_{2n}a_{11})x_n}{a_{11}a_{22} - a_{12}a_{21}} \right| + \sum_{i=3}^{n} |x_i|
$$

with $a_{11}a_{22} - a_{12}a_{21} \neq 0$, and $a_{1i}^2 + a_{2i}^2 \neq 0, i = 1, \ldots, n$. Then $g_n(x_1, x_2)$ can be expressed explicitly

as

$$g_n(x_1, x_2) = \min_{\substack{i,j=1,\ldots,n \\ i \neq j}} \frac{|a_{2i}x_1 - a_{1i}x_2| + |a_{2j}x_1 - a_{1j}x_2|}{|a_{2i}a_{1j} - a_{1i}a_{2j}|}.$$

Proof. We prove the statement by induction. Denote the objective function in $g_n(x_1, x_2)$ as $h_n(x_1, \ldots, x_n)$. For $n = 3$, we have

$$g_3(x_1, x_2) = \min_{x_3} h_3(x_1, x_2, x_3) = \min_{x_3} \frac{|a_{22}x_1 - a_{12}x_2 + (a_{12}a_{23} - a_{13}a_{22})x_3|}{|a_{11}a_{22} - a_{12}a_{21}|} +$$
$$\frac{|-a_{21}x_1 + a_{11}x_2 + (a_{21}a_{13} - a_{23}a_{11})x_3|}{|a_{11}a_{22} - a_{12}a_{21}|} + |x_3|.$$

Note that if we fix x_1, x_2, then one can partition the real line $\mathbb{R}$ into four parts (two rays and two segments) such that the objective function h_3 is linear in each part. Obviously the minimizer cannot be achieved when $x_3 = \pm\infty$. Hence, it must be achieved at the boundary of one of these segments. That is, if $a_{12}a_{23} - a_{13}a_{22} \neq 0$ and $a_{21}a_{13} - a_{23}a_{11} \neq 0$, then

$$g_3(x_1, x_2) = \min\left\{ h_3(x_1, x_2, 0), h_3\left(x_1, x_2, \frac{a_{22}x_1 - a_{12}x_2}{a_{22}a_{13} - a_{12}a_{23}}\right), h_3\left(x_1, x_2, \frac{a_{21}x_1 - a_{11}x_2}{a_{21}a_{13} - a_{11}a_{23}}\right) \right\}.$$

Clearly, $h_3(x_1, x_2, 0) = (|a_{22}x_1 - a_{12}x_2| + |a_{21}x_1 - a_{11}x_2|)/|a_{22}a_{11} - a_{12}a_{21}|$. Furthermore,

$$h_3\left(x_1, x_2, \frac{a_{22}x_1 - a_{12}x_2}{a_{22}a_{13} - a_{12}a_{23}}\right) = \frac{\left| -a_{21}x_1 + a_{11}x_2 + \frac{(a_{22}x_1 - a_{12}x_2)(a_{21}a_{13} - a_{23}a_{11})}{a_{22}a_{13} - a_{12}a_{23}} \right|}{|a_{11}a_{22} - a_{12}a_{21}|} + \left| \frac{a_{22}x_1 - a_{12}x_2}{a_{22}a_{13} - a_{12}a_{23}} \right|$$
$$= \left| \frac{a_{23}x_1 - a_{13}x_2}{a_{22}a_{13} - a_{12}a_{23}} \right| + \left| \frac{a_{22}x_1 - a_{12}x_2}{a_{22}a_{13} - a_{12}a_{23}} \right|.$$

Similarly, we have

$$h_3\left(x_1, x_2, \frac{a_{21}x_1 - a_{11}x_2}{a_{21}a_{13} - a_{11}a_{23}}\right) = \left| \frac{a_{23}x_1 - a_{13}x_2}{a_{13}a_{21} - a_{11}a_{23}} \right| + \left| \frac{a_{21}x_1 - a_{11}x_2}{a_{13}a_{21} - a_{11}a_{23}} \right|.$$

Hence,

$$g_3(x_1, x_2) = \min_{\substack{i,j=1,2,3 \\ i \neq j}} \frac{|a_{2i}x_1 - a_{1i}x_2| + |a_{2j}x_1 - a_{1j}x_2|}{|a_{2i}a_{1j} - a_{1i}a_{2j}|}.$$

If $a_{12}a_{23} - a_{13}a_{22} = 0$ and $a_{21}a_{13} - a_{23}a_{11} = 0$, the above statement clearly also holds. If only one of them is zero, without loss of generality we assume $a_{12}a_{23} - a_{13}a_{22} \neq 0$ and $a_{21}a_{13} - a_{23}a_{11} = 0$. Since $a_{11}^2 + a_{21}^2 \neq 0, a_{13}^2 + a_{23}^2 \neq 0$, we further assume $a_{11} \neq 0, a_{13} \neq 0$ without losing generality.

Then $a_{21} = a_{23}a_{11}/a_{13}$, $g_3(x_1, x_2) = \min\left\{h_3(x_1, x_2, 0), h_3\left(x_1, x_2, \frac{a_{22}x_1 - a_{12}x_2}{a_{22}a_{13} - a_{12}a_{23}}\right)\right\}$, and

$$
\begin{aligned}
h_3\left(x_1, x_2, \frac{a_{22}x_1 - a_{12}x_2}{a_{22}a_{13} - a_{12}a_{23}}\right) &= \frac{|-a_{21}x_1 + a_{11}x_2|}{|a_{11}a_{22} - a_{12}a_{21}|} + \left|\frac{a_{22}x_1 - a_{12}x_2}{a_{22}a_{13} - a_{12}a_{23}}\right| \\
&= \left|\frac{a_{23}a_{11}x_1/a_{13} - a_{11}x_2}{a_{11}a_{22} - a_{12}a_{23}a_{11}/a_{13}}\right| + \left|\frac{a_{22}x_1 - a_{12}x_2}{a_{22}a_{13} - a_{12}a_{23}}\right| \\
&= \frac{|a_{23}x_1 - a_{13}x_2| + |a_{22}x_1 - a_{12}x_2|}{|a_{22}a_{13} - a_{12}a_{23}|}.
\end{aligned}
$$

Hence, the statement $g_3(x_1, x_2) = \min_{\substack{i,j=1,2,3 \\ i \neq j}} \frac{|a_{2i}x_1 - a_{1i}x_2| + |a_{2j}x_1 - a_{1j}x_2|}{|a_{2i}a_{1j} - a_{1i}a_{2j}|}$ also holds in this case. Above all, we have shown that

$$
g_3(x_1, x_2) = \min_{\substack{i,j=1,2,3 \\ i \neq j}} \frac{|a_{2i}x_1 - a_{1i}x_2| + |a_{2j}x_1 - a_{1j}x_2|}{|a_{2i}a_{1j} - a_{1i}a_{2j}|}.
$$

Now we assume that $g_n(x_1, x_2) = \min_{\substack{i,j=1,\ldots,n \\ i \neq j}} \frac{|a_{2i}x_1 - a_{1i}x_2| + |a_{2j}x_1 - a_{1j}x_2|}{|a_{2i}a_{1j} - a_{1i}a_{2j}|}$ holds for some $n \geq 3$ and consider the case $n+1$. Since there is no constraint in our optimization problem (thus the constraints are independent), we can minimize the function h_{n+1} by first minimizing over some variables, and then minimizing over the remaining variables; see Section 4.1.3 of Boyd & Vandenberghe (2004) for more details. Hence,

$$
g_{n+1}(x_1, x_2) = \min_{x_3,\ldots,x_{n+1}} h_{n+1}(x_1, \ldots, x_{n+1}) = \min_{x_3,\ldots,x_n} \min_{x_{n+1}} h_{n+1}(x_1, \ldots, x_{n+1}).
$$

Then using the same arguments as $n = 3$, if $a_{12}a_{2(n+1)} - a_{1(n+1)}a_{22} \neq 0$, $a_{21}a_{1(n+1)} - a_{2(n+1)}a_{11} \neq 0$,

we have

$$g_{n+1}(x_1, x_2) = \min_{x_3,\ldots,x_n} \min \Big\{ h_{n+1}(x_1,\ldots,x_n,0),$$

$$h_{n+1}\Big(x_1,\ldots,x_n, \frac{a_{22}x_1 - a_{12}x_2 + \sum_{i=3}^n (a_{12}a_{2i} - a_{1i}a_{22})x_i}{a_{1(n+1)}a_{22} - a_{12}a_{2(n+1)}}\Big),$$

$$h_{n+1}\Big(x_1,\ldots,x_n, \frac{a_{21}x_1 - a_{11}x_2 + \sum_{i=3}^n (a_{11}a_{2i} - a_{1i}a_{21})x_i}{a_{1(n+1)}a_{21} - a_{11}a_{2(n+1)}}\Big)\Big\}$$

$$= \min \Big\{ \min_{x_3,\ldots,x_n} h_{n+1}(x_1,\ldots,x_n,0),$$

$$\min_{x_3,\ldots,x_n} h_{n+1}\Big(x_1,\ldots,x_n, \frac{a_{22}x_1 - a_{12}x_2 + \sum_{i=3}^n (a_{12}a_{2i} - a_{1i}a_{22})x_i}{a_{1(n+1)}a_{22} - a_{12}a_{2(n+1)}}\Big),$$

$$\min_{x_3,\ldots,x_n} h_{n+1}\Big(x_1,\ldots,x_n, \frac{a_{21}x_1 - a_{11}x_2 + \sum_{i=3}^n (a_{11}a_{2i} - a_{1i}a_{21})x_i}{a_{1(n+1)}a_{21} - a_{11}a_{2(n+1)}}\Big)\Big\}.$$

Note that $\min_{x_3,\ldots,x_n} h_{n+1}(x_1,\ldots,x_n,0) = \min_{x_3,\ldots,x_n} h_n(x_1,\ldots,x_n) = g_n(x_1, x_2)$, and

$$\min_{x_3,\ldots,x_n} h_{n+1}\Big(x_1,\ldots,x_n, \frac{a_{22}x_1 - a_{12}x_2 + \sum_{i=3}^n (a_{12}a_{2i} - a_{1i}a_{22})x_i}{a_{1(n+1)}a_{22} - a_{12}a_{2(n+1)}}\Big)$$

$$= \min_{x_3,\ldots,x_n} \Big| \frac{-a_{21}x_1 + a_{11}x_2 + \sum_{i=1}^n (a_{21}a_{1i} - a_{2i}a_{11})x_i + \frac{a_{21}a_{1(n+1)} - a_{2(n+1)}a_{11}}{a_{1(n+1)}a_{22} - a_{12}a_{2(n+1)}}\{a_{22}x_1 - a_{12}x_2 + }{a_{11}a_{22} - a_{12}a_{21}}$$

$$\frac{\sum_{i=3}^n (a_{12}a_{2i} - a_{1i}a_{22})x_i\}}{a_{11}a_{22} - a_{12}a_{21}} \Big| + \sum_{i=3}^n |x_i| + \Big| \frac{a_{22}x_1 - a_{12}x_2 + \sum_{i=3}^n (a_{12}a_{2i} - a_{1i}a_{22})x_i}{a_{1(n+1)}a_{22} - a_{12}a_{2(n+1)}} \Big|$$

$$= \min_{x_3,\ldots,x_n} \Big| \frac{a_{2(n+1)}x_1 - a_{1(n+1)}x_2 + \sum_{i=3}^n (a_{1(n+1)}a_{2i} - a_{1i}a_{2(n+1)})x_i}{a_{1(n+1)}a_{22} - a_{12}a_{2(n+1)}} \Big| + \sum_{i=3}^n |x_i| +$$

$$\Big| \frac{a_{22}x_1 - a_{12}x_2 + \sum_{i=3}^n (a_{12}a_{2i} - a_{1i}a_{22})x_i}{a_{1(n+1)}a_{22} - a_{12}a_{2(n+1)}} \Big|,$$

which is equal to $g_n(x_1, x_2)$ if we replace the coefficients a_{11}, a_{21} in $g_n(x_1, x_2)$ with $a_{1(n+1)}$ and $a_{2(n+1)}$, respectively. Hence,

$$\min_{x_3,\ldots,x_n} h_{n+1}\Big(x_1,\ldots,x_n, \frac{a_{22}x_1 - a_{12}x_2 + \sum_{i=3}^n (a_{12}a_{2i} - a_{1i}a_{22})x_i}{a_{1(n+1)}a_{22} - a_{12}a_{2(n+1)}}\Big)$$

$$= \min_{\substack{i,j=2,\ldots,n+1 \\ i \neq j}} \frac{|a_{2i}x_1 - a_{1i}x_2| + |a_{2j}x_1 - a_{1j}x_2|}{|a_{2i}a_{1j} - a_{1i}a_{2j}|}.$$

Similarly, we have

$$\min_{x_3,\dots,x_n} h_{n+1}\Big(x_1,\dots,x_n, \frac{a_{21}x_1 - a_{11}x_2 + \sum_{i=3}^{n}(a_{11}a_{2i} - a_{1i}a_{21})x_i}{a_{1(n+1)}a_{21} - a_{11}a_{2(n+1)}}\Big)$$

$$= \min_{\substack{i,j=1,3,\dots,n+1 \\ i\neq j}} \frac{|a_{2i}x_1 - a_{1i}x_2| + |a_{2j}x_1 - a_{1j}x_2|}{|a_{2i}a_{1j} - a_{1i}a_{2j}|}.$$

Above all,

$$g_{n+1}(x_1, x_2) = \min\Big\{ \min_{\substack{i,j=1,\dots,n \\ i\neq j}} \frac{|a_{2i}x_1 - a_{1i}x_2| + |a_{2j}x_1 - a_{1j}x_2|}{|a_{2i}a_{1j} - a_{1i}a_{2j}|},$$

$$\min_{\substack{i,j=2,\dots,n+1 \\ i\neq j}} \frac{|a_{2i}x_1 - a_{1i}x_2| + |a_{2j}x_1 - a_{1j}x_2|}{|a_{2i}a_{1j} - a_{1i}a_{2j}|}, \min_{\substack{i,j=1,3,\dots,n+1 \\ i\neq j}} \frac{|a_{2i}x_1 - a_{1i}x_2| + |a_{2j}x_1 - a_{1j}x_2|}{|a_{2i}a_{1j} - a_{1i}a_{2j}|}\Big\}$$

$$= \min_{\substack{i,j=1,\dots,n+1 \\ i\neq j}} \frac{|a_{2i}x_1 - a_{1i}x_2| + |a_{2j}x_1 - a_{1j}x_2|}{|a_{2i}a_{1j} - a_{1i}a_{2j}|}.$$

If $a_{12}a_{2(n+1)} - a_{1(n+1)}a_{22} = a_{21}a_{1(n+1)} - a_{2(n+1)}a_{11} = 0$ or only one of them is zero, it is straightforward to show that the above statement also holds by using the argument as in the case $n = 3$. Therefore, the proof is complete. $\square$

Lemma 5. Let $a, b, c, d \in [0, \infty)$ with $ad - bc \neq 0$, and $f(x_1, x_2) = \frac{|ax_1 - bx_2| + |cx_1 - dx_2|}{|ad - bc|}$, then we have

$$\min_{x_1 \geq 1, x_2 \geq 1} f(x_1, x_2) = \min\Big\{ \max(1/a, 1/b), \max(1/c, 1/d), \frac{|a - b| + |c - d|}{|ad - bc|}\Big\}.$$

Proof. Since one can always partition $\mathbb{R}^2$ into four subregions such that the function $f(x_1, x_2)$ is linear in each subregion, we know that the minimizer of this function on $[1, \infty)^2$ can only be achieved at the intersection of one subregion and $[1, \infty)^2$. That is, if $a \leq b, c \leq d$, then

$$\min_{x_1 \geq 1, x_2 \geq 1} f(x_1, x_2) = \min\{f(1, 1), f(b/a, 1), f(d/c, 1)\}$$

$$= \min\Big\{ \frac{|a - b| + |c - d|}{|ad - bc|}, 1/a, 1/c\Big\}$$

$$= \min\Big\{ \frac{|a - b| + |c - d|}{|ad - bc|}, \max(1/a, 1/b), \max(1/c, 1/d)\Big\}.$$

Similarly, for $b \leq a, c \leq d$, or $a \leq b, d \leq c$, or $b \leq a, d \leq c$, we also obtain

$$\min_{x_1 \geq 1, x_2 \geq 1} f(x_1, x_2) = \min\left\{ \frac{|a - b| + |c - d|}{|ad - bc|}, \max(1/a, 1/b), \max(1/c, 1/d) \right\}.$$

Therefore, above all, the conclusion always holds. $\qquad\square$

Lemma 6. *Let $a, b, c, d \in [0, 1]$ with $ad - bc \neq 0$, $ab < 1$ and $cd < 1$, then*

$$\frac{|a - b| + |c - d|}{|ad - bc|} > 1.$$

Proof. Since $ad - bc \neq 0$, we know that $a - b$ and $c - d$ cannot be both equal to zero. Hence,

$$|ad - bc| = |ad - ac + ac - bc| \leq a|c - d| + c|a - b| < |c - d| + |a - b|,$$

where the last inequality is due to the fact that $ab < 1$ and $cd < 1$, which means that a and b cannot be both equal to 1, and c and d cannot be both equal to 1. This implies that $\frac{|a-b|+|c-d|}{|ad-bc|} > 1$. $\quad\square$

Lemma 7. *Let Y be a random variable with distribution function F_Y such that $\bar{F}_Y \in \mathscr{L}_\beta, \beta > 0$. If F_Y is absolutely continuous with density function f_Y, then we have the asymptotic relationship*

$$\log \bar{F}_Y(x) \sim \log f_Y(x) \sim -\beta x, \quad x \to \infty.$$

Proof. Since $\bar{F}_Y \in \mathscr{L}_\beta$, using L'Hôpital's rule, we know that $f_Y \in \mathscr{L}_\beta$. Now suppose that $g : \mathbb{R} \to \mathbb{R}_+$ is an exponential-tailed function with index β, i.e., $g \in \mathscr{L}_\beta$. Using the representation (5), we have

$$g(x) = a(x) \exp\left\{ -\int_0^x \beta(v) \mathrm{d}v \right\},$$

where $a(x) \to a \in (0, \infty)$ and $\beta(x) \to \beta$ as $x \to \infty$. The fact $\beta(x) \to \beta$ implies that $\beta(\cdot)$ is slowly regularly. By Karamata's theorem (see Theorem 2.1 in Resnick (2007)), we have that

$$\int_0^x \beta(v) \mathrm{d}v \sim \beta x, \quad x \to \infty.$$

Hence,

$$\log g(x) = \log a(x) - \int_0^x \beta(v)\mathrm{d}v \sim - \int_0^x \beta(v)\mathrm{d}v \sim -\beta x, \quad x \to \infty.$$

Since both f_Y and $\bar{F}_Y$ have exponential tails with index β, we have the asymptotic expansion $\log \bar{F}_Y(x) \sim \log f_Y(x) \sim -\beta x$, $x \to \infty$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \square$

We are now ready to prove Proposition 3.3.

Proof of Proposition 3.3. If (6), then clearly $\eta = 1$, which coincides with the result in Proposition 3.1. Otherwise, without losing generality we assume that $a_{11} = \max_{i \in \{1,\dots,n\}} a_{1i}$ and $a_{22} = \max_{i \in \{1,\dots,n\}} a_{2i}$. Since extremal dependence is a copula property, $(X_1/a_{11}, X_2/a_{22})$ has the same extremal dependence structure as $\boldsymbol{X}$. Hence, we can further assume that $a_{11} = a_{22} = 1$ and $0 \le a_{ji} \le 1$, $j = 1, 2$, $i = 1, \dots, n$. Since $\arg\max_{i \in \{1,\dots,n\}} a_{1i} \cap \arg\max_{i \in \{1,\dots,n\}} a_{2i} = \emptyset$, we know that $a_{1i}a_{2i} < 1$.

The strategy for this proof relies on augmenting the model $X_1 = Y_1 + a_{12}Y_2 + \cdots + a_{1n}Y_n$, $X_2 = a_{21}Y_1 + Y_2 + a_{23}Y_3 + \cdots + a_{2n}Y_n$ in the following way

$$\begin{cases} X_1 = Y_1 + a_{12}Y_2 + \cdots + a_{1n}Y_n, \\ X_2 = a_{21}Y_1 + Y_2 + a_{23}Y_3 + \cdots + a_{2n}Y_n, \\ X_3 = Y_3, \\ \quad \vdots \\ X_n = Y_n. \end{cases}$$

Since $Y_i, i = 1, \dots, n$ are independent and have common distribution function $\bar{F}_Y \in \mathscr{L}_\beta$, by Theorem 3 in Embrechts & Goldie (1980), we know that $\bar{F}_{X_i} \in \mathscr{L}_\beta, i = 1, \dots, n$. Using Lemma 7, we have

$$1 - F_{X_i}(x) = e^{-h(x)}, \quad h(x) \sim \beta x, \quad \text{as } x \to \infty.$$

Now for the square matrix

$$
A = \begin{pmatrix}
1 & a_{12} & a_{13} & \cdots & a_{1n} \\
a_{21} & 1 & a_{23} & \cdots & a_{2n} \\
 & & 1 & & \\
 & & & \ddots & \\
 & & & & 1
\end{pmatrix} \in \mathbb{R}^{n \times n},
$$

it can be shown that its determinant is $|A| = 1 - a_{12}a_{21}$, and its inverse is

$$
A^{-1} = \frac{1}{1 - a_{12}a_{21}} \begin{pmatrix}
1 & -a_{12} & a_{12}a_{23} - a_{13} & \cdots & a_{12}a_{2n} - a_{1n} \\
-a_{21} & 1 & a_{21}a_{13} - a_{23} & \cdots & a_{21}a_{1n} - a_{2n} \\
 & & 1 - a_{12}a_{21} & & \\
 & & & \ddots & \\
 & & & & 1 - a_{12}a_{21}
\end{pmatrix}.
$$

Denote $\boldsymbol{X} = (X_1, X_2)^T$, $\tilde{\boldsymbol{X}} = (X_1, \ldots, X_n)^T$ and the probability density function of $\tilde{\boldsymbol{X}}$ as $f_{\tilde{\boldsymbol{X}}}$. Then we have

$$
\lim_{t \to \infty} \frac{-\log f_{\tilde{\boldsymbol{X}}}(t\boldsymbol{x})}{h(t)} = \lim_{t \to \infty} \frac{\beta t(|\boldsymbol{\alpha}_1^T \boldsymbol{x}| + \cdots + |\boldsymbol{\alpha}_n^T \boldsymbol{x}|)}{\beta t}
$$
$$
= |\boldsymbol{\alpha}_1^T \boldsymbol{x}| + \cdots + |\boldsymbol{\alpha}_n^T \boldsymbol{x}|
$$
$$
=: \tilde{g}(\boldsymbol{x}),
$$

where $\boldsymbol{\alpha}_i, i = 1, \ldots, n$ are the ith row vector of the matrix A^{-1}. By Proposition 2 in Nolde & Wadsworth (2022), a sequence of scaled random samples $N_n = \{\tilde{\boldsymbol{X}}_1/r_n, \ldots, \tilde{\boldsymbol{X}}_n/r_n\}$ from $f_{\tilde{\boldsymbol{X}}}$ converges in probability onto a limit set $\tilde{G}$ with $\tilde{G} = \{\boldsymbol{x} \in \mathbb{R}^d : \tilde{g}(\boldsymbol{x}) \leq 1\}$. Then using Proposition 4 in Nolde & Wadsworth (2022) we know that, for $\boldsymbol{X}$, which is a two-dimensional subvector of $\tilde{\boldsymbol{X}}$, sample clouds from $\boldsymbol{X}$ converge onto the limit set $G = \{x_1, x_2 \in \mathbb{R} : g(x_1, x_2) \leq 1\}$ with gauge function

$$
g(x_1, x_2) = \min_{x_3, \ldots, x_n} \tilde{g}(\boldsymbol{x}).
$$

Therefore, by Proposition 8 in Nolde & Wadsworth (2022), we have

$$\eta^{-1} = \min_{x_1,x_2\geq 1} g(x_1, x_2)$$

$$= \min_{x_1,x_2\geq 1} \min_{x_3,\ldots,x_n} \frac{|x_1 - a_{12}x_2 + (a_{12}a_{23} - a_{13})x_3 + \cdots + (a_{12}a_{2n} - a_{1n})x_n|}{1 - a_{12}a_{21}}_+$$

$$\frac{|x_2 - a_{21}x_1 + (a_{21}a_{13} - a_{23})x_3 + \cdots + (a_{21}a_{1n} - a_{2n})x_n)|}{1 - a_{12}a_{21}} + \sum_{i=3}^{n} |x_i|.$$

Using Lemma 4 and 5, we have

$$\eta = \left\{ \min_{x_1,x_2\geq 1} \min_{\substack{i,j=1,\ldots,n \\ i\neq j}} \frac{|a_{2i}x_1 - a_{1i}x_2| + |a_{2j}x_1 - a_{1j}x_2|}{|a_{2i}a_{1j} - a_{1i}a_{2j}|} \right\}^{-1}$$

$$= \left\{ \min_{\substack{i,j=1,\ldots,n \\ i\neq j}} \min_{x_1,x_2\geq 1} \frac{|a_{2i}x_1 - a_{1i}x_2| + |a_{2j}x_1 - a_{1j}x_2|}{|a_{2i}a_{1j} - a_{1i}a_{2j}|} \right\}^{-1}$$

$$= \left[\min_{\substack{i,j=1,\ldots,n \\ i\neq j}} \min \left\{ \frac{|a_{2i} - a_{1i}| + |a_{2j} - a_{1j}|}{|a_{2i}a_{1j} - a_{1i}a_{2j}|}, \max(1/a_{1i}, 1/a_{2i}), \max(1/a_{1j}, 1/a_{2j}) \right\} \right]^{-1}.$$

If (7), then by Lemma 6, we have $\eta < 1$ and thus X_1 and X_2 are asymptotically independent.

Notice that although we have assumed $a_{11} = \max_{i\in\{1,\ldots,n\}} a_{1i}$ and $a_{22} = \max_{i\in\{1,\ldots,n\}} a_{2i}$ at the beginning of this proof, they are only needed to ensure that $a_{11}a_{22} - a_{12}a_{21} \neq 0$, which implies that the matrix A is nonsingular, and they do not affect the expression of η. Therefore, the proof is complete. □

Appendix B   Proofs for Section 4

We first present the proof of Theorem 4.1.

Proof of Theorem 4.1. By Proposition 3.3, we know that the residual tail dependence coefficient of $(u_n(\boldsymbol{s}_1), u_n(\boldsymbol{s}_2))^\top$ is

$$\eta_n(\boldsymbol{s}_1, \boldsymbol{s}_2) = \max_{i,j\in\{1,\ldots,J\},i\neq j} \left\{ \frac{|\tilde{a}_{1i}\tilde{a}_{2j} - \tilde{a}_{1j}\tilde{a}_{2i}|}{|\tilde{a}_{1i} - \tilde{a}_{2i}| + |\tilde{a}_{1j} - \tilde{a}_{2j}|}, \tilde{a}_{1i} \wedge \tilde{a}_{2i}, \tilde{a}_{1j} \wedge \tilde{a}_{2j} \right\},$$

where $\tilde{a}_{ri} = \frac{G(\|\boldsymbol{s}_r - \boldsymbol{d}_i\|)}{G(l_r)}, r = 1, 2, i = 1, \ldots, J$ and $l_r = \min_{i\in\{1,\ldots,J\}} \|\boldsymbol{s}_r - \boldsymbol{d}_i\|$. If $l_1 \neq l_2$, say $l_1 < l_2$, then one can add one or more mesh nodes close to $\boldsymbol{s}_2$ such that $l_1 = l'_2$ holds in the resulting mesh.
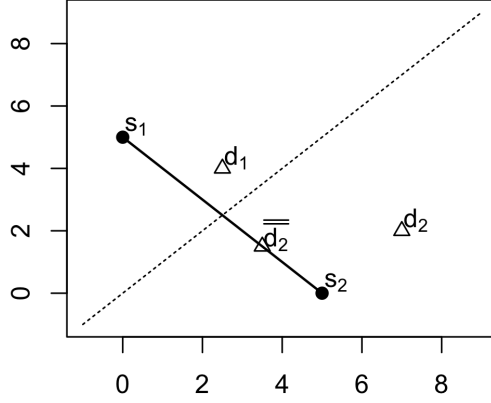
Figure 6: Explaination for the derivation of the limiting η in the type G model.

So without loss of generality, we assume that $l_1 = l_2 = l$.

For fixed $l > 0$, it is clear that $\tilde{a}_{1i} \wedge \tilde{a}_{2i} \leq G(\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|/2)/G(l)$, where the equality is achieved when $\boldsymbol{d}_i$ is at the midpoint of $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$, i.e., $\|\boldsymbol{s}_1 - \boldsymbol{d}_i\| = \|\boldsymbol{s}_2 - \boldsymbol{d}_i\| = \|\boldsymbol{s}_1 - \boldsymbol{s}_2\|/2$. Since our interest is in the case when $l \to 0$, we assume that l is small and $l < \|\boldsymbol{s}_1 - \boldsymbol{s}_2\|/2$. We now focus on the term $\frac{|\tilde{a}_{1i}\tilde{a}_{2j} - \tilde{a}_{1j}\tilde{a}_{2i}|}{|\tilde{a}_{1i} - \tilde{a}_{2i}| + |\tilde{a}_{1j} - \tilde{a}_{2j}|}$. Substitute $\tilde{a}_{ri} = \frac{G(\|\boldsymbol{s}_r - \boldsymbol{d}_i\|)}{G(l)}$, and denote this term by f, i.e.,

$$f(\boldsymbol{d}_1, \boldsymbol{d}_2) = \frac{|G(\|\boldsymbol{d}_1 - \boldsymbol{s}_1\|)G(\|\boldsymbol{d}_2 - \boldsymbol{s}_2\|) - G(\|\boldsymbol{d}_1 - \boldsymbol{s}_2\|)G(\|\boldsymbol{d}_2 - \boldsymbol{s}_1\|)|}{|G(\|\boldsymbol{d}_1 - \boldsymbol{s}_1\|) - G(\|\boldsymbol{d}_1 - \boldsymbol{s}_2\|)|G(l) + |G(\|\boldsymbol{d}_2 - \boldsymbol{s}_1\|) - G(\|\boldsymbol{d}_2 - \boldsymbol{s}_2\|)|G(l)}.$$

Then we need to find the maximum of this function for $\boldsymbol{d}_1, \boldsymbol{d}_2 \in \mathbb{R}^2 \setminus (\mathcal{B}(\boldsymbol{s}_1, l) \cup \mathcal{B}(\boldsymbol{s}_2, l))$, where $\mathcal{B}(\boldsymbol{s}_r, l) = \{\boldsymbol{x} \in \mathbb{R}^2 : \|\boldsymbol{x} - \boldsymbol{s}_r\| < l\}$ is the open ball of radius l centered at $\boldsymbol{s}_r$.

Denote the line segment from $\boldsymbol{s}_1$ to $\boldsymbol{s}_2$ by $LS(\boldsymbol{s}_1, \boldsymbol{s}_2)$, and its perpendicular bisector by $PB(\boldsymbol{s}_1, \boldsymbol{s}_2)$. If $\boldsymbol{d}_1 \in PB(\boldsymbol{s}_1, \boldsymbol{s}_2)$ and $\boldsymbol{d}_2 \in PB(\boldsymbol{s}_1, \boldsymbol{s}_2)$, then clearly the function f is not well-defined. By Proposition 3.3 and the discussion followed, we know that these points can be neglected. If only one of $\boldsymbol{d}_1$ and $\boldsymbol{d}_2$ is on the line $PB(\boldsymbol{s}_1, \boldsymbol{s}_2)$, then we have $f(\boldsymbol{d}_1, \boldsymbol{d}_2) = G(\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|/2)/G(l)$.

For $\boldsymbol{d}_1, \boldsymbol{d}_2 \notin PB(\boldsymbol{s}_1, \boldsymbol{s}_2)$, without loss of generality we assume that $\boldsymbol{d}_1$ is closer to $\boldsymbol{s}_1$ than to $\boldsymbol{s}_2$, i.e., $\|\boldsymbol{d}_1 - \boldsymbol{s}_1\| < \|\boldsymbol{d}_1 - \boldsymbol{s}_2\|$, as shown in Figure 6. One can show that if $\boldsymbol{d}_2$ is also closer to $\boldsymbol{s}_1$ than to $\boldsymbol{s}_2$, then its symmetric point $\bar{\boldsymbol{d}}_2$ about the line $PB(\boldsymbol{s}_1, \boldsymbol{s}_2)$ satisfies $f(\boldsymbol{d}_1, \bar{\boldsymbol{d}}_2) \geq f(\boldsymbol{d}_1, \boldsymbol{d}_2)$.

39

This implies that to obtain the maximum of f, it is sufficient to consider points $\boldsymbol{d}_1, \boldsymbol{d}_2$ located on the two sides of the line $PB(\boldsymbol{s}_1, \boldsymbol{s}_2)$. Hence we can assume that $\boldsymbol{d}_1$ is closer to $\boldsymbol{s}_1$ than to $\boldsymbol{s}_2$, and $\boldsymbol{d}_2$ is closer to $\boldsymbol{s}_2$ than to $\boldsymbol{s}_1$. Since G is strictly decreasing on $[0, \infty)$, the function f can then be simplified to

$$f(\boldsymbol{d}_1, \boldsymbol{d}_2) = \frac{G(\|\boldsymbol{d}_1 - \boldsymbol{s}_1\|)G(\|\boldsymbol{d}_2 - \boldsymbol{s}_2\|) - G(\|\boldsymbol{d}_1 - \boldsymbol{s}_2\|)G(\|\boldsymbol{d}_2 - \boldsymbol{s}_1\|)}{\{G(\|\boldsymbol{d}_1 - \boldsymbol{s}_1\|) - G(\|\boldsymbol{d}_1 - \boldsymbol{s}_2\|) + G(\|\boldsymbol{d}_2 - \boldsymbol{s}_1\|) - G(\|\boldsymbol{d}_2 - \boldsymbol{s}_2\|)\}G(l)}.$$

Furthermore, due to the convexity of G, one can show that for $\boldsymbol{d} \in \mathbb{R}^2 \setminus (\mathcal{B}(\boldsymbol{s}_1, l) \cup \mathcal{B}(\boldsymbol{s}_2, l))$, the range of the function $|G(\|\boldsymbol{d} - \boldsymbol{s}_1\|) - G(\|\boldsymbol{d} - \boldsymbol{s}_2\|)|$ is $[0, G(l) - G(\|\boldsymbol{s}_1 - \boldsymbol{s}_2\| - l)]$ with its minimum obtained on the line $PB(\boldsymbol{s}_1, \boldsymbol{s}_2)$ and its maximum at the intersection points of the line passing through $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ and the two circles with radius l centered at $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$. Now if we fix the point $\boldsymbol{d}_1$, then by the intermediate value theorem (Tao, 2016, Theorem 9.7.1), for any $\boldsymbol{d}_2$ such that $\|\boldsymbol{d}_2 - \boldsymbol{s}_2\| < \|\boldsymbol{d}_2 - \boldsymbol{s}_1\|$ and $\boldsymbol{d}_2 \notin \mathcal{B}(\boldsymbol{s}_2, l)$, there always exists one point $\bar{\bar{\boldsymbol{d}}}_2$ on the line segment $LS((\boldsymbol{s}_1 + \boldsymbol{s}_2)/2, \boldsymbol{s}_2) \setminus \mathcal{B}(\boldsymbol{s}_2, l)$ such that $G(\|\boldsymbol{d}_2 - \boldsymbol{s}_1\|) - G(\|\boldsymbol{d}_2 - \boldsymbol{s}_2\|) = G(\|\bar{\bar{\boldsymbol{d}}}_2 - \boldsymbol{s}_1\|) - G(\|\bar{\bar{\boldsymbol{d}}}_2 - \boldsymbol{s}_2\|)$, and $G(\|\bar{\bar{\boldsymbol{d}}}_2 - \boldsymbol{s}_i\|) \geq G(\|\boldsymbol{d}_2 - \boldsymbol{s}_i\|)$, $i = 1, 2$, which yields $f(\boldsymbol{d}_1, \bar{\bar{\boldsymbol{d}}}_2) \geq f(\boldsymbol{d}_1, \boldsymbol{d}_2)$. Hence, it is sufficient to consider only points $\boldsymbol{d}_1 \in LS(\boldsymbol{s}_1, (\boldsymbol{s}_1 + \boldsymbol{s}_2)/2) \setminus \mathcal{B}(\boldsymbol{s}_1, l)$ and $\boldsymbol{d}_2 \in LS(\boldsymbol{s}_2, (\boldsymbol{s}_1 + \boldsymbol{s}_2)/2) \setminus \mathcal{B}(\boldsymbol{s}_2, l)$, and the optimization problem can be rewritten as

$$\sup_{x_1, x_2 \in [l, h)} \bar{f}(x_1, x_2) = \sup_{x_1, x_2 \in [l, h)} \frac{G(x_1)G(x_2) - G(2h - x_1)G(2h - x_2)}{\{G(x_1) - G(2h - x_1) + G(x_2) - G(2h - x_2)\}G(l)}.$$

Since G is absolutely continuous, we have the partial derivative of $\bar{f}$ with respect to x_1 as

$$\frac{\partial \bar{f}}{\partial x_1} = a(x_1, x_2)[G'(x_1)\{G(x_2) - G(2h - x_2)\} - G'(2h - x_1)\{G(x_1) - G(2h - x_2)\}],$$

where $a(x_1, x_2) = \frac{G(x_2) - G(2h - x_2)}{\{G(x_1) - G(2h - x_1) + G(x_2) - G(2h - x_2)\}^2 G(l)} > 0$. For $x \in [l, h)$, due to the convexity of G, we have that $G'(x) \leq G'(2h - x)$. Hence,

$$\frac{\partial \bar{f}}{\partial x_1} \leq a(x_1, x_2)G'(2h - x_1)\{G(x_2) + G(2h - x_2) - G(x_1) - G(2h - x_1)\}.$$

As G is monotonically decreasing, we know that $G'(2h - x_1) \leq 0$. Furthermore, the convexity of G implies that the function $g(x) = G(x) + G(2h - x)$ is monotonically decreasing on $[l, h)$. Therefore,

for fixed $x_2 \in [l, h)$, we know that if $x_1 \geq x_2$, then $\frac{\partial \bar{f}}{\partial x_1} \leq 0$. This implies that

$$\bar{f}(x_1, x_2) \leq \bar{f}(x_2, x_2) = \{G(x_2) + G(2h - x_2)\}/\{2G(l)\} \leq \{G(l) + G(2h - l)\}/\{2G(l)\},$$

where the maximum is obtained when $x_1 = x_2 = l$.

Above all, we have that for $\boldsymbol{d}_1, \boldsymbol{d}_2 \in \mathbb{R}^2 \setminus (\mathcal{B}(\boldsymbol{s}_1, l) \cup \mathcal{B}(\boldsymbol{s}_2, l))$,

$$f(\boldsymbol{d}_1, \boldsymbol{d}_2) \leq \max\left\{\frac{G(\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|/2)}{G(l)}, \frac{G(l) + G(\|\boldsymbol{s}_1 - \boldsymbol{s}_2\| - l)}{2G(l)}\right\} = 1/2 + G(\|\boldsymbol{s}_1 - \boldsymbol{s}_2\| - l)/G(l),$$

where the last equality holds due to the convexity of G, and the maximum of f is achieved when $\boldsymbol{d}_1, \boldsymbol{d}_2$ are the intersection points of the line segment $LS(\boldsymbol{s}_1, \boldsymbol{s}_2)$ and the two circles with radius l centered at $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$. Hence, as the mesh becomes finer and the set of mesh nodes M_n becomes denser in $\mathcal{D}$, which means that $l \to 0$, the residual tail dependence function $\eta(\boldsymbol{s}_1, \boldsymbol{s}_2)$ tends to $1/2 + G(\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|)/G(0)$. This completes the proof. $\qquad\square$

We now prove Theorem 4.2.

Proof of Theorem 4.2. By Proposition 3.3, the residual tail dependence coefficient of the approximation $(u_n(s_1), u_n(s_2))$ of the form (10) is

$$\eta = \max_{i,j=0,\ldots,n_2-1, i \neq j} \left\{\frac{|a_{2i}a_{1j} - a_{1i}a_{2j}|}{|a_{2i} - a_{1i}| + |a_{2j} - a_{1j}|}, \min(a_{1i}, a_{2i}), \min(a_{1j}, a_{2j})\right\},$$

with $a_{1i} = G(s_1 - t_i)/G(s_1 - t_{n_1-1})$ for $0 \leq i \leq n_1 - 1$ and $a_{1i} = 0$ for $n_1 \leq i \leq n_2 - 1$, and $a_{2i} = G(s_2 - t_i)/G(s_2 - t_{n_2-1})$ for $0 \leq i \leq n_2 - 1$.

Since G is strictly decreasing, the sequence $\{a_{1i}, 0 \leq i \leq n_1 - 1\}$ and $\{a_{2i}, 0 \leq i \leq n_2 - 1\}$ are strictly increasing. Hence, $\max_{i=0,\ldots,n_2-1} \min(a_{1i}, a_{2i}) = a_{2(n_1-1)}$. For the first term inside the maximum operation of the expression of η, we have

$$\frac{|a_{2i}a_{1j} - a_{1i}a_{2j}|}{|a_{2i} - a_{1i}| + |a_{2j} - a_{1j}|} \leq \frac{\max(a_{1j}, a_{2j})|a_{2i} - a_{1i}|}{|a_{2i} - a_{1i}| + |a_{2j} - a_{1j}|} = \frac{\max(a_{1j}, a_{2j})}{1 + |a_{2j} - a_{1j}|/|a_{2i} - a_{1i}|}$$

$$\leq \frac{\max(a_{1j}, a_{2j})}{1 + |a_{2j} - a_{1j}|} = \frac{1}{1 + \{1 - \min(a_{1j}, a_{2j})\}/\max(a_{1j}, a_{2j})}$$

$$\leq \frac{1}{2 - a_{2(n_1-1)}},$$

where the equality is obtained when $i = n_2 - 1$ and $j = n_1 - 1$. Thus we have η for $(u_n(s_1), u_n(s_2))^\top$ as

$$\eta = \max\left(\frac{1}{2 - a_{2(n_1-1)}}, a_{2(n_1-1)}\right) = \frac{1}{2 - a_{2(n_1-1)}}.$$

Therefore, as $m_n \to \infty$ and the set of mesh nodes M_n tries to fill up the space in $\mathcal{D}$, we have the limiting residual tail dependence function as $1/\{2 - G(h)/G(0)\}$. $\qquad\square$

Next, we prove Theorem 4.3.

Proof of Theorem 4.3. Without loss of generality we assume that $0 \le s_1 < s_2$. The representation (12) implies that $u(s_2)$ can be represented as the convolution of two independent random variables

$$u(s_2) = e^{-a(s_2-s_1)}u(s_1) + \int_{s_1}^{s_2} e^{-a(s_2-s)}\mathrm{d}z(at).$$

Since $u(s) = V$ and $\bar{F}_V \in \mathscr{L}_\beta$, we know that $u(s_1), u(s_2)$ have exponential tails with index β and $e^{-a(s_2-s_1)}u(s_1)$ has an exponential tail with index $e^{a(s_2-s_1)}\beta$. Denote $V_1 = e^{-a(s_2-s_1)}u(s_1)$. The self-decomposability of V_1 yields its infinite divisibility, which further implies that the Wiener condition is satisfied, i.e. $M_{V_1}(\beta + it) \ne 0, t \in \mathbb{R}$, where M_{V_1} is the moment generating function of V_1; see Theorem 25.17 of (Sato, 1999) for more details. Then Lemma 2.5 in Pakes (2004) yields that $\int_{s_1}^{s_2} e^{-a(s_2-s)}\mathrm{d}z(at)$ must have an exponential tail with index β. Therefore, the result in Proposition 3.3 and Example 2 gives the asymptotic independence between $u(s_1)$ and $u(s_2)$, and their residual tail dependence coefficient is $\eta = 1/(2 - e^{-a(s_2-s_1)})$. $\qquad\square$

In the following we prove Proposition 4.4.

Proof of Proposition 4.4. By Sgibnev (1990) and Shimura & Watanabe (2005), we know that an infinitely divisible distribution is convolution tail equivalent if and only if its normalized Lévy measure is convolution tail equivalent. That is, for any $t \in \mathbb{R}$,

$$\bar{F}_{z(1)} \in \mathscr{S}_\beta \qquad \Leftrightarrow \qquad \frac{\mathbb{1}_{x>1}U_{z(1)}([x,\infty))}{U_{z(1)}([1,\infty))} \in \mathscr{S}_\beta.$$

From Barndorff-Nielsen (1998) we know that if $U_{z(1)}([x,\infty))$ is continuous in x on $(0,\infty))$, then we

have the relation $U_u([x, \infty)) = \int_x^\infty s^{-1} U_{z(1)}([s, \infty)) \mathrm{d}s$. Hence, for $x > 1$,

$$U_u([\log x, \infty)) = \int_{\log x}^\infty s^{-1} U_{z(1)}([s, \infty)) \mathrm{d}s = \int_x^\infty U_{z(1)}([\log t, \infty))/(t \log t) \mathrm{d}t.$$

Since $\frac{\mathbb{1}_{x>1} U_{z(1)}([x, \infty))}{U_{z(1)}([1, \infty))} \in \mathscr{S}_\beta$, we have that $U_{z(1)}([\log t, \infty)) \in \mathrm{R}V_{-\beta}$, which further implies that $U_{z(1)}([\log t, \infty))/(t \log t) \in \mathrm{R}V_{-\beta-1}$. By Karamata's theorem (cf. Theorem 2.1 in Resnick (2007)),

$$\lim_{x \to \infty} \frac{U_{z(1)}([\log x, \infty))/\log x}{\int_x^\infty U_{z(1)}([\log t, \infty))/(t \log t) \mathrm{d}t} = \beta.$$

Hence,

$$\lim_{x \to \infty} \frac{U_u([\log x, \infty))}{U_{z(1)}([\log t, \infty))} = \lim_{x \to \infty} \frac{1}{\beta \log x} = 1/\beta.$$

Therefore, using Theorem 1 in Cline (1986), we have that $\frac{\mathbb{1}_{x>1} U_u([x, \infty))}{U_u([1, \infty))} \in \mathscr{S}_\beta$ and thus $\bar{F}_u \in \mathscr{S}_\beta$.

Now we denote the Lévy measure of $z(t)$ as $U_{z(t)}$. Since $F_{z(1)} \in \mathscr{S}_\beta$, we know that $\frac{\mathbb{1}_{x>0} U_{z(1)}}{U_{z(1)}([1, \infty))} \in \mathscr{S}_\beta$. By the definition of Lévy process, we have $U_{z(t)} = t U_{z(1)}$, for any $t > 0$. Hence, using Theorem 1 in Cline (1986), we know that $\frac{\mathbb{1}_{x>0} U_{z(t)}([x, \infty))}{U_{z(t)}([1, \infty))} \in \mathscr{S}_\beta$ and thus $\bar{F}_{z(t)} \in \mathscr{S}_\beta$. $\qquad \square$

Appendix C   Finite Element Approximations

In this section we introduce the finite element approximation which is the key for efficient simulation and inference of the type G Matérn fields (Bolin, 2014; Bolin & Wallin, 2020).

Similarly to the non-Gaussian OU processes, the type G Matérn SPDE random fields are an important extension of the well-known SPDE-based formulation of Gaussian random fields (Lindgren et al., 2011), aiming to capture more flexible marginal behaviors, different sample path properties, and non-Gaussian dependence structures. Specifically, the generalization from Gaussian white noise to general Lévy noise in the SPDE (13) provides a natural way to construct non-Gaussian random fields, while the differential operator $(\kappa^2 - \Delta)^{\alpha/2}$ ensures the Matérn covariance for the resulting processes.

The main advantage of the SPDE-based representation is that one can approximate its solution by using the finite element method to achieve computationally efficient simulation and inference. To make sense of the SPDE (13), one can think of $\mathcal{M}$ as a homogeneous Lévy basis, namely an

infinitely divisible and independently scattered random measure, and interpret the Equation (13) in a weak sense. That is, for a test function g in an appropriate space, we have

$$\langle g(\boldsymbol{s}), (\kappa^2 - \Delta)^{\alpha/2} u(\boldsymbol{s}) \rangle = \int g(\boldsymbol{s}) \mathcal{M}(\mathrm{d}\boldsymbol{s}),$$

where the equality holds in distribution, and $\langle a(\boldsymbol{s}), b(\boldsymbol{s}) \rangle = \int a(\boldsymbol{s}) b(\boldsymbol{s}) \mathrm{d}\boldsymbol{s}$. To approximate the weak solution numerically efficiently, we consider a bounded domain $\mathcal{D} \in \mathbb{R}^d$ and assume that the operator $(\kappa^2 - \Delta)^{\alpha/2}$ is equipped with suitable boundary conditions (Lindgren et al., 2011; Bolin, 2014). Then one can choose the space of the test functions to be the linear span of a finite number of basis functions $\{\phi_i, i = 1, \ldots, n\}$. For instance, similarly as in Lindgren et al. (2011), ϕ_i can be chosen as the commonly used piecewise linear basis functions induced by a triangulation of the domain $\mathcal{D}$. We then obtain a finite element approximation of the weak solution as

$$u_n(\boldsymbol{s}) = \sum_{i=1}^{n} \omega_i \phi_i(\boldsymbol{s}),$$

where the stochastic weights $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^\top$ satisfy

$$K_\alpha \boldsymbol{\omega} = \left(\int_{\mathcal{D}} \phi_1(\boldsymbol{s}) \mathcal{M}(\mathrm{d}\boldsymbol{s}), \ldots, \int_{\mathcal{D}} \phi_n(\boldsymbol{s}) \mathcal{M}(\mathrm{d}\boldsymbol{s}) \right)^\top,$$

with $K_\alpha = C\{C^{-1}(\kappa^2 C + G)\}^{\alpha/2}$, and the matrices $C, G \in \mathbb{R}^{n \times n}$ having elements $C_{ij} = \langle \phi_i, \phi_j \rangle_{\mathcal{D}}$, and $G_{ij} = \langle \nabla \phi_i, \nabla \phi_j \rangle_{\mathcal{D}}$, respectively.

The random measure $\mathcal{M}$ is said to be of type G if $\mathcal{M}$ evaluated on the unit square $[0, 1]^d$ can be represented as a normal mean-variance mixture, i.e., $\mathcal{M}([0, 1]^d) = \mu + \gamma v + \sqrt{v} Z$, where v is a non-negative infinitely divisible random variable and Z is a Gaussian random variable. This representation of the type G random measure admits a Gaussian distribution for $\boldsymbol{\omega}$ conditioning on the variance process $v(\boldsymbol{s})$. By assuming that v is closed under convolution, the distribution of $\boldsymbol{\omega}$ can be approximated by

$$\boldsymbol{\omega} \mid \bar{\boldsymbol{v}} \sim \mathcal{N}\left(K_\alpha^{-1}(\mu + \gamma \bar{\boldsymbol{v}}), K_\alpha^{-1} \mathrm{diag}(\bar{\boldsymbol{v}}) K_\alpha^{-1} \right)$$

with $\bar{\boldsymbol{v}} = (\bar{v}_1, \ldots, \bar{v}_n)^\top$ and $\bar{v}_j = M_v(\mathcal{D}_j)$, where M_v is the random measure associated with v, and

$\mathcal{D}_j = \{\boldsymbol{s} : \phi_j(\boldsymbol{s}) \geq \phi_i(\boldsymbol{s}), \forall i \neq j\}$. For the GIG distribution introduced in Section 2, the only two subclasses which are closed under convolution are the Gamma distribution and the inverse Gaussian distribution, which lead to the variance Gamma and normal inverse Gaussian random measures, respectively, for $\mathcal{M}$, and these are the two non-Gaussian SPDE models considered in Bolin (2014), Wallin & Bolin (2015) and Bolin & Wallin (2020).

If v has an inverse Gaussian distribution, then $\bar{v}_i, i = 1, \ldots, n$, are independent and inverse Gaussian distributed, and have exponential tails with the same index. Consequently, the approximate weak solution $u_n(\boldsymbol{s})$ can be rewritten as

$$u_n(\boldsymbol{s}) = (\phi_1(\boldsymbol{s}), \ldots, \phi_n(\boldsymbol{s}))\boldsymbol{\omega} = (\phi_1(\boldsymbol{s}), \ldots, \phi_n(\boldsymbol{s}))K_\alpha^{-1}\boldsymbol{Y}, \tag{17}$$

where $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top$, $Y_i, i = 1, \ldots, n$ are independent and normal inverse Gaussian distributed, and have exponential tails with the same index. If v has a Gamma distribution, then we have the same representation as in (17) for the approximate weak solution, but with the difference that $Y_i, i = 1, \ldots, n$ have a variance Gamma distribution.

The preceding analysis shows that characterizing the extremal dependence of the finite element approximation $u_n(\boldsymbol{s})$ is equivalent to characterizing the extremal dependence of the random vector $\boldsymbol{X} = (X_1, X_2)^\top$ constructed as in (4). From the results in Section 3 we know that the extremal dependence between $u_n(\boldsymbol{s}_1)$ and $u_n(\boldsymbol{s}_2)$ depends on whether $\arg\max_{i=1,\ldots,n}\{(\phi_1(\boldsymbol{s}_1), \ldots, \phi_n(\boldsymbol{s}_1))K_\alpha^{-1}\}$ is equal to $\arg\max_{i=1,\ldots,n}\{(\phi_1(\boldsymbol{s}_2), \ldots, \phi_n(\boldsymbol{s}_2))K_\alpha^{-1}\}$. And for any given mesh and basis functions $\{\phi_i\}$, the residual tail dependence coefficient of the finite element approximation can be easily computed using the formula in Proposition 3.3.

Appendix D   Additional Simulation Studies

Here we examine the effect of different mesh constructions on the extremal dependence of the induced finite element approximation of NIG Matérn SPDE models. We choose the same SPDE parameters and NIG noise parameters as in the first simulation study in Section 4.3. As shown in Figure 7, we randomly select 100 sites in the unit square and consider three different mesh constructions covering the sites. The meshes are constructed based on lattices with outer extensions, where meshes 1, 2, and 3 have 25, 100, and 625 lattice nodes, respectively. We then simulate 10^5
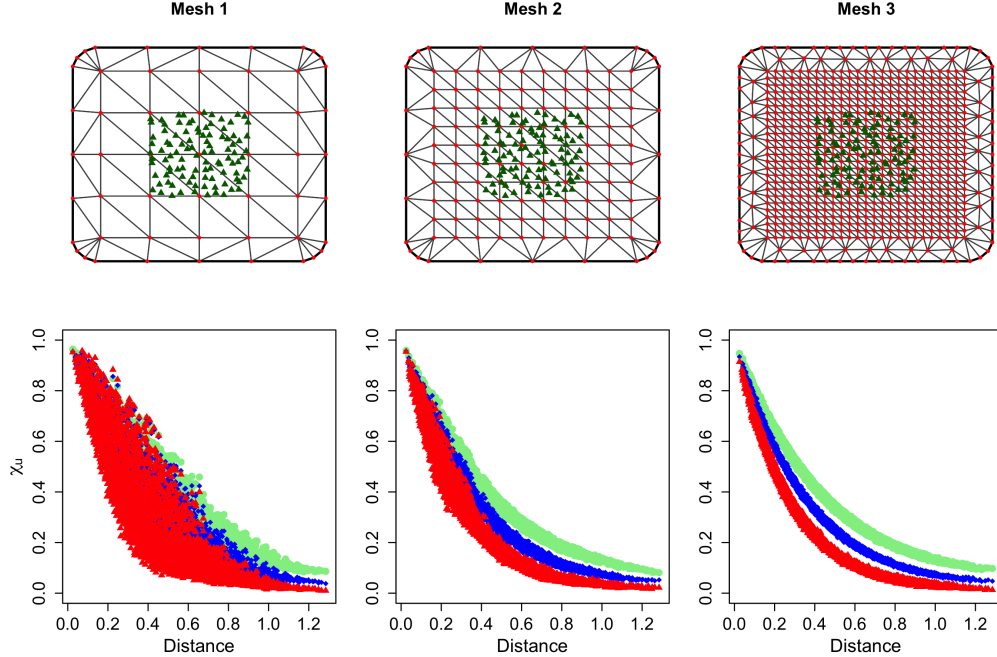
Figure 7: Extremal dependence of the finite element approximations of the NIG Matérn SPDE model with respect to different mesh constructions. Mesh 1, 2, and 3 are constructed on a lattice with respective 25, 100, and 625 nodes. The empirical tail dependence coefficient $\chi(q)$ with different q (green points: $q = 0.95$; blue points: $q = 0.975$; red points: $q = 0.99$) of the finite element approximation model based on mesh 1, 2, and 3 are shown in the lower left, middle, and right panels, respectively.

observations from the finite element approximation model and the lower panel in Figure 7 depicts the empirical tail dependence coefficient of all pairs of sites with respect to different quantile levels $u = 0.95$ (green points), 0.975 (blue points), and 0.99 (red points). This figure clearly illustrates the nonstationarity induced by a coarse mesh (mesh 1 and 2), i.e., various values of $\chi(q)$ exist for the same distance, but this nonstationarity tends to vanish as the mesh becomes finer. One can also observe in the lower right panel that as the quantile level u increases to 1, $\chi(q)$ tends to decrease. This seems to coincide with our theoretical findings of the integral approximation in Section 4.3, whose limiting extremal dependence class is asymptotic independence and which necessarily implies that $\chi = \lim_{q \uparrow 1} \chi(q) = 0$ for all pairs.

Appendix E   Additional Results

Proposition E.1. *Let χ be the tail dependence coefficient in Example 1. If $a_{12} \neq a_{22}$, then $\chi < 1$.*

Proof. Without loss of generality, we assume that $0 \leq a_{12} < a_{22} < 1$. To prove $\chi < 1$, we first show

46

that $c > 0$. Note that for $0 < t < \sqrt{\psi}$, we have

$$
\begin{aligned}
M'_{Y_1}(t) &= \int_{-\infty}^{\infty} y \exp(ty) F_{Y_1}(\mathrm{d}y) \\
&= \int_{-\infty}^{0} y \exp(ty) F_{Y_1}(\mathrm{d}y) + \int_{0}^{\infty} y \exp(ty) F_{Y_1}(\mathrm{d}y) \\
&> \int_{-\infty}^{0} y F_{Y_1}(\mathrm{d}y) + \int_{0}^{\infty} y F_{Y_1}(\mathrm{d}y) = 0,
\end{aligned}
$$

where the last equality is due to the symmetry of the distribution $\mathrm{GH}(\lambda, \tau, \psi, \mu = 0, \gamma = 0)$. Hence, $c = \{\log M_{Y_1}(a_{22}\sqrt{\psi}) - \log M_{Y_1}(a_{12}\sqrt{\psi})\}/\sqrt{\psi} > 0$.

Therefore,

$$
\begin{aligned}
1 - \chi &= 1 - \frac{\int_{(a_{22}-a_{12})y \le c} \exp(a_{22}\sqrt{\psi}y) F_{Y_1}(\mathrm{d}y)}{M_{Y_1}(a_{22}\sqrt{\psi})} - \frac{\int_{(a_{22}-a_{12})y > c} \exp(a_{12}\sqrt{\psi}y) F_{Y_1}(\mathrm{d}y)}{M_{Y_1}(a_{12}\sqrt{\psi})} \\
&= \int_{c/(a_{22}-a_{12})}^{\infty} e^{\sqrt{\psi}a_{22}y - \log M_{Y_1}(a_{22}\sqrt{\psi})} - e^{\sqrt{\psi}a_{12}y - \log M_{Y_1}(a_{12}\sqrt{\psi})} F_{Y_1}(\mathrm{d}y) \\
&> 0,
\end{aligned}
$$

where the last inequality holds since $\sqrt{\psi}a_{22}y - \log M_{Y_1}(a_{22}\sqrt{\psi}) > \sqrt{\psi}a_{12}y - \log M_{Y_1}(a_{12}\sqrt{\psi})$ for $y > c/(a_{22} - a_{12})$. $\qquad\square$

Proposition E.2. *Let* $f(x) = x^\nu K_\nu(x), \nu > 0, x > 0$. *Then* $f(x)$ *is strictly decreasing on* $(0, \infty)$.

Proof. Using the formula $K'_\nu(x) = -\{K_{\nu-1}(x) + K_{\nu+1}(x)\}/2$ (cf. Abramowitz & Stegun (1972), formula 9.6.26) we have

$$
\begin{aligned}
f'(x) &= \nu x^{\nu-1} K_\nu(x) + x^\nu \left\{ -\frac{K_{\nu-1}(x) + K_{\nu+1}(x)}{2} \right\} \\
&= x^{\nu-1} \frac{2\nu K_\nu(x) - x K_{\nu-1}(x) - x K_{\nu+1}(x)}{2}.
\end{aligned}
$$

Since $x K_{\nu+1}(x) = x K_{\nu-1}(x) + 2\nu K_\nu(x)$ (cf. Abramowitz & Stegun (1972), formula 9.6.26), we have

$$
f'(x) = x^{\nu-1} \frac{2\nu K_\nu(x) - x K_{\nu-1}(x) - x K_{\nu-1}(x) - 2\nu K_\nu(x)}{2} = -x^\nu K_{\nu-1}(x) < 0.
$$

Therefore, $f(x)$ is strictly decreasing on $(0, \infty)$. $\qquad\square$

We now give more details of the GH distribution. The Lebesgue density function of $\text{GH}(\lambda, \tau, \psi, \mu, \gamma)$ is

$$f_{\text{GH}}(x) = \frac{\psi^{\lambda/2}(\psi + \gamma^2)^{1/2-\lambda}}{(2\pi)^{1/2}\tau^{\lambda/2}K_\lambda(\sqrt{\tau\psi})} \cdot \frac{K_{\lambda-1/2}\left[\sqrt{\{\tau + (x-\mu)^2\}(\psi + \gamma^2)}\right]e^{\gamma(x-\mu)}}{\left[\sqrt{\{\tau + (x-\mu)^2\}(\psi + \gamma^2)}\right]^{1/2-\lambda}}, \quad x \in \mathbb{R}.$$

To see why the GH density function also has exponential tails, using the asymptotic expansion of the modified Bessel function of the second kind for large arguments (Abramowitz & Stegun, 1972, Formula 9.7.2), we obtain

$$f_{\text{GH}}(x) \sim c_0 x^{\lambda-1} e^{-(\sqrt{\psi+\gamma^2}-\gamma)x}, \quad x \to \infty,$$

where $c_0 > 0$ is a constant, and $f(x) \sim g(x)$ as $x \to \infty$ means that g is eventually non-zero and $f(x)/g(x) \to 1$ as $x \to \infty$. A simple application of L'Hôpital's rule gives that the GH distribution functions also have exponential tails.

References

Abramowitz, M., & Stegun, I. A. (Eds.) (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. National Bureau of Standards, United States of America, 10 ed.

Asar, Ö., Bolin, D., Diggle, P. J., & Wallin, J. (2020). Linear mixed effects models for non-Gaussian continuous repeated measurement data. *Journal of the Royal Statistical Society (Series C)*, *69*(5), 1015–1065.

Barndorff-Nielsen, O. E. (1977). Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society of London, Series A*, *353*, 401–419.

Barndorff-Nielsen, O. E. (1998). Processes of normal inverse Gaussian type. *Finance and Stochastics*, *2*, 41–68.

Barndorff-Nielsen, O. E., & Shephard, N. (2001). Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society (Series B)*, *63*(2), 167–241.

Bolin, D. (2014). Spatial Matérn fields driven by non-Gaussian noise. *Scandinavian Journal of Statistics*, *41*, 557–579.

Bolin, D., & Wallin, J. (2020). Multivariate type G Matérn stochastic partial differential equation random fields. *Journal of the Royal Statistical Society (Series B)*, *82*(1), 215–239.

Boyd, S., & Vandenberghe, L. (2004). *Convex Optimization*. Cambrige University Press.

Breiman, L. (1965). On some limit theorems similar to the arc-sine law. *Theory of Probability and its Applications*, *10*(2), 323–331.

Buldygin, V. V., Indlekofer, K., Klesov, O. I., & Steinebach, J. G. (2018). *Pseudo-Regularly Varying Functions and Generalized Renewal Processes*. Springer Cham.

Cline, D. B. H. (1986). Convolution tails, product tails and domains of attraction. *Probability Theory and Related Fields*, *72*, 529–557.

Coles, S., Heffernan, J., & Tawn, J. A. (1999). Dependence measures for extreme value analyses. *Extremes*, *2*(4), 339–365.

Cressie, N., & Pavlicová, M. (2002). Calibrated spatial moving average simulations. *Statistical Modelling*, *2*, 267–279.

Eberlein, E. (2001). Application of generalized hyperbolic Lévy motions to finance. In O. E. Barndorff-Nielsen, S. I. Resnick, & T. Mikosch (Eds.) *Lévy Processes – Theory and Applications*, (pp. 319–336). Birkhäuser, Boston, MA.

Embrechts, P., & Goldie, C. M. (1980). On closure and factorization properties of subexponential and related distributions. *Journal of the Australian Mathematical Society (Series A)*, *29*, 243–256.

Embrechts, P., & Goldie, C. M. (1982). On convolution tails. *Stochastic Processes and their Applications*, *13*, 263–278.

Engelke, S., Opitz, T., & Wadsworth, J. L. (2019). Extremal dependence of random scale constructions. *Extremes*, *22*, 623–666.

Gnecco, N., Meinshausen, N., Peters, J., & Engelke, S. (2021). Causal discovery in heavy-tailed models. *The Annals of Statistics*, *49*(3), 1755–1778.

Higdon, D. (2002). Space and space-time modeling using process convolutions. In C. W. Anderson, V. Barnett, P. C. Chatwin, & A. H. El-Shaarawi (Eds.) *Quantitative Methods for Current Environmental Issues*, (pp. 37–56). Springer.

Huser, R., & Wadsworth, J. L. (2019). Modeling spatial processes with unknown extremal dependence class. *Journal of the American Statistical Association*, *114*(525), 434–444.

Jurek, Z. J., & Vervaat, W. (1983). An integral representation for selfdecomposable Banach space valued random variables. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, *62*, 247–262.

Kallenberg, O. (2017). *Random Measures, Theory and Applications*. Springe.

Krupskii, P., & Huser, R. (2022). Modeling spatial tail dependence with Cauchy convolution processes. *Electronic Journal of Statistics*, *16*, 6135–6174.

Kuusela, M., & Stein, M. L. (2018). Locally stationary spatio-temporal interpolation of Argo profiling float data. *Proceedings of the Royal Society A*, *474*, 20180400.

Ledford, A. W., & Tawn, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika*, *83*(1), 169–187.

Lindgren, F., Rue, H., & Lindström, J. (2011). An explicit link between the Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society (Series B)*, *73*, 423–498.

McNeil, A. J., Frey, R., & Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press.

Nolde, N. (2014). Geometric interpretation of the residual dependence coefficient. *Journal of Multivariate Analysis*, *123*, 85–95.

Nolde, N., & Wadsworth, J. L. (2022). Linking representations for multivariate extremes via a limit set. *Advances in Applied Probability*, *54*(3), 688–717.

Nolde, N., & Zhou, C. (2021). Extreme value analysis for financial risk management. *Annual Review of Statistics and Its Application*, *8*, 217–240.

Opitz, T. (2018). Spatial random field models based on Lévy indicator convolutions. ArXiv: 1710.06826.

Pakes, A. G. (2004). Convolution equivalence and infinity divisibility. *Journal of Applied Probability*, *41*(2), 407–424.

Prause, K. (1999). *The Generalized Hyperbolic Model: Estimation, Financial Derivatives, and Risk Measures*. Ph.D. thesis, Albert-Ludwigs-Universität Freiburg.

Rajput, B. S., & Rosinski, J. (1989). Spectral representations of infinitely divisible processes. *Probability Theory and Related Fields*, *82*, 451–487.

Resnick, S. I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer.

Rodrigues, A., & Diggle, P. J. (2010). A class of convolution-based models for spatio-temporal processes with non-seperable covariance structure. *Scandinavian Journal of Statistics*, *37*, 553–567.

Rosinski, J. (1991). On a class of infinitely divisible processes represented as mixtures of Gaussian processes. In S. Cambanis, G. Samorodnitsky, & M. S. Taqqu (Eds.) *Stable Processes and Related Topics: A Selection of Papers from the Mathematical Sciences Institute Workshop, January 9–13, 1990*, (pp. 27–41). Boston, MA: Birkhäuser Boston.

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

Sgibnev, M. S. (1990). Asymptotics of infinitely divisible distributions on $\mathbb{R}$. *Siberian Mathematical Journal*, *31*, 115–119.

Shimura, T., & Watanabe, T. (2005). Infinite divisibility and generalized subexponentiality. *Bernoulli*, *11*(3), 445–469.

Tao, T. (2016). *Analysis I*. Springer, 3 ed.

Ver Hoef, J. M., & Peterson, E. E. (2010). A moving average approach for spatial statistical models for stream networks. *Journal of the American Statistical Association*, *105*(489), 6–18.

Wallin, J., & Bolin, D. (2015). Geostatistical modelling using non-Gaussian Matérn fields. *Scandinavian Journal of Statistics*, *42*, 872–890.

Wang, Y., & Stoev, S. A. (2011). Conditional sampling for spectrally discrete max-stable random fields. *Advances in Applied Probability*, *43*, 461–483.