Scandinavian Journal of Statistics Theory and Applications

Scandinavian Journal of Statistics, Vol. 44: 21–45, 2017
doi: 10.1111/sjos.12240
© 2016 Board of the Foundation of the Scandinavian Journal of Statistics. Published by Wiley Publishing Ltd.

Models for Extremal Dependence Derived from Skew-symmetric Families

BORIS BERANGER
School of Mathematics and Statistics, University of New South Wales
SIMONE A. PADOAN
Department of Decision Sciences, Bocconi University
SCOTT A. SISSON
School of Mathematics and Statistics, University of New South Wales

ABSTRACT. Skew-symmetric families of distributions such as the skew-normal and skew-*t* represent supersets of the normal and *t* distributions, and they exhibit richer classes of extremal behaviour. By defining a non-stationary skew-normal process, which allows the easy handling of positive definite, non-stationary covariance functions, we derive a new family of max-stable processes – the extremal skew-*t* process. This process is a superset of non-stationary processes that include the stationary extremal-*t* processes. We provide the spectral representation and the resulting angular densities of the extremal skew-*t* process and illustrate its practical implementation.

Key words: angular density, asymptotic independence, extremal coefficient, extreme values, max-stable distribution, non-central extended skew-*t* distribution, non-stationarity, skew-normal distribution, skew-normal process, skew-*t* distribution

1. Introduction

The modern day analysis of extremes is based on results from the theory of stochastic processes. In particular, max-stable processes (de Haan, 1984) are a popular and useful tool when modelling extremal responses in environmental, financial and engineering applications. Let $\mathbb{S} \subseteq \mathbb{R}^k$ denote a k-dimensional region of space (or space-time) over which a real-valued stochastic process $\{Y(s)\}_{s\in\mathbb{S}}$ with a continuous sample path on $\mathbb{S}$ can be defined. Considering a sequence $Y_1, \ldots, Y_n$ of independent and identically distributed (iid) copies of Y, the pointwise partial maximum can be defined as

$$M_n(s) = \max_{i=1,\ldots,n} Y_i(s), \quad s \in \mathbb{S}.$$

If there are sequences of real-valued functions, $a_n(s) > 0$ and $b_n(s)$, for $s \in \mathbb{S}$ and $n = 1, 2, \ldots$, such that

$$\left\{ \frac{M_n(s) - b_n(s)}{a_n(s)} \right\}_{s\in\mathbb{S}} \Rightarrow \{U(s)\}_{s\in\mathbb{S}},$$

converges weakly as $n \to \infty$ to a process $U(s)$ with non-degenerate marginal distributions for all $s \in \mathbb{S}$, then $U(s)$ is known as a max-stable process (de Haan & Ferreira, 2006, Ch. 9). In this setting, for a finite sequence of points $(s_j)_{j\in I}$ in $\mathbb{S}$, where $I = \{1, \ldots, d\}$ is an index set, the finite-dimensional distribution of U is then a multivariate extreme value distribution (de Haan & Ferreira, 2006, Ch. 6). This distribution has generalized extreme value univariate margins and, when parameterized with unit Fréchet margins, has a joint distribution function of the form

$$G(x_j, j \in I) = \exp\{-V(x_j, j \in I)\}, \quad x_j > 0,$$

where $x_j \equiv x(s_j)$. The exponent function V describes the dependence between extremes and can be expressed as

$$V(x_j, j \in I) = \int_{\mathbb{W}} \max_{j \in I}(w_j/x_j) H(\mathrm{d}w_1, \ldots, \mathrm{d}w_d),$$

where the angular measure H is a finite measure defined on the d-dimensional unit simplex $\mathbb{W} = \{w \in \mathbb{R}^d : w_1 + \cdots + w_d = 1\}$, satisfying the moment conditions $\int_{\mathbb{W}} w_j \; H(\mathrm{d}w) = 1, j \in I$, (de Haan & Ferreira, 2006, Ch. 6).

In recent years, a variety of specific max-stable processes have been developed, many of which have become popular as they can be practically amenable to statistical modelling (Davison *et al.*, 2012). The extremal-*t* process (Opitz, 2013) is one of the best-known and widely used max-stable processes, from which the Brown–Resnick process (Brown & Resnick, 1977; Kabluchko *et al.*, 2009), the Gaussian extreme-value process (Smith, 1990) and the extremal-Gaussian processes (Schlather, 2002) can be seen as special cases. In their most basic form, the Brown–Resnick and the extremal-*t* processes can be respectively understood as the limiting extremal processes of strictly stationary Gaussian and Student-*t* processes. However, in practice, data may be non-stationary and exhibit asymmetric distributions in many applications. In these scenarios, skew-symmetric distributions (Azzalini & Captianio A., 2014; Arellano-Valle & Azzalini, 2006; Azzalini, 2005; Genton, 2004; Azzalini, 1985) provide simple models for modelling asymmetrically distributed data. However, the limiting extremal behaviour of these processes has not yet been established.

In this paper, we characterize and develop statistical models for the extremal behaviour of skew-normal and skew-*t* distributions. The joint tail behaviours of these skew distributions are capable of describing a far wider range of dependence levels than that obtained under the symmetric normal and *t* distributions. We provide a definition of a skew-normal process, which is in turn a non-stationary process. This provides an accessible approach to constructing positive definite, non-stationary covariance functions when working with non-Gaussian processes. Recently, some forms of non-stationary dependent structures embedded into max-stable processes have been studied by Huser & Genton (2015). We show that on the basis of the skew-normal process, a new family of max-stable processes – the extremal skew-*t* process – can be obtained. This process is a superset of non-stationary processes that include the stationary extremal-*t* processes (Opitz, 2013). From the extremal skew-*t* process, a rich family of non-stationary, isotropic or anisotropic extremal coefficient functions can be obtained.

This paper is organized as follows: in Section 2, we first introduce a new variant of the extended skew-*t* class of distributions, before developing a non-stationary version of the skew-normal process. In both cases, we discuss the stochastic behaviour of their extreme values. In Section 3, we derive the spectral representation of the extended extremal skew-*t* process. Section 4 discusses inferential aspects of the extremal skew-*t* dependence model, and Section 5 provides a real data application. We conclude with a Discussion.

2. Preliminary results on skew-normal processes and skew-*t* distributions

We introduce two preliminary results that will be used in order to present our main contribution in Section 3: the extremal skew-*t* process. In Section 2.1, we define the *non-central* extended skew-*t* family of distributions, which is a new variant of the class introduced by Arellano–Valle & Genton (2010) that allows a non-centrality parameter. In Section 2.2, we present the development of a new non-stationary, skew-normal random process.

Hereafter, we use $Y \sim \mathcal{D}_d(\theta_1, \theta_2, \ldots)$ to denote that Y is a d-dimensional random vector with probability law $\mathcal{D}$ and parameters $\theta_1, \theta_2, \ldots$. When $d = 1$, the subscript is omitted for

© 2016 Board of the Foundation of the Scandinavian Journal of Statistics.

14679469, 2017, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.12240 by University of New South Wales, Wiley Online Library on [04/08/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

brevity. Similarly, when a parameter is equal to zero or a scale matrix is equal to the identity (both in a vector and scalar sense) so that $\mathcal{D}_d$ reduces to an obvious sub-family, it is also omitted.

2.1. The non-central, extended skew-t distribution

While several skew-symmetric distributions have been developed (e.g. Genton, 2004; Azzalini, 2014), we focus on the skew-normal and skew-*t* distributions.

Denote a d-dimensional skew-normally distributed random vector by $Y \sim \mathcal{SN}_d(\mu, \Omega, \alpha, \tau)$ (Arellano-Valle & Genton, 2010). This random vector has probability density function (pdf)

$$\phi_d(y; \mu, \Omega, \alpha, \tau) = \frac{\phi_d(y; \mu, \Omega)}{\Phi\{\tau/\sqrt{1 + Q_{\bar{\Omega}}(\alpha)}\}} \Phi(\alpha^\top z + \tau), \qquad y \in \mathbb{R}^d, \tag{1}$$

where $\phi_d(y; \mu, \Omega)$ is a d-dimensional normal pdf with mean $\mu \in \mathbb{R}^d$ and $d \times d$ covariance matrix Ω, $z = (y - \mu)/\omega$, $\omega = \text{diag}(\Omega)^{1/2}$, $\bar{\Omega} = \omega^{-1}\Omega\,\omega^{-1}$, $Q_{\bar{\Omega}}(\alpha) = \alpha^\top \bar{\Omega}\alpha$ and $\Phi(\cdot)$ is the standard univariate normal cumulative distribution function (cdf). The shape parameters $\alpha \in \mathbb{R}^d$ and $\tau \in \mathbb{R}$ are, respectively, *slant* and *extension* parameters. The cdf associated with (1) is termed the extended skew-normal distribution (Arellano-Valle & Genton, 2010) of which the skew-normal and normal distributions are special cases (Arellano-Valle & Genton, 2010; Azzalini & Captianio A., 2014). For example, in the case where $\alpha = 0$ and $\tau = 0$, the standard normal pdf is recovered.

Definition 1. Y is a d-dimensional, non-central extended skew-*t* distributed random vector, denoted by $Y \sim \mathcal{ST}_d(\mu, \Omega, \alpha, \tau, \kappa, \nu)$, if for $y \in \mathbb{R}^d$ it has pdf

$$\psi_d(y; \mu, \Omega, \alpha, \tau, \kappa, \nu) = \frac{\psi_d(y; \mu, \Omega, \nu)}{\Psi\left(\frac{\tau}{\sqrt{1+Q_{\bar{\Omega}}(\alpha)}}; \frac{\kappa}{\sqrt{1+Q_{\bar{\Omega}}(\alpha)}}, \nu\right)} \Psi\left\{(\alpha^\top z + \tau)\sqrt{\frac{\nu + d}{\nu + Q_{\bar{\Omega}^{-1}}(z)}}; \kappa, \nu + d\right\}, \tag{2}$$

where $\psi_d(y; \mu, \Omega, \nu)$ is the pdf of a d-dimensional t-distribution with location $\mu \in \mathbb{R}^d$, $d \times d$ scale matrix Ω and $\nu \in \mathbb{R}^+$ degrees of freedom, $\Psi(\cdot; a, \nu)$ denotes a univariate non-central t cdf with non-centrality parameter $a \in \mathbb{R}$ and ν degrees of freedom and $Q_{\bar{\Omega}^{-1}}(z) = z^\top \bar{\Omega}^{-1} z$. The remaining terms are as defined in (1). The associated cdf is

$$\Psi_d(y; \mu, \Omega, \alpha, \tau, \kappa, \nu) = \frac{\Psi_{d+1}\{\bar{z}; \Omega^*, \kappa^*, \nu\}}{\Psi(\bar{\tau}; \bar{\kappa}, \nu)}, \tag{3}$$

where $\bar{z} = (z^\top, \bar{\tau})^\top$, Ψ_{d+1} is a $(d+1)$-dimensional (non-central) t cdf with covariance matrix and non-centrality parameters

$$\Omega^* = \begin{pmatrix} \bar{\Omega} & -\delta \\ -\delta^\top & 1 \end{pmatrix}, \quad \kappa^* = \begin{pmatrix} 0 \\ \bar{\kappa} \end{pmatrix},$$

and ν degrees of freedom, and where

$$\delta = \left\{1 + Q_{\bar{\Omega}}(\alpha)\right\}^{-1/2} \bar{\Omega}\alpha, \quad \bar{\kappa} = \left\{1 + Q_{\bar{\Omega}}(\alpha)\right\}^{-1/2} \kappa, \quad \bar{\tau} = \left\{1 + Q_{\bar{\Omega}}(\alpha)\right\}^{-1/2} \tau. \tag{4}$$

When the non-centrality parameter κ is zero, then the extended skew-*t* family of Arellano–Valle & Genton (2010) is obtained. For the non-central skew-*t* family, we now demonstrate modified properties to those discussed in Arellano-Valle & Genton (2010).

© 2016 Board of the Foundation of the Scandinavian Journal of Statistics.

14679469, 2017, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.12240 by University of New South Wales, Wiley Online Library on [04/08/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Proposition 1 (Properties). *Let* $Y \sim \mathcal{ST}_d(\mu, \Omega, \alpha, \tau, \kappa, \nu)$.

(1) Marginal and conditional distributions. Let $I \subset \{1, \ldots, d\}$ *and* $\bar{I} = \{1, \ldots, d\} \backslash I$ *identify the* d_I*-dimensional and* $d_{\bar{I}}$*-dimensional subvector partition of* Y *such that* $Y = \left(Y_I^\top, Y_{\bar{I}}^\top\right)^\top$, *with corresponding partitions of the parameters* (μ, Ω, α). *Then*

(a) $Y_I \sim \mathcal{ST}_{d_I}\left(\mu_I, \Omega_{II}, \alpha_I^*, \tau_I^*, \kappa_I^*, \nu\right)$, *where*

$$\alpha_I^* = \frac{\alpha_I + \bar{\Omega}_{II}^{-1}\bar{\Omega}_{I\bar{I}}\alpha_{\bar{I}}}{\sqrt{1 + Q_{\tilde{\Omega}_{\bar{I}\bar{I}\cdot I}}(\alpha_{\bar{I}})}}, \quad \tau_I^* = \frac{\tau}{\sqrt{1 + Q_{\tilde{\Omega}_{\bar{I}\bar{I}\cdot I}}(\alpha_{\bar{I}})}}, \quad \kappa_I^* = \frac{\kappa}{\sqrt{1 + Q_{\tilde{\Omega}_{\bar{I}\bar{I}\cdot I}}(\alpha_{\bar{I}})}}, \tag{5}$$

given $\tilde{\Omega}_{\bar{I}\bar{I}\cdot I} = \bar{\Omega}_{\bar{I}\bar{I}} - \bar{\Omega}_{\bar{I}I}\bar{\Omega}_{II}^{-1}\bar{\Omega}_{I\bar{I}}$.

(b) $(Y_{\bar{I}} \mid Y_I = y_I) \sim \mathcal{ST}_{d_{\bar{I}}}(\mu_{\bar{I}\cdot I}, \Omega_{\bar{I}\cdot I}, \alpha_{\bar{I}\cdot I}, \tau_{\bar{I}\cdot I}, \kappa_{\bar{I}\cdot I}, \nu_{\bar{I}\cdot I})$, *where* $\mu_{\bar{I}\cdot I} = \mu_{\bar{I}} + \Omega_{\bar{I}I}\Omega_{II}^{-1}(y_I - \mu_I)$, $\Omega_{\bar{I}\cdot I} = \zeta_I \Omega_{\bar{I}\bar{I}\cdot I}$, $\zeta_I = \{\nu + Q_{\Omega_{II}^{-1}}(z_I)\}/(\nu + d_I)$, $z_I = \omega_I^{-1}(y_I - \mu_I)$, $\omega_I = \mathrm{diag}(\omega_{II})^{1/2}$, $Q_{\Omega_{II}^{-1}}(z_I) = z_I^\top \Omega_{II}^{-1} z_I$, $\Omega_{\bar{I}\bar{I}\cdot I} = \Omega_{\bar{I}\bar{I}} - \Omega_{\bar{I}I}\Omega_{II}^{-1}\Omega_{I\bar{I}}$, $\alpha_{\bar{I}\cdot I} = \omega_{\bar{I}\cdot I}\omega_{\bar{I}}^{-1}\alpha_{\bar{I}}$, $\omega_{\bar{I}\cdot I} = \mathrm{diag}(\Omega_{\bar{I}\bar{I}\cdot I})^{1/2}$, $\omega_{\bar{I}} = \mathrm{diag}(\omega_{\bar{I}\bar{I}})^{1/2}$, $\tau_{\bar{I}\cdot I} = \zeta_I^{-1/2}\left\{\left(\alpha_{\bar{I}}^\top \bar{\Omega}_{\bar{I}I}\bar{\Omega}_{II}^{-1} + \alpha_I^\top\right) z_I + \tau\right\}$, $\kappa_{\bar{I}\cdot I} = \zeta_I^{-1/2}\kappa$ *and* $\nu_{\bar{I}\cdot I} = \nu + d_I$.

(2) Conditioning-type stochastic representation. We can write $Y = \mu + \Omega Z$, *where* $Z = (X \mid \alpha^\top X + \tau > X_0)$ *and where* $X \sim \mathcal{T}_d(\bar{\Omega}, \nu)$ *is independent of* $X_0 \sim \mathcal{T}(\kappa, \nu)$.

(3) Additive-type stochastic representation. We can write $Y = \mu + \Omega Z$, *where* $Z = \sqrt{\frac{\nu + \tilde{X}_0^2}{\nu+1}} X_1 + \delta \tilde{X}_0$, $X_1 \sim \mathcal{T}_d(\Omega - \delta\delta^\top, \bar{\kappa}, \nu + 1)$ *is independent of* $\tilde{X}_0 = (X_0 \mid X_0 + \bar{\tau} > 0)$, $X_0 \sim \mathcal{T}(\bar{\kappa}, \nu)$, $\delta \in (-1, 1)^d$ *and where* $\bar{\tau}$ *and* $\bar{\kappa}$ *are as in (4).*

Proof in Appendix A.1

We conclude by presenting a final property of the non-central skew-*t* family. The next result describes the extremal behaviour of observations drawn from a member of this class.

Proposition 2. *Let* $Z_1, \ldots, Z_n$ *be iid copies of* $Z \sim \mathcal{ST}_d(\bar{\Omega}, \alpha, \tau, \kappa, \nu)$ *and* M_n *be the componentwise sample maxima. Define* $a_n = (a_{n,1}, \ldots, a_{n,d})^\top$, *where*

$$a_{n,j} = \left\{ \frac{n\{(\nu/2)\}^{-1}\{(\nu+1)/2\}\nu^{(\nu-2)/2}\,\Psi\left(\alpha_j^*\sqrt{\nu+1}; \kappa, \nu+1\right)}{\sqrt{\pi}\Psi\left(\tau_j^*/\left\{1 + Q_{\bar{\Omega}}\left(\alpha_j^*\right)\right\}^{1/2}; \kappa_j^*/\left\{1 + Q_{\bar{\Omega}}\left(\alpha_j^*\right)\right\}, \nu\right)} \right\}^{1/\nu}$$

where $\alpha_j^* = \alpha_{\{j\}}^*$, $\tau_j^* = \tau_{\{j\}}^*$ *and* $\kappa_j^* = \kappa_{\{j\}}^*$ *are the marginal parameters (5) under Proposition 1(1). Then* $M_n/a_n \Rightarrow U$ *as* $n \to +\infty$, *where* U *has univariate* ν*-Fréchet marginal distributions (i.e.* $e^{-x^{-\nu}}$, $x > 0$*) and exponent function*

$$V(x_j, j \in I) = \sum_{j=1}^{d} x_j^{-\nu}\Psi_{d-1}\left(\left(\sqrt{\frac{\nu+1}{1-\omega_{i,j}^2}}\left(\frac{x_i^+}{x_j^+} - \omega_{i,j}\right), i \in I_j\right)^\top; \bar{\Omega}_j^+, \alpha_j^+, \tau_j^+, \nu+1\right), \tag{6}$$

where Ψ_{d-1} *is a* $(d-1)$*-dimensional central extended skew-t distribution with correlation matrix,* $\bar{\Omega}_j^+$, *shape and extension parameters ,*α_j^+ *and* τ_j^+, *and* $\nu + 1$ *degrees of freedom,* $I = \{1, \ldots, d\}$, $I_j = I \backslash \{j\}$ *and* $\omega_{i,j}$ *is the* (i, j)*-th element of* $\bar{\Omega}$.

© 2016 Board of the Foundation of the Scandinavian Journal of Statistics.

14679469, 2017, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.12240 by University of New South Wales, Wiley Online Library on [04/08/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Proof (and further details) in Appendix A.2.

As the limiting distribution (6) is the same as that of the classic skew-t distribution (Padoan, 2011), it exhibits identical upper and lower tail dependence coefficients (e.g. Joe, 1997, Ch. 5). That is, the extension and non-centrality parameters, τ and κ, do not affect the extremal behaviour.

2.2. *A non-stationary, skew-normal random process*

While there are several definitions of a stationary skew-normal process (e.g. Minozzo & Ferracuti, 2012), stationarity is incompatible with the requirement that all finite-dimensional distributions of the process are skew-normal. We now construct a non-stationary version of the skew-normal process through the additive-type stochastic representation (e.g. Azzalini, 2014, Ch. 5). A similar approach was explored by Zhang & El-Shaarawi (2010) for the stationary case.

Definition 2. Let $\{X(s)\}_{s\in\mathbb{S}}$ be a stationary Gaussian random process on $\mathbb{S}$ with zero-mean, unit variance and correlation function $\rho(h) = \mathbb{E}\{X(s)X(s+h)\}$ for $s \in \mathbb{S}$ and $h \in \mathbb{R}^k$. For $X' \sim \mathcal{N}(0,1)$ independent of $X(s)$, $\varepsilon \in \mathbb{R}$ and a function $\delta : \mathbb{S} \mapsto (-1,1)$, define

$$\begin{aligned} X''(s) &:= X' \mid X' + \varepsilon > 0, \qquad \forall\, s \in \mathbb{S} \\ Z(s) &:= \sqrt{1-\delta(s)^2}\,X(s) + \delta(s)X''(s), \quad s \in \mathbb{S}. \end{aligned} \tag{7}$$

Then $Z(s)$ is a skew-normal random process.

We refer to $\delta(s)$ as the slant function. From (7), if $\delta(s) \equiv 0$ for all $s \in \mathbb{S}$, then Z is a Gaussian random process. Note that Z is a random process with a consistent family of distribution functions, because $Z(s) = a(s)X(s) + b(s)Y(s)$, where a and b are bounded functions and X and Y are random processes with a consistent family of distribution functions. For any finite sequence of points $s_1, \ldots, s_d \in \mathbb{S}$, the joint distribution of $Z(s_1), \ldots, Z(s_d)$ is $\mathcal{SN}_d(\bar{\Omega}, \alpha, \tau)$, where

$$\begin{aligned} \bar{\Omega} &= D_\delta(\bar{\Sigma} + (D_\delta^{-1}\delta)(D_\delta^{-1}\delta)^\top)D_\delta \\ \alpha &= \{1 + (D_\delta^{-1}\delta)^\top \bar{\Sigma}^{-1}(D_\delta^{-1}\delta)\}^{-1/2} D_\delta^{-1}\bar{\Sigma}^{-1}(D_\delta^{-1}\delta) \\ \tau &= \{1 + Q_{\bar{\Omega}}(\alpha)\}^{1/2}\,\varepsilon \end{aligned} \tag{8}$$

and where $\bar{\Sigma}$ is the $d \times d$ correlation matrix of X, $\delta = (\delta(s_1), \ldots, \delta(s_d))^\top$ and $D_\delta = \{1_d - \mathrm{diag}(\delta^2)\}^{1/2}$, where 1_d is the identity matrix (Azzalini & Captianio A., 2014, Ch. 5). As a result, for any lag $h \in \mathbb{R}^k$, the distributions of $\{Z(s_1), \ldots, Z(s_d)\}$ and $\{Z(s_1+h), \ldots, Z(s_d+h)\}$ will differ unless $\delta(s) = 0$ for all $s \in \mathbb{S}$. Hence, the distribution of Z is not translation invariant, and the process is not strictly stationary. For $s \in \mathbb{S}$ and $h \in \mathbb{R}^k$, the mean $m(s)$ and covariance function $c_s(h)$ of the skew-normal random process are

$$m(s) = \mathbb{E}\{Z(s)\} = \delta(s)\phi(\varepsilon)/\Phi(\varepsilon)$$

and

$$c_s(h) = \mathrm{Cov}\{Z(s), Z(s+h)\} = \rho(h)\sqrt{\{1-\delta^2(s)\}\{1-\delta^2(s+h)\}} + \delta(s)\delta(s+h)(1-r), \tag{9}$$

where $r = \left\{\frac{\phi(\varepsilon)}{\Phi(\varepsilon)}\left(\varepsilon + \frac{\phi(\varepsilon)}{\Phi(\varepsilon)}\right)\right\}$. Hence, the mean is not constant, and the covariance does not depend only on the lag h, unless $\delta(s) = \delta_0 \in (-1,1)$ for all $s \in \mathbb{S}$. In the latter case, the skew-normal random process is weakly stationary (Zhang & El-Shaarawi, 2010).

© 2016 Board of the Foundation of the Scandinavian Journal of Statistics.

14679469, 2017, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.12240 by University of New South Wales, Wiley Online Library on [04/08/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

One benefit of working with a skew-normal random field is that the non-stationary covariance function (9) is positive definite if the covariance function of X is positive definite and if $-1 < \delta(s) < 1$ for all $s \in \mathbb{S}$. Hence, a valid model is directly obtainable by means of standard parametric correlation models $\rho(h)$ and any bounded function δ in $(-1, 1)$. If the Gaussian process correlation function satisfies $\rho(0) = 1$ and $\rho(h) \to 0$ as $\|h\| \to +\infty$, then the correlation of the skew-normal process satisfies $\rho_S(0) = 1$ and

$$\rho_S(h) = \frac{c_S(h)}{\sqrt{c_S(0)c_S(h)}} \approx \frac{\delta(s)\delta(s+h)(1-r)}{\sqrt{(1-\delta^2(s)r)(1-\delta^2(s+h)r)}},$$

as $\|h\| \to +\infty$. Hence, $\rho_S(h) = 0$ if either $\delta(s)$ or $\delta(s+h)$ is zero. Conversely, if both $\delta(s) \to \pm 1$ and $\delta(s+h) \to \pm 1$, then $\rho_S(h) \to \pm 1$.

The increments $Z(s+h) - Z(s)$ are skew-normal distributed for any fixed $s \in \mathbb{S}$ and $h \in \mathbb{R}^k$ (Azzalini & Captianio A., 2014, Ch. 5), and the variogram $2\gamma_S(h) = \mathrm{Var}\{Z(s+h) - Z(s)\}$ is equal to

$$2\gamma_S(h) = 2\left(1 - c_S(h) - \frac{\delta^2(s+h) + \delta^2(s)}{2/r}\right).$$

When $h = 0$, the variogram is zero, and when $\|h\| \to +\infty$, the variogram approaches a constant ≤ 2, respectively resulting in spatial independence or dependence for large distances h. We can now infer the conditions required so that $Z(s)$ has a continuous sample path.

Proposition 3. *Assume that $\mathbb{S} \subseteq \mathbb{R}$. A skew-normal process $\{Z(s), s \in \mathbb{S}\}$ has a continuous sample path if $\delta(s+h) - \delta(s) = o(1)$ and $1 - \rho(h) = O(|\log|h||^{-a})$ for some $a > 3$, as $h \to 0$.*

This result follows by noting that $r_S(h) = \rho(h) + \delta^2(s)(1-\rho(h)) + o(1)$ as $h \to 0$, and this is a consequence of the continuity assumption on $\delta(s)$, where $r_S(h) = c_S(h) + r\{\delta^2(s+h) + \delta^2(s)\}/2$. Therefore, $1 - r_S(h) = O(|\log|h||^{-a})$ as $h \to 0$. Thus, the proof follows from the results in Lindgren (2012, page 48). This means that continuity of the skew-normal process is assured if $\delta(s)$ is a continuous function, in addition to the usual condition on the correlation function of the generating Gaussian process (e.g. Lindgren, 2012, Ch. 2).

Figure 1 illustrates trajectories of the skew-normal process for $k = 1$, with $X(s)$ a zero-mean, unit variance Gaussian process on [0, 1] with isotropic power-exponential correlation function

$$\rho(h; \vartheta) = \exp\{-(h/\lambda)^{\xi}\}, \quad \vartheta = (\lambda, \xi),\ \lambda > 0,\ 0 < \xi \leq 2,\ h > 0, \tag{10}$$

with $\xi = 1.5$, $\lambda = 0.3$ and $h \in [0, 1]$.

The first row shows the standard stationary case. The second row illustrates the non-stationary correlation function obtained with $s = 0.1$ (solid line) behaving close to the stationary correlation, however, decaying more slowly as s increases and approaching, but not reaching zero exactly. The third row demonstrates that both points may be negatively correlated and that $\rho_S(h)$ is not necessarily a decreasing function in h. The bottom row highlights this even more clearly – correlation functions need not be monotonically decreasing – implying that pairs of points far apart can be more dependent than nearby points.

Simulating a skew-normal random process is computationally cheap through Definition 2, with the simulation of the required stationary Gaussian process achievable through many fast algorithms (e.g. Wood & Chan, 1994; Chan & Wood, 1997). Rather than relying on (8), for practical purposes, to directly simulate from a skew-normal process with given parameters α, $\bar{\Omega}$ and τ, a conditioning sampling approach can be adopted (Azzalini & Captianio A., 2014, Ch. 5).

© 2016 Board of the Foundation of the Scandinavian Journal of Statistics.

14679469, 2017, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.12240 by University of New South Wales, Wiley Online Library on [04/08/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14679469, 2017, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.12240 by University of New South Wales, Wiley Online Library on [04/08/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

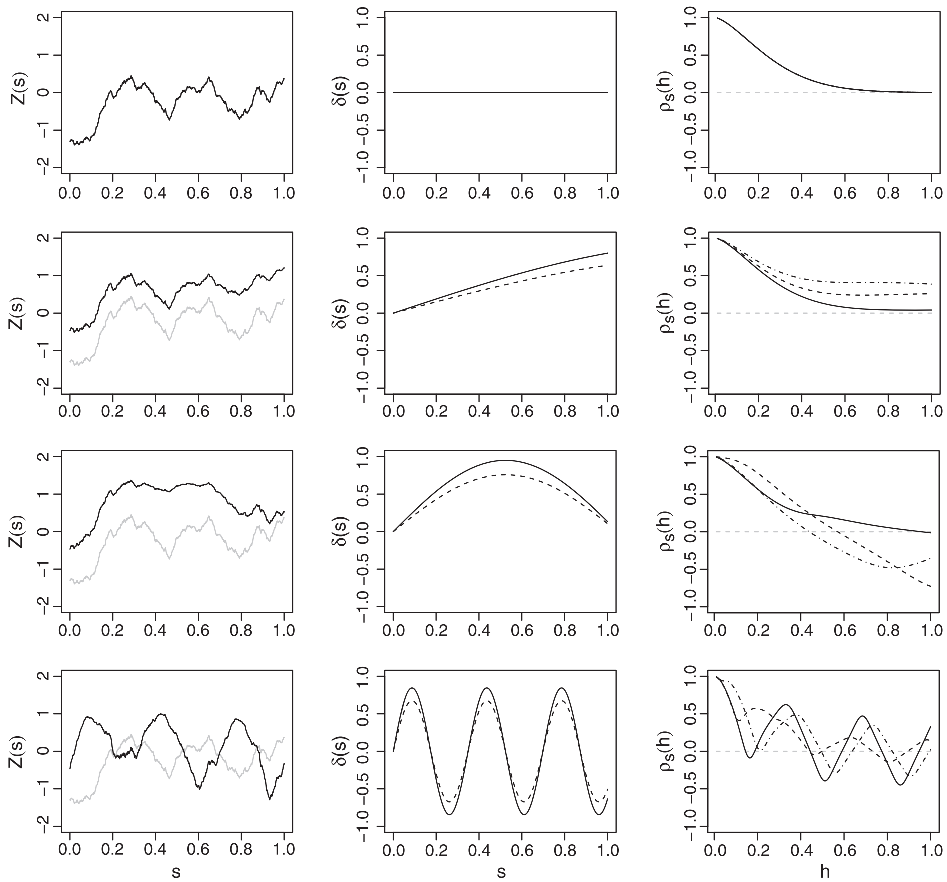

Fig. 1. Simulations from four univariate skew-normal random processes on $[0, 1]$ with $\varepsilon = 0$. The left column shows the sample path (solid line) of the simulated process $Z(s)$ and of the generating Gaussian process $X(s)$ (grey line). The middle column illustrates the slant function $\delta(s)$ (solid line) and the mean $m(s)$ of the process (dashed line). The right column displays the non-stationary correlation functions at locations $s = 0.1$ (solid line), 0.5 and 0.75 (dot-dash). Rows 1–3 use slant function $\delta(s) = a\sin(bs)$ with $a = 0.95$ and $b = 0, 1$ and 3 respectively, whereas row 4 uses $\delta(s) = a^2\sin(bs)\cos(bs)$ with $a = 1.3$ and $b = 0.9$.

Specifically, let $X(s)$ define a zero-mean, unit variance stationary Gaussian random field on $\mathbb{S}$ with correlation function $\omega(h) = \mathbb{E}\{X(s)X(s+h)\}$, and let $\bar{\Omega}$ be the $d \times d$ correlation matrix of $X(s_1), \ldots, X(s_d)$. Specify $\alpha : \mathbb{S} \mapsto \mathbb{R}$ to be a continuous square-integrable function, and let $\langle \alpha, X \rangle = \int_{\mathbb{S}} \alpha(s)X(s)\,\mathrm{d}s$ be the inner product. Let X' be a standard normal random variable independent of X and $\tau \in \mathbb{R}$. If we define

$$Z(s) = \left\{X(s) | \langle \alpha, X \rangle > X' - \tau\right\}, \qquad s \in \mathbb{S} \tag{11}$$

then, for any finite set $s_1, \ldots, s_d \in \mathbb{S}$, the distribution of $Z(s_1), \ldots, Z(s_d)$ is $\mathcal{SN}(\bar{\Omega}, \alpha, \tau)$, where $\alpha \equiv \{\alpha(s_1), \ldots, \alpha(s_d)\}$. For simplicity, we also refer to $\alpha(s)$ as the slant function. More efficient simulation of skew-normal processes can be achieved by considering the form $Z(s) = X(s)$ if $\langle \alpha, X \rangle > X' - \tau$ and $Z(s) = -X(s)$ otherwise (e.g. Azzalini, 2014, Ch.5).

We conclude this section by discussing some extremal properties of the skew-normal process $Z(s)$. For a finite sequence of points $s_1, \ldots, s_d \in \mathbb{S}$, with $d \geq 2$. Each margin $Z(s_i)$ follows a

© 2016 Board of the Foundation of the Scandinavian Journal of Statistics.

skew-normal distribution (Azzalini & Captianio A., 2014) and so is in the domain of attraction of a Gumbel distribution (Chang & Genton, 2007; Padoan, 2011). Further, each pair $(Z(s_i), Z(s_j))$ is asymptotically independent (Bortot, 2010; Lysenko *et al.*, 2009). However, in this case, a broad class of tail behaviours can still be obtained by assuming that the joint survival function is regularly varying at $+\infty$ with index $-1/\eta$ (Ledford & Tawn, 1996), so that

$$\Pr(Z(s_i) > x, Z(s_j) > x) = x^{-1/\eta}\mathscr{L}(x), \qquad x \to +\infty, \tag{12}$$

where $\eta \in (0, 1]$ is the coefficient of tail dependence and $\mathscr{L}(x)$ is a slowly varying function, that is, $\mathscr{L}(ax)/\mathscr{L}(x) \to 1$ as $x \to +\infty$, for fixed $a > 0$. Considering $\mathscr{L}$ as a constant, at extreme levels, margins are negatively associated when $\eta < 1/2$, independent when $\eta = 1/2$ and positively associated when $1/2 < \eta < 1$. When $\eta = 1$ and $\mathscr{L}(x) \nrightarrow 0$, asymptotic dependence is obtained. We derive the asymptotic behaviour of the joint survival function (12) for a pair of skew-normal margins. As our primary interest is in spatial applications, we focus on the joint upper tail of the skew-normal distribution when the variables are positively correlated or uncorrelated.

Proposition 4. *Let $Z \sim \mathcal{SN}_2(\bar{\Omega}, \alpha)$, where $\alpha = (\alpha_1, \alpha_2)^\top$ and $\bar{\Omega}$ is a correlation matrix with off-diagonal term $\omega \in [0, 1)$. The joint survivor function of the bivariate skew-normal distribution with unit Fréchet margins behaves asymptotically as (12), where*

(1) when either $\alpha_1, \alpha_2 \geq 0$ or $\omega > 0$ and $\alpha_j \leq 0$ and $\alpha_{3-j} \geq -\omega^{-1}\alpha_j$ for $j = 1, 2$, then

$$\eta = (1+\omega)/2, \qquad \mathscr{L}(x) = \frac{2\,(1+\omega)}{1-\omega}(4\pi \log x)^{-\omega/(1+\omega)};$$

(2) when $\omega > 0$, $\alpha_j < 0$ and $-\omega\,\alpha_j \leq \alpha_{3-j} < -\omega^{-1}\alpha_j$, for $j = 1, 2$, then

(a) If $\alpha_{3-j} > -\alpha_j/\bar{\alpha}_j$, then

$$\eta = \frac{(1-\omega^2)\bar{\alpha}_j^2}{1-\omega^2+(\bar{\alpha}_j-\omega)^2}, \qquad \mathscr{L}(x) = \frac{2\,\bar{\alpha}_j^2(1-\omega^2)}{(\bar{\alpha}_j^2-\omega)(1-\omega\bar{\alpha}_j)}(4\pi \log x)^{1/2\eta-1};$$

(b) If $\alpha_{3-j} < -\alpha_j/\bar{\alpha}_j$, then

$$\eta = \left[\frac{1-\omega^2+(\bar{\alpha}_j-\omega)^2}{(1-\omega^2)\bar{\alpha}_j^2} + \left(\alpha_{3-j}+\frac{\alpha_j}{\bar{\alpha}_j}\right)^2\right]^{-1},$$

$$\mathscr{L}(x) = \frac{-2^{3/2}\pi^{1/2}\bar{\alpha}_j^2(1-\omega^2)(\alpha_{3-j}+\alpha_j/\bar{\alpha}_j)^{-1}}{(\bar{\alpha}_j-\omega)\{1-\omega\bar{\alpha}_j+\alpha_j(\alpha_j+\alpha_{3-j}\bar{\alpha}_j)(1-\omega^2)\}}(4\pi \log x)^{1/2\eta-3/2};$$

1. when either $\alpha_1, \alpha_2 < 0$ or $\omega > 0$, $\alpha_j < 0$ and $0 < \alpha_{3-j} < -\omega\,\alpha_j$ for $j = 1, 2$, then

$$\eta = \left\{\frac{1}{1-\omega^2}\left(\frac{\alpha_{3-j}^2(1-\omega^2)+1}{\bar{\alpha}_{3-j}^2} + \frac{\alpha_j^2(1-\omega^2)+1}{\bar{\alpha}_j^2} + \frac{2(\alpha_{3-j}\alpha_j(1-\omega^2)-\omega)}{\bar{\alpha}_{3-j}\bar{\alpha}_j}\right)\right\}^{-1},$$

$$\mathscr{L}(x) = \frac{-2^{3/2}\pi^{1/2}\bar{\alpha}_j^{3/2}\bar{\alpha}_{3-j}^2(1-\omega^2)(\alpha_i\bar{\alpha}_j+\alpha_j\bar{\alpha}_{3-j})^{-1}}{(\bar{\alpha}_j-\omega\bar{\alpha}_{3-j})\{1-\omega\bar{\alpha}_j+\alpha_j(\alpha_j+\alpha_{3-j}\bar{\alpha}_j/\bar{\alpha}_{3-j})(1-\omega^2)\}}(4\pi \log x)^{1/2\eta-3/2};$$

where $\bar{\alpha}_j = \sqrt{1+\alpha_j^{*2}}$ and $\alpha_j^* := \alpha_{\{j\}}^* = \frac{\alpha_j+\omega\alpha_{3-j}}{\sqrt{1+\alpha_{3-j}(1-\omega^2)}}$.

© 2016 Board of the Foundation of the Scandinavian Journal of Statistics.

14679469, 2017, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.12240 by University of New South Wales, Wiley Online Library on [04/08/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Proof in Appendix A.3.

As a result, when both marginal parameters are non-negative (case 1), then $1/2 \leq \eta < 1$, with $\eta = 1/2$ occurring when $\omega = 0$. As a consequence, as for the Gaussian distribution (for which $\alpha = 0$), the marginal extremes are either positively associated or exactly independent. The marginal extremes are also completely dependent when $\omega = 1$, regardless of the values of the slant parameters, α. When one marginal parameter is positive and one is negative (case 2), then $\eta > (1 + \omega)/2$. In this case, the extreme marginals are also positively associated, but the dependence is greater than when the random variables are normally distributed. Finally, when both marginal parameters are negative (case 3), then $0 < \eta < 1/2$, implying that the extreme marginals are negatively associated, although $\omega > 0$. It should be noted that differently from the Gaussian case ($\alpha = 0$) where $\omega > 0$ implies a positive association, in this case, it is not necessarily true. In summary, the degree of dependence in the upper tail of the skew-normal distribution ranges from negative to positive association and includes independence.

3. Spectral representation for the extremal skew-t process

The spectral representation of stationary max-stable processes with common unit Fréchet margins can be constructed using the fundamental procedures introduced by de Haan (1984) and Schlather (2002) (see also de Haan & Ferreria, 2006, Ch. 9). This representation can be formulated in broader terms resulting in max-stable processes with ν-Fréchet univariate marginal distributions, with $\nu > 0$ (Opitz, 2013). In order to state our result, we rephrase the spectral representation to also take into account non-stationary processes.

Let $\{Y(s)\}_{s\in\mathbb{S}}$ be a non-stationary real-valued stochastic process with continuous sample path on $\mathbb{S}$ such that $\mathbb{E}\{\sup_{s\in\mathbb{S}} Y(s)\} < \infty$ and $m^+(s) = \mathbb{E}[\{Y^+(s)\}^\nu] < \infty, \forall s \in \mathbb{S}$ for $\nu > 0$, where $Y^+(\cdot) = \max\{Y(\cdot), 0\}$ denotes the positive part of Y. Let $\{R_i\}_{i\geq 1}$ be the points of an inhomogeneous Poisson point process on $(0,\infty)$ with intensity $\nu r^{-(\nu+1)}$, $\nu > 0$, which are independent of Y. Define

$$U(s) = \max_{i=1,2,\dots} \{R_i Y_i^+(s)\}/\{m^+(s)\}^{1/\nu}, \quad s \in \mathbb{S}, \tag{13}$$

where $Y_1, Y_2, \dots$ are iid copies of Y. Then U is a max-stable process with common ν-Fréchet univariate margins. In particular, for fixed $s \in \mathbb{S}$ and $x(s) > 0$, we have

$$\Pr(U(s) \leq x(s)) = \exp\left[-\frac{\mathbb{E}\{Y^+(s)\}^\nu}{x^\nu(s)m^+(s)}\right] = \exp\{-1/x^\nu(s)\},$$

and for fixed $s_1, \dots, s_d$, the finite-dimensional distribution of U has exponent function

$$V(x(s_1), \dots, x(s_d)) = \mathbb{E}\left(\max_j \left[\frac{\{Y^+(s_j)/x(s_j)\}^\nu}{m^+(s_j)}\right]\right), \quad x(s_j) > 0,\ j = 1, \dots, d \tag{14}$$

(de Haan & Ferreria, 2006, Ch. 9).

In this construction, the impact of a non-stationary process $Y(s)$ would be that the dependence structure of the max-stable process $U(s + h)$ depends on both the separation h and the location $s \in \mathbb{S}$ and would therefore itself be non-stationary. The succeeding theorem derives a max-stable process $U(s)$ when $Y(s)$ is the skew-normal random field introduced in Section 2.2.

© 2016 Board of the Foundation of the Scandinavian Journal of Statistics.

14679469, 2017, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.12240 by University of New South Wales, Wiley Online Library on [04/08/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Theorem 1 (Extremal skew-t process). *Let $Y(s)$ be a skew-normal random field on $s \in \mathbb{S}$ with finite-dimensional distribution $\mathcal{SN}_d(\bar{\Omega}, \alpha, \tau)$, as defined in (11). Then the max-stable process $U(s)$, given by (13), has ν-Fréchet univariate marginal distributions and exponent function*

$$V(x_j, j \in I) = \sum_{j=1}^{d} x_j^{-\nu} \Psi_{d-1}\left(\left(\sqrt{\frac{\nu+1}{1-\omega_{i,j}^2}}\left(\frac{x_i^\circ}{x_j^\circ} - \omega_{i,j}\right), i \in I_j\right)^\top ; \bar{\Omega}_j^\circ, \alpha_j^\circ, \tau_j^\circ, \kappa_j^\circ, \nu+1\right), \tag{15}$$

where $x_j \equiv x(s_j)$, Ψ_{d-1} is a $(d-1)$-dimensional non-central extended skew-t distribution (Definition 1) with correlation matrix $\bar{\Omega}_j^\circ$, shape, extension and non-centrality parameters α_j°, τ_j° and κ_j°, $\nu+1$ degrees of freedom, $I = \{1, \dots, d\}$, $I_j = I \backslash \{j\}$ and $\omega_{i,j}$ is the (i, j)-th element of $\bar{\Omega}$.

Proof (and further details) in Appendix A.4.

We call the process $U(s)$ with exponent function (15) an extremal skew-t process.

Note that in Theorem 1, when $\tau = 0$, and the slant function is such that $\alpha(s) \equiv 0$ for all $s \in \mathbb{S}$, then the exponent function (15) becomes

$$V(x_j, j \in I) = \sum_{j \in I} x_j^{-\nu} \Psi_{d-1}\left[\left(\sqrt{\frac{\nu+1}{1-\omega_{i,j}^2}}\left(\frac{x_i}{x_j} - \omega_{i,j}\right), i \in I_j\right)^\top ; \bar{\Omega}_j^\circ, \nu+1\right]. \tag{16}$$

This is the exponent function of the extremal-t process as discussed in Opitz (2013).

If we assume $\tau = 0$ in (11), then the bivariate exponent function of the extremal skew-t process seen as a function of the separation h is equal to

$$V\{x(s), x(s+h)\} = \frac{\Psi(b(x_s^*(h)); \alpha_s^*(h), \tau_s^*(h), \nu+1)}{x^\nu(s)} + \frac{\Psi(b(x_s^+(h)); \alpha_s^+(h), \tau_s^+(h), \nu+1)}{x^\nu(s+h)}$$

where Ψ is a univariate extended skew-t distribution, $b(\cdot) = \sqrt{\frac{\nu+1}{1-\omega^2(h)}}(\cdot - \omega(h))$,

$$x_s^*(h) = \frac{x(s+h)\Gamma_s(h)}{x(s)}, \qquad x_s^+(h) = \frac{x(s)}{x(s+h)\Gamma_s(h)},$$

$$\alpha_s^*(h) = \alpha(s+h)\sqrt{1-\omega^2(h)}, \qquad \alpha_s^+(h) = \alpha(s)\sqrt{1-\omega^2(h)},$$

$$\tau_s^*(h) = \sqrt{\nu+1}\{\alpha(s) + \alpha(s+h)\omega(h)\}, \qquad \tau_s^+(h) = \sqrt{\nu+1}\{\alpha(s+h) + \alpha(s)\omega(h)\},$$

and

$$\Gamma_s(h) = \left(\frac{\Psi\left[\alpha(s) + \alpha(s+h)\omega(h)\sqrt{\frac{\nu+1}{\alpha^2(s+h)\{1-\omega^2(h)\}}}; \nu+1\right]}{\Psi\left[\alpha(s+h) + \alpha(s)\omega(h)\sqrt{\frac{\nu+1}{\alpha^2(s)\{1-\omega^2(h)\}}}; \nu+1\right]}\right)^{1/\nu}.$$

Clearly, as the dependence structure depends on both correlation function $\omega(h)$ and the slant function $\alpha(s)$, and therefore on the value of $s \in \mathbb{S}$, it is a non-stationary dependence structure.

© 2016 Board of the Foundation of the Scandinavian Journal of Statistics.

14679469, 2017, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.12240 by University of New South Wales, Wiley Online Library on [04/08/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

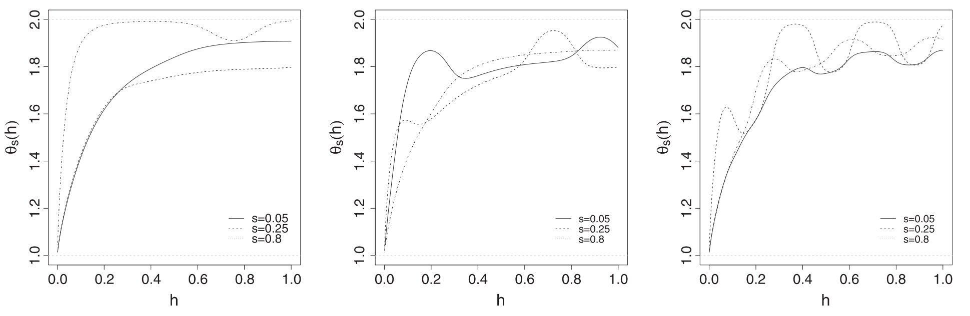

Fig. 2. Examples of univariate ($k = 1$) non-stationary isotropic extremal coefficient functions $\theta_s(h)$, for the extremal skew-t process over $s \in [0, 1]$, using correlation function (10), where $h \in [0, 1]$, $\lambda = 1.5$ and $\gamma = 0.3$. Slant functions are as follows (left to right panels): $\alpha(s) = -1 - s + \exp\{\sin(5s)\}, \alpha(s) = 1 + 1.5s - \exp\{\sin(8s)\}$ and $\alpha(s) = 2.25\sin(9s)\cos(9s)$. Solid, dashed and dot-dashed lines represent the fixed locations $s = 0.05, 0.25$ and 0.8, respectively.

From the bivariate exponent function, we can derive the non-stationary extremal coefficient function, using the relation $\theta_s(h) = V(1, 1)$, which gives

$$\theta_s(h) = \Psi(b(\Gamma_s(h))\,\alpha_s^*(h), \tau_s^*(h), \nu + 1) + \Psi\left(b(1/\Gamma_s(h)); \alpha_s^+(h), \tau_s^+(h), \nu + 1\right). \quad (17)$$

Figure 2 shows some examples of univariate ($k = 1$) non-stationary isotropic extremal coefficient functions obtained from (17) using the power-exponential correlation function (10). Each panel illustrates a different slant function $\alpha(s)$, with the line-types indicating the fixed location value of $s \in \mathbb{S}$. The extremal coefficient functions $\theta_s(h)$ increase as the value of h increases, meaning that the dependence of extremes decreases with the distance. $\theta_s(h)$ grows with different rates depending on the location $s \in \mathbb{S}$. Although the ergodicity and mixing properties of the process must be investigated, numerical results show that for some s, $\theta_s(h) \to 2$ as $|h| \to +\infty$. By increasing the complexity of the slant function (e.g. centre and right panels), it is possible to construct extremal coefficient functions, which exhibit stronger dependence for larger distances, h, compared with shorter distances. Similarly, Fig. 3 illustrates examples of bivariate ($k = 2$) non-stationary geometric anisotropic extremal coefficient functions, $\theta_s(h)$, also obtained from (17). Similar interpretations to the univariate case can be made (Fig. 2), in addition to noting that the level of dependence is affected by the direction (from the origin).

4. Inference for extremal skew-t processes

Parametric inference for the extremal skew-t process can be performed in two ways. The first uses the marginal composite likelihood approach (e.g. Padoan *et al.*, 2010; Davison & Gholamrezaee, 2012; Huser & Davison, 2013), because only marginal densities of dimension up to $d = 4$ are practically available (Supporting Information).

Let $\vartheta \in \Theta \subseteq \mathbb{R}^p$, $p = 1, 2, \ldots$, denote the vector of dependence parameters of the extremal skew-t process. Consider a sample $(x_i, i = 1, \ldots, n)$ with $x_i \in \mathbb{R}_+^d$ of n iid replicates of the process observed over a finite number of points $(s_j, j \in I)$ with $s_j \in \mathbb{S}$. For simplicity, it is assumed that the univariate marginal distributions are unit Fréchet. The pairwise or triplewise ($m = 2, 3$) log-composite likelihood is defined by

$$\ell_m(\vartheta; x) = \sum_{i=1}^{n} \sum_{E \in E_m} \log f(x_i \in E; \vartheta), \quad m = 2, 3,$$

© 2016 Board of the Foundation of the Scandinavian Journal of Statistics.

14679469, 2017, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.12240 by University of New South Wales, Wiley Online Library on [04/08/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

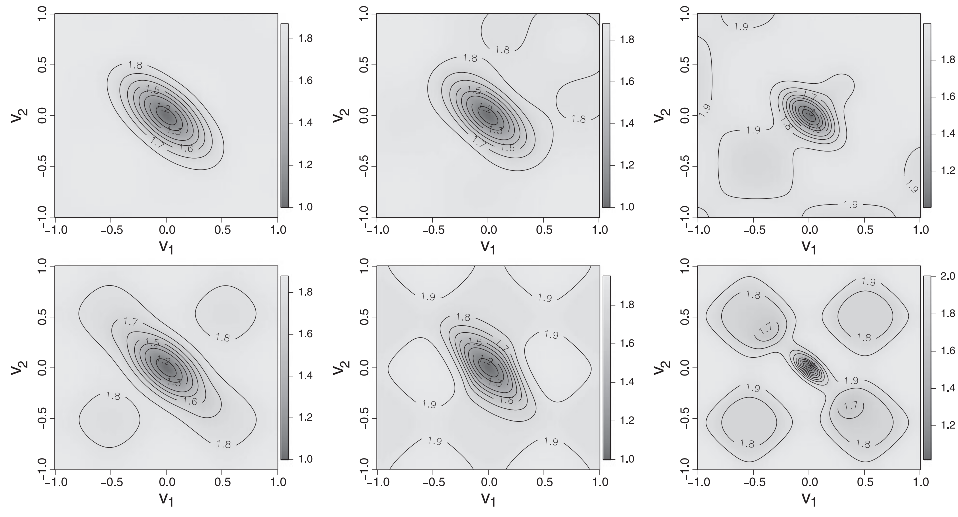

Fig. 3. Bivariate ($k = 2$) geometric anisotropic non-stationary extremal coefficient functions $\theta_s(h)$, for the extremal skew-t process on $s \in [0,1]^2$, based on extremal coefficient function (17) with $\lambda = 1.5$ and $\gamma = 0.3$, where $h = v^\top R v$, $v = (v_1, v_2)^\top \in [-1,1]^2$ and R is a 2×2 matrix whose diagonal elements are 2.5 and off-diagonal elements 1.5. Slant functions are $\alpha(s) = \exp\{\sin(4s_1)\sin(4s_2) - s_1 s_2 - 1\}$ (top panels) and $\alpha(s) = 2.25\{\sin(3s_1)\cos(3s_1) + \sin(3s_2)\cos(3s_2)\}$ (bottom), with $s = (s_1, s_2)^\top \in [0,1]^2$. Left to right: panels are based on fixing $s = (0.2, 0.2)^\top$, $s = (0.4, 0.4)^\top$ and $s = (0.85, 0.85)^\top$ (top panels) and $s = (0.25, 0.25)^\top$, $s = (0.25, 0.8)^\top$ and $s = (0.8, 0.8)^\top$ (bottom).

where $x = (x_1, \ldots, x_n)^\top$ with $x_i \in \mathbb{R}_+^m$ and f is a marginal extremal skew-t pdf associated with each member of a set of marginal events E_m. See, for example, Varin *et al.* (2011) for a complete description of composite likelihood methods.

A second approach is to use the approximate likelihood function introduced by Coles & Tawn (1994), which is constructed on the space of angular densities. The angular measure of the extremal skew-t dependence model (15) places mass on the interior as well as on all the other subspaces of the simplex, such as the edges and the vertices. We derive some of these densities following the results in Coles & Tawn (1991).

Let J be an index set that takes values in $\mathbb{I} = \mathbb{P}(\{1, \ldots, d\}) \backslash \emptyset$, where $\mathbb{P}(I)$ is the power set of I. For any fixed d and all $J \in \mathbb{I}$, the sets

$$\mathbb{W}_{d,J} = (w \in \mathbb{W} : w_j = 0, \text{ if } j \notin J;\ w_j > 0 \text{ if } j \in J)$$

provide a partition of the d-dimensional simplex $\mathbb{W}$ into $2^d - 1$ subsets. Let $k = |J|$ be the size of J. Let $h_{d,J}$ denote the density that lies on the subspace $\mathbb{W}_{d,J}$, which has $k-1$ free parameters w_j such that $j \in J$. When $J = \{1, \ldots, d\}$, the angular density in the interior of the simplex is

$$h(w) = \frac{\psi_{d-1}\left(\left(\left[\sqrt{\frac{\nu+1}{1-\omega_{i,1}^2}}\left\{\left(\frac{w_i^\circ}{w_1^\circ}\right)^{1/\nu} - \omega_{i,1}\right\}, i \in I_1\right]^\top; \Omega_1^\circ, \alpha_1^\circ, \tau_1^\circ, \kappa_1^\circ, \nu+1\right)\right)}{w_1^{(d+1)}\left\{\prod_{i=2}^d \frac{1}{\nu}\sqrt{\frac{\nu+1}{1-\omega_{i,1}^2}}\left(\frac{w_i^\circ}{w_1^\circ}\right)^{\frac{1}{\nu}-1}\frac{m_i^+}{m_1^+}\right\}^{-1}}, \quad w \in \mathbb{W} \tag{18}$$

© 2016 Board of the Foundation of the Scandinavian Journal of Statistics.

14679469, 2017, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.12240 by University of New South Wales, Wiley Online Library on [04/08/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

where ψ_{d-1} denotes the $d-1$-dimensional skew-t density, $I_j = \{1,\ldots,d\}\backslash j$ and where the parameters $\Omega_1^\circ, \alpha_1^\circ, \tau_1^\circ, \kappa_1^\circ$ and $w_i^\circ = w_i(m_i^+)^{1/\nu}$ are given in the proof to Theorem 1 (Appendix A.4). When $J = \{i_1,\ldots,i_k\} \subset \{1,\ldots,d\}$, the angular density for any $x \in \mathbb{R}_+^d$ is

$$h_{d,J}\left(\frac{x_{i_1}}{\sum_{i\in J} x_i},\cdots,\frac{x_{i_{k-1}}}{\sum_{i\in J} x_i}\right) = -\left(\sum_{i\in J} x_i\right)^{k+1} \lim_{\substack{x_j\to 0,\\ j\notin J}} \frac{\partial^k V}{\partial x_{i_1}\cdots\partial x_{i_k}}(x). \tag{19}$$

Thus, when $J = \{j\}$ for any $j \in \{1,\ldots,d\}$, then $\mathbb{W}_{d,J}$ is a vertex $\mathbf{e}_j$ of the simplex, and the density is a point mass, denoted $h_{d,J} = H(\{\mathbf{e}_j\})$. In this case, (19) reduces to

$$h_{d,J} = \Psi_{d-1}\left\{\left(-\sqrt{\frac{\nu+1}{1-\omega_{i,j}^2}}\omega_{i,j}, i \in I_j\right)^\top; \Omega_j^\circ, \alpha_j^\circ, \tau_j^\circ, \kappa_j^\circ, \nu+1\right\}, \tag{20}$$

where Ψ_{d-1} denotes the $d-1$-dimensional skew-t distribution with parameters again given in the proof to Theorem 1 (Appendix A.4).

Computations of all $2^d - 1$ densities that lie on the edges and vertices of the simplex are available for $d = 3$. In this case, the angular densities on the interior and vertices of the simplex can be deduced from (18) and (20). For all $i, j \in J = \{1,2,3\}$, with $i \neq j$, the angular density on the edges of $\mathbb{W}_{d,J}$ for $w \in \mathbb{W}_{d,J}$ is given by

$$\begin{aligned}
h_{3,\{i,j\}}(w) = \sum_{u,v\in\{i,j\},u\neq v} &\left(\frac{\psi(b_{u,v}^\circ;\nu+1)}{\Psi(\bar\tau_u;\nu+1)}\Psi_2\left[\left\{y_1^\circ(u,v), y_2^\circ(u,v)\right\}^\top; \bar\Omega_u^{\circ\circ}, \nu+2\right]\right.\\
&\times \frac{1}{w_1}\left\{\frac{\mathrm{d}^2 b_{u,v}^\circ}{\mathrm{d}w_u\mathrm{d}w_v} + \frac{\mathrm{d}b_{u,v}^\circ}{\mathrm{d}w_v}\left(\frac{\mathrm{d}b_{u,v}^\circ}{\mathrm{d}w_u}\frac{(\nu+2)b_{u,v}^\circ}{\nu+1+b_{u,v}^{\circ 2}} - \frac{1}{w_1}\right)\right\}\\
&+ \psi\{y_1^\circ(u,v);\nu+2\}\sqrt{\frac{\nu+2}{1-\Omega_{u,[1,2]}^{\circ 2}}}\frac{b_{u,v}^\circ c_{u,\bar k} + \Omega_{u,[1,2]}^{\circ 2}(\nu+1)}{\left(\nu+1+b_{u,v}^{\circ 2}\right)^{3/2}}\\
&\times \Psi\left(\frac{\sqrt{\nu+3}\left\{z_2^\circ(u,v)\Omega_{u,[1,1]}^{\circ\circ} - z_1^\circ(u,v)\Omega_{u,[1,2]}^{\circ\circ}\right\}}{\sqrt{\left[\Omega_{u,[1,1]}^{\circ\circ}\left\{\nu+1+b_{u,v}^{\circ 2}\right\} + z_1^{\circ 2}(u,v)\right]\det(\Omega_u^{\circ\circ})}}; \nu+3\right)\\
&+ \psi\{y_2^\circ(u,v);\nu+2\}\sqrt{\frac{\nu+2}{1-\Omega_{u,[1,3]}^{*2}}}\frac{x(u,v)\bar\tau_u + \Omega_{u,[1,3]}^{*2}(\nu+1)}{\left\{\nu+1+b_{u,v}^{\circ 2}\right\}^{3/2}}\\
&\left.\times\Psi\left\{\frac{\sqrt{\nu+3}\left\{z_1^\circ(u,v)\Omega_{u,[2,2]}^{\circ\circ} - z_2^\circ(u,v)\Omega_{u,[1,2]}^{\circ\circ}\right\}}{\sqrt{\left(\Omega_{u,[2,2]}^{\circ\circ}\left\{\nu+1+b_{u,v}^{\circ 2}\right\} + z_2^\circ(u,v)^2\right)\det(\Omega_u^{\circ\circ})}}; \nu+3\right\}\right),
\end{aligned} \tag{21}$$

where for all $u, v \in J$, with $u \neq v$, and $\bar k \notin \{i,j\}$,

$$y_\ell^\circ(u,v) = \frac{z_\ell^\circ(u,v)}{\sqrt{\Omega_{u,[\ell,\ell]}^\circ}}\sqrt{\frac{\nu+2}{\nu+1+b_{u,v}^{\circ 2}}},\ \ell = 1,2, \quad z_1^\circ(u,v) = c_{u,\bar k} - \Omega_{u,[1,2]}^\circ b_{u,v}^\circ,$$

$$c_{u,v} = -\omega_{u,v}\sqrt{\frac{\nu+1}{1-\omega_{u,v}^2}}, \quad z_2^\circ(u,v) = \bar\tau_u - \Omega_{u,[1,3]}^\circ, \quad b_{u,v}^\circ = \sqrt{\frac{\nu+1}{1-\omega_{u,v}^2}}\left(\left(\frac{w_v^\circ}{w_u^\circ}\right)^{1/\nu} - \omega_{u,v}\right),$$

$$\Omega_u^\circ = \begin{bmatrix}\bar\Omega_u & -\delta_u\\ -\delta_u^\top & 1\end{bmatrix}, \quad \delta_u^\top = \bar\Omega_u\left(\alpha_v\sqrt{1-\omega_{u,v}^2}, \alpha_k\sqrt{1-\omega_{u,k}^2}\right)^\top, \quad \bar\Omega_u^{\circ\circ} = \omega_u^{\circ-1/2}\Omega_u^{\circ\circ}\omega_u^{\circ-1/2},$$

© 2016 Board of the Foundation of the Scandinavian Journal of Statistics.

14679469, 2017, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.12240 by University of New South Wales, Wiley Online Library on [04/08/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$\omega_u^\circ = \text{diag}(\Omega_u^{\circ\circ})$, $\Omega_u^{\circ\circ} = \Omega_{u,[-1,-1]}^\circ - \Omega_{u,[-1,1]}^\circ \Omega_{u,[1,-1]}^\circ$. Components of Ω_u° and $\Omega_u^{\circ\circ}$ are, respectively, given by $\Omega_{u,[i,j]}^\circ$ and $\Omega_{u,[i,j]}^{\circ\circ}$ for $i, j \in J$. See also Appendix A.4 for further details. When $\tau = 0$ and $\alpha(s) = 0$, then the densities (18), (20) and (21) reduce to the densities of the extremal-t dependence model. A graphical illustration that shows the difference between the two dependence models is provided in the Supporting Information.

Therefore, for $d = 3$, the estimation of dependence parameters can be based on the following approach. Let $\{(r_i, w_i) : i = 1, \ldots, n\}$ be the set of observations, where $r_i = \sum_{j \in I} x_{i,j}$ and $w_i = x_i / r_i$, with $x_i = (x_{i,j})_{j \in I}$, are pseudo-polar radial and angular components. Then the approximate log-likelihood is defined by

$$\ell(\vartheta; \tilde{w}) = \sum_{\substack{i=1,\ldots,n: \\ r_i > r_0}} \log h(w_i; \vartheta), \tag{22}$$

where $\tilde{w} = (w_1, \ldots, w_n)^\top$, for some radial threshold $r_0 > 0$, and where h is the angular density function of the extremal skew-t dependence model. The components of the sum in (22) comprise the three types of angular densities lying on the interior, edges and vertices of the simplex. Whether an angular component belongs either to the interior, an edge or a vertex of the simplex, producing the associated density is determined according to the following criterion. We select a threshold $c \in [0, 0.1]$, and we construct the following partitions for an arbitrary observation $w_i = (w_{i,j}, w_{i,k}, w_{i,l})$, $i = 1, \ldots, n$. Set $w \equiv w_i$ for simplicity. When $\mathcal{C}_j = \{w_j > 1 - c; j \in I\}$, then an observation belongs to vertex $\boldsymbol{e}_j$. When $\mathcal{E}_{j,k} = \{w_j, w_k < 1 - c, w_l < c, w_j > 1 - 2w_k, w_k > 1 - 2w_j; j \in I, k \in I_j, l \in I_j \backslash \{k\}\}$, then an observation belongs to edge between the jth and kth components. When $\mathcal{I} = \{w_j > c; j \in I\}$, then an observation belongs to the interior (see the Supporting Information for more details). The components of the angular density $h(w)$ then require rescaling so that they satisfy the constraints of valid angular densities – namely that they integrate to the number of components of w (three in the trivariate case) – while also respecting the partition of $\mathbb{W}$ implied by c. Without this rescaling, then the likelihood of, for example, the model that places mass on all subsets of the simplex is not comparable with that of models that places mass only on subsets of the simplex. Specifically

$$\int_{\mathbb{W}} h(w) \mathrm{d}w = K_{\mathcal{C}} \sum_{j \in I} \int_{\mathcal{C}_j} h_{3,\{j\}} \mathrm{d}w + \sum_{\substack{j=1,2 \\ k=j+1,3}} K_{\mathcal{E}_{j,k}} \int_{\mathcal{E}_{j,k}} h_{3,\{j,k\}}(w) \mathrm{d}w + K_{\mathcal{I}} \int_{\mathcal{I}} h_{3,\{1,2,3\}}(w) \mathrm{d}w = 3,$$

where

$$K_{\mathcal{C}} = \frac{4}{\sqrt{3}c^2}, \quad K_{\mathcal{E}_{j,k}} = \frac{2\int_0^1 h_{3,\{j,k\}}(w) \mathrm{d}w}{c\sqrt{3}(1 - 2c)}, \quad K_{\mathcal{I}} = \frac{\int_0^1 \int_0^1 h_{3,\{1,2,3\}}(w) \mathrm{d}w}{\int_c^{1-2c} \int_c^{1-2c} h_{3,\{1,2,3\}}(w) \mathrm{d}w},$$

and $h_{3,\{j\}}$, $h_{3,\{j,k\}}(w)$ and $h_{3,\{1,2,3\}}(w)$ are defined earlier. Note that for $j, k \in I$ with $j \neq k$, we have that $h_{3,\{j,k\}}(w) = h_{3,\{k,j\}}(w)$. In the bivariate case ($d = 2$), the appropriate modification only considers the mass on the vertices and interior.

We now illustrate the ability of the approximate likelihood in estimating the extremal dependence parameters in the bivariate and trivariate cases. We generate 500 replicate datasets of sizes 5000 (bivariate) and 1000 (trivariate), with parameters $\vartheta_2 = (\omega, \nu) = (0.6, 1.5)$ and $\vartheta_3 = (\omega_{1,2}, \omega_{1,3}, \omega_{2,3}, \nu) = (0.6, 0.8, 0.7, 1)$. Each dataset is transformed to pseudo-polar

© 2016 Board of the Foundation of the Scandinavian Journal of Statistics.

14679469, 2017, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.12240 by University of New South Wales, Wiley Online Library on [04/08/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

coordinates, and the 100 observations with the largest radial component are retained. Parameters are estimated through the profile likelihood where the dependence parameter ω is the parameter of interest and the degree of freedom ν is considered as a nuisance parameter. Parameters are estimated for different values of the threshold $c = 0, 0.02, 0.04, 0.06, 0.08, 0.1$. In order to compare likelihoods for different values of c, the likelihood functions are evaluated using those data points considered to belong to the interior of the simplex, multiplied by the mass at the corners and/or edges in proportion to their rescaling constants.

Figures 4 and 5 provide (left to right) boxplots of the resulting estimates of the dependence parameter(s) ω, the degree of freedom ν and of the likelihood function for increasing values of c, for the 500 replicate datasets for both bivariate and trivariate cases. The true parameter values are indicated by the horizontal lines.

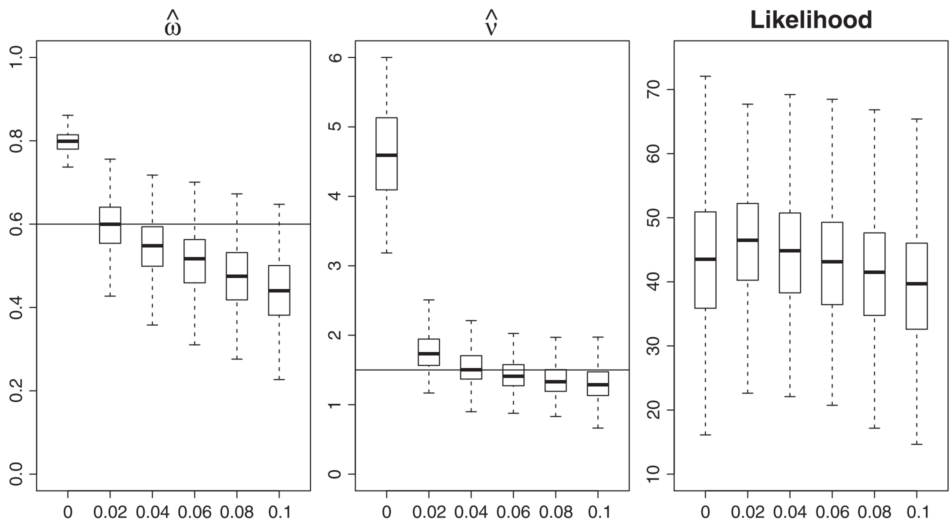

Fig. 4. Left to right: Boxplots of the estimates of the dependence parameter ω, the degree of freedom ν and the associated maximum of the likelihood function based on the rescaled angular density, when $c = 0, 0.02, 0.04, 0.06, 0.08$ and 0.1. Boxplots are constructed from 500 replicate datasets of size 5000. Horizontal lines indicate the true values $\omega = 0.6$ and $\nu = 1.5$.

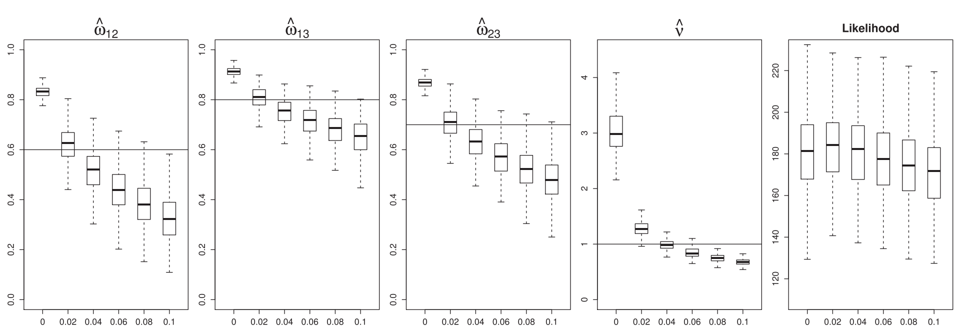

Fig. 5. Left to right: Boxplots of the estimates of the dependence parameter $\omega = (\omega_{1,2}, \omega_{1,3}, \omega_{2,3})$, the degree of freedom ν and the associated maximum of the likelihood function based on the rescaled angular density, when $c = 0, 0.02, 0.04, 0.06, 0.08$ and 0.1. Boxplots are constructed from 500 replicate datasets of size 1000. Horizontal lines indicate the true values $\omega_{1,2} = 0.6, \omega_{1,3} = 0.7, \omega_{2,3} = 0.7$ and $\nu = 1$.

© 2016 Board of the Foundation of the Scandinavian Journal of Statistics.

14679469, 2017, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.12240 by University of New South Wales, Wiley Online Library on [04/08/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

In the rightmost panel of each figure, the largest values of the log-likelihood are globally obtained for $c = 0.02$, for which the most accurate estimates of ω and ν are also obtained. Conditional on $c = 0.02$, the mean estimates are $\hat{\omega} = 0.55$ and $\hat{\nu} = 1.79$ in the bivariate case and $\hat{\omega} = (0.62, 0.80, 0.71)$ and $\hat{\nu} = 1.27$ in the trivariate case. Note that the degree of freedom ν appears to be slightly overestimated and appears to be better estimated for slightly larger values of c. Overall, this procedure appears capable of efficiently estimating the model parameters. Note that increased precision of estimates can be obtained by considering a denser range of threshold values c.

An independent study comparing the efficiency of the maximum pairwise and triplewise composite likelihood estimators is provided in the Supporting Information.

5. Application to wind speed data

We illustrate the use of the extremal skew-t process using wind speed data (the weekly maximum wind speed in km/h), collected from four monitoring stations across Oklahoma, USA, over the March–May period during 1996–2012, as part of a larger dataset of 99 stations. An analysis establishing the significant marginal, station-specific skewness of these data is presented in the Supporting Information. Here, we focus on the dependence structure between stations, where for simplicity, the data are marginally transformed to unit Fréchet distributions. Only extremal-t and extremal skew-t models are considered, and parameter estimation is performed via pairwise composite likelihoods as detailed at the beginning of Section 4.

Model comparison is performed through the composite likelihood information criterion (CLIC; Varin *et al.*, 2011) given by

$$\text{CLIC} = -2\left[\ell_2(\hat{\vartheta}; x) - \text{tr}\left\{\hat{J}(\hat{\vartheta})\hat{H}(\hat{\vartheta})^{-1}\right\}\right],$$

where $\hat{\vartheta}$ is the maximum composite likelihood estimate of ϑ, $\ell_2(\hat{\vartheta}; x)$ is the maximized pairwise composite likelihood and $\hat{J}$ and $\hat{H}$ are estimates of $J(\vartheta) = \text{Var}_U(\nabla \ell_2(\vartheta; U))$ and $H(\vartheta) = \mathbb{E}_U(-\nabla^2 \ell_2(\vartheta; U))$, the variability and sensibility (hessian) matrices, where U is a bivariate random vector with extremal skew-t distribution.

Table 1 presents the pairwise composite likelihood estimates of $\omega = (\omega_{12}, \omega_{13}, \omega_{23})$, $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ and ν for the extremal-t and extremal skew-t models, obtained for all triplewise combinations of the four locations CLOU, CLAY, PAUL and SALL. For each triple, the

Table 1. *Pairwise composite likelihood estimates* $\hat{\vartheta} = (\hat{\omega}, \hat{\nu})$ *and* $\hat{\vartheta} = (\hat{\omega}, \hat{\alpha}, \hat{\nu})$ *of the extremal-t (ext-t) and extremal skew-t (ex-skew-t) models, respectively, for all possible triplets of the four locations CLOU, CLAY, PAUL and SALL. Standard errors (se) are shown for* $\hat{\alpha}$ *only*

Stations	Model	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\nu}$	CLIC
(CLOU,CLAY,SALL)	ex-t	$(0.67, 0.57, 0.69)$	—	2.89	5395.73
	ex-skew-t	$(0.42, 0.74, 0.52)$	$(-0.80, 2.88, -0.23)$	2.06	5385.07
			se: $(0.04, 0.14, 0.03)$		
(CLOU,CLAY,PAUL)	ex-t	$(0.59, 0.50, 0.69)$	—	2.53	5503.54
	ex-skew-t	$(0.45, 0.29, 0.65)$	$(-0.68, 21.07, 23.41)$	2.16	5496.90
			se: $(0.05, 0.97, 1.09)$		
(CLAY,SALL,PAUL)	ex-t	$(0.65, 0.61, 0.53)$	—	1.55	5086.13
	ex-skew-t	$(0.56, 0.51, 0.39)$	$(3.55, 2.36, 8.49)$	1.29	5075.87
			se: $(0.17, 0.15, 0.63)$		
(CLOU,SALL,PAUL)	ex-t	$(0.37, 0.40, 0.42)$	—	1.88	5428.04
	ex-skew-t	$(0.29, 0.30, 0.37)$	$(-0.14, 1.04, 34.70)$	2.11	5419.27
			se: $(0.03, 0.02, 3.49)$		

© 2016 Board of the Foundation of the Scandinavian Journal of Statistics.

14679469, 2017, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.12240 by University of New South Wales, Wiley Online Library on [04/08/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14679469, 2017, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.12240 by University of New South Wales, Wiley Online Library on [04/08/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 2. *Extremal-t and extremal skew-t conditional probabilities of exceeding particular fixed thresholds of the form $Pr(X > x \mid Y > y, Z > z)$ and $Pr(X > x, Y > y \mid Z > z)$, along with empirical estimates. The wind speed thresholds (x, y, z) are constructed from the marginal quantiles $q^{70} = (q^{70}_{CO}, q^{70}_{CA}, q^{70}_{SA}, q^{70}_{PA}) = (18.04, 20.33, 24.18, 23.61)$ and $q^{90} = (q^{90}_{CO}, q^{90}_{CA}, q^{90}_{SA}, q^{90}_{PA}) = (22.11, 24.33, 29.05, 28.26)$ at each location*

	Threshold	Extremal-t	Extremal skew-t	Empirical (95% confidence intervals)
$X \mid Y, Z$	$(q^{90}_{\mathrm{CO}}, q^{70}_{\mathrm{CA}}, q^{70}_{\mathrm{PA}})$	0.2587	0.2737	0.3333 (0.2706, 0.3960)
	$(q^{90}_{\mathrm{SA}}, q^{70}_{\mathrm{CA}}, q^{70}_{\mathrm{PA}})$	0.3268	0.3305	0.2973 (0.2356, 0.3590)
	$(q^{90}_{\mathrm{PA}}, q^{70}_{\mathrm{CA}}, q^{70}_{\mathrm{SA}})$	0.3752	0.3356	0.2857 (0.2247, 0.3467)
	$(q^{90}_{\mathrm{CO}}, q^{70}_{\mathrm{SA}}, q^{70}_{\mathrm{PA}})$	0.2686	0.3150	0.3333 (0.2706, 0.3960)
$X, Y \mid Z$	$(q^{90}_{\mathrm{CO}}, q^{90}_{\mathrm{CA}}, q^{70}_{\mathrm{SA}})$	0.1196	0.0789	0.0781 (0.0420, 0.1142)
	$(q^{90}_{\mathrm{CA}}, q^{90}_{\mathrm{PA}}, q^{70}_{\mathrm{CO}})$	0.1236	0.0776	0.0938 (0.0546, 0.1330)
	$(q^{90}_{\mathrm{CO}}, q^{90}_{\mathrm{SA}}, q^{70}_{\mathrm{PA}})$	0.0896	0.1048	0.0938 (0.0550, 0.1326)
	$(q^{90}_{\mathrm{SA}}, q^{90}_{\mathrm{PA}}, q^{70}_{\mathrm{CO}})$	0.1038	0.1071	0.0769 (0.0415, 0.1123)

extremal skew-t model achieves a lower CLIC score than the extremal-t model, indicating its greater suitability. Moreover, the standard errors of the estimated slant parameters $\hat{\alpha}$ clearly indicate that these parameters are non-zero, strengthening the argument of a significantly better fit from the extremal skew-t model.

For each location triple (X, Y, Z), we can also evaluate the conditional probability of exceeding some fixed threshold (x, y, z) using each parametric model. Table 2 presents estimated probabilities of the two cases $\Pr(X > x \mid Y > y, Z > z)$ and $\Pr(X > x, Y > y \mid Z > z)$, along with the associated empirical probabilities and their 95% confidence intervals (CI) for a range of thresholds. For these specific thresholds, the extremal skew-t model provides estimates of the conditional probabilities that fall within the 95% empirical CI. However, four probabilities estimated with the extremal-t model are not consistent with the empirical CI. This indicates that the additional flexibility of the extremal skew-t model allows it to more accurately characterize the dependence structure evident in the observed data.

Finally, Fig. 6 provides examples of univariate (top panels) and bivariate (bottom) conditional return levels for each triple of sites. The return levels are computed conditionally on the wind at the remaining station(s) being higher than their upper 70% marginal quantile. For the univariate conditional return levels (top panels), both the extremal-t and extremal skew-t model fits are strongly influenced by the windspeed outlier of $\sim$ 40 km/h observed at CLAY station (centre two panels). This phenomenon, whereby the far tails of extremal model fits can be dominated by a single extreme outlier, is not uncommon in practice (e.g. Coles *et al.*, 2003). Being the more flexible model, the extremal skew-t model is better able to follow this extreme outlier compared with the extremal t. When the outlier is not present (in the two outer panels), the extremal skew-t model provides a better visual fit to the observed data and spends more time within the empirical CI, indicating a superior model fit.

The primary differences in the bivariate conditional return levels (bottom panels, Fig. 6) are the possibility of asymmetric contour levels under the extremal skew-t model (dot-dashed blue line) in contrast with symmetric contours under the extremal-t model (dotted red line). The difference is more noticeable in the leftmost and rightmost panel. The leftmost panel indicates lower return levels for the extremal skew-t model, which occurs because (CLOU, SALL) have negative slant parameters (Table 1, top row), and so the joint tail is shorter than that of the extremal t. Conversely, the rightmost panel exhibits larger return levels for the extremal skew-t model, as a result of the small negative and very large slant parameters for (CLOU, PAUL) (Table 1, bottom row), and so the joint tail is longer than that of the extremal-t. The differences

© 2016 Board of the Foundation of the Scandinavian Journal of Statistics.

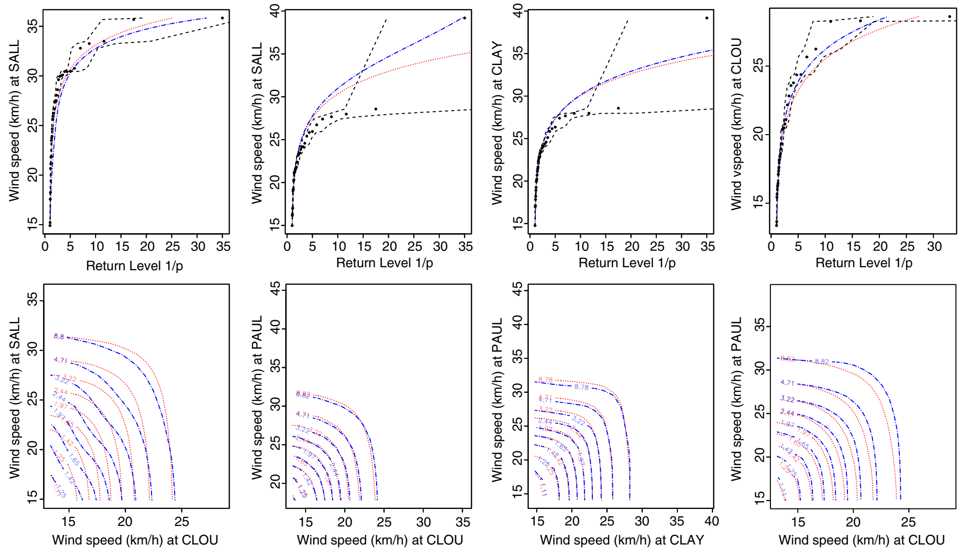

Fig. 6. Univariate (top row) and bivariate (bottom) conditional return levels for the triples (left to right): (CLOU, CLAY, SALL), (CLOU, CLAY, PAUL), (CLAY, SALL, PAUL) and (CLOU, SALL, PAUL). Dotted red and dot-dashed blue lines, respectively, indicate return levels calculated from extremal-*t* and extremal skew-*t* models. Points indicate the empirical observations and the black dashed lines their 95% confidence interval. [Colour figure can be viewed at wileyonlinelibrary.com]

in the centre two panels are less pronounced. For the second panel, the slant parameters of (CLOU, PAUL) similarly take a large positive and a small negative value (Table 1, row 2). However, as the parameter for CLAY is also a large positive value, this means that there is little difference between the joint tails of the two models. Finally, for the third panel, the slant parameters of (CLAY, PAUL, SALL) are relatively small and positive (Table 1, row 3), and so there is little difference between the joint tails of the two models.

In summary, for these wind speed data, the more flexible extremal skew-*t* model is demonstrably superior to the extremal-*t* model in describing the extremes of both the univariate marginal distributions and the extremal dependence between locations.

6. Discussion

Appropriate modelling of extremal dependence is critical for producing realistic and precise estimates of future extreme events. In practice, this is a hugely challenging task, as extremes in different application areas may exhibit different types of dependence structures, asymptotic dependence levels, exchangeability and stationary or non-stationary behaviour.

Working with families of skew-normal distributions and processes, we have derived flexible new classes of extremal dependence models. Their flexibility arises as they include a wide range of dependence structures, while also incorporating several previously developed and popular models, such as the stationary extremal-*t* process and its sub-processes, as special cases. These include dependence structures that are asymptotically independent, which is useful for describing the dependence of variables that are not exchangeable, and a wide class of non-stationary, asymptotically dependent models, suitable for the modelling of spatial extremes.

In terms of future development, semi-parametric estimation methods would provide powerful techniques to fully take advantage of the flexibility offered by non-stationary max-stable models. Such methods can be computationally demanding, however. An interesting further

© 2016 Board of the Foundation of the Scandinavian Journal of Statistics.

14679469, 2017, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.12240 by University of New South Wales, Wiley Online Library on [04/08/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

direction would be to design simple and interpretable families of covariance functions for skew-normal processes for which it is then possible to derive max-stable dependence models that are useful in practical applications.

The code used to perform the simulations studies and real data analysis in Sections 4 and 5 as well as in the Supporting Information is available in the `R` (R Core Team, 2014) package `ExtremalDep` (Beranger *et al.*, 2015) available at `https://r-forge.r-project.org/R/?group_id=1998`.

Acknowledgements

We would like to thank the referees, associate editor and editor for useful comments that led to improved presentation of the material.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web site.

References

Arellano-Valle, R. B. & Azzalini, A. (2006). On the unification of families of skew-normal distributions. *Scand. J. Statist.* **33**, 561–574.

Arellano-Valle, R. B. & Genton, M.G. (2010). Multivariate extended skew-*t* distributions and related families. *Metron.* **68**, (3), 201–234.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statist.* **12**, 171–178.

Azzalini, A. (2005). The skew-normal distribution and related multivariate families. *Scand. J. Statist.* **32**, (2), 159–200. With discussion by Marc G. Genton and a rejoinder by the author.

Azzalini, A. & Captianio A. (2014). *The skew normal and related families, IMS Monographs Series*, Cambridge University Press.

Beranger, B., Marcon, G. & Padoan, S. (2015). *Extremaldep: Extremal dependence modeling*. R package version 0.1-2/r76.

Bortot, P. (2010). *Tail dependence in bivariate skew-normal and skew-t distributions*. Unpublished manuscript.

Brown, B. M. & Resnick, S. I. (1977). Extreme values of independent stochastic processes. *J. Appl. Probab.* **14**, 732–739.

Chan, G. & Wood, A. T. (1997). Algorithm AS 312: An algorithm for simulating stationary Gaussian random fields. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **46**, 171–181.

Chang, S.-M. & Genton, M. G. (2007). Extreme value distributions for the skew-symmetric family of distributions. *Comm. Statist. Theory Methods* **36**, (9), 1705–1717.

Coles, S. G., Pericchi, L. R. & Sisson, S. A. (2003). A fully probabilistic approach to extreme value modelling. *Journal of Hydrology* **273**, 35–50.

Coles, S. G. & Tawn, J. A. (1991). Modelling extreme multivariate events. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **53**, (2), 377–392.

Coles, S. G. & Tawn, J. A. (1994). Statistical methods for multivariate extremes: An application to structural design. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **43**, (1), 1–48.

Davison, A. C. & Gholamrezaee, M. M. (2012). Geostatistics of extremes. *Proceedings of the Royal Society of London Series A: Mathematical and Physical Sciences* **468**, 581–608.

Davison, A. C., Padoan, S. A. & Ribatet, M. (2012). Statistical modeling of spatial extremes. *Statist. Sci.* **27**, 161–186.

Dutt, J. E. (1973). A representation of multivariate normal probability integrals by integral transforms. *Biometrika* **60**, (3), 637–645.

Feller, W. (1968). *An introduction to probability theory and its applications*, Vol. I, John Wiley & Sons, London-New York-Sydney-Toronto.

Genton, M. (2004). *Skew-elliptical distributions and their applications*, Chapman & Hall/CRC, Boca Raton, FL.

de Haan, L. (1984). A spectral representation for max-stable processes. *Ann. Appl. Probab.* **12**, 1194–1204.

© 2016 Board of the Foundation of the Scandinavian Journal of Statistics.

14679469, 2017, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.12240 by University of New South Wales, Wiley Online Library on [04/08/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

de Haan, L. & Ferreira, A. (2006). *Extreme value theory*, Springer Series in Operations Research and Financial Engineering, Springer, New York.
Huser, R. & Davison, A. C. (2013). Composite likelihood estimation for the Brown–Resnick process. *Biometrika* **100**, 511–518.
Huser, R. & Genton, M. (2015). *Non-stationary dependence structures for spatial extremes*. arXiv:1411.3174v1.
Jamalizadeh, A., Mehrali, Y. & Balakrishnan, N. (2009). Recurrence relations for bivariate t and extended skew-t distributions and an application to order statistics from bivariate t. *Comput. Statist. Data Anal.* **53**, (12), 4018–4027.
Joe, H. (1997). *Multivariate models and dependence concepts*, Monographs on Statistics and Applied Probability, vol. 73, Chapman & Hall, London.
Kabluchko, Z., Schlather, M. & de Haan, L. (2009). Stationary max-stable fields associated to negative definite functions. *Ann. Appl. Probab.* **37**, 2042–2065.
Ledford, A. W. & Tawn, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika* **83**, (1), 169–187.
Lindgren, G. (2012). *Stationary stochastic processes: Theory and applications*, CRC Press.
Lysenko, N., Roy, P. & Waeber, R. (2009). Multivariate extremes of generalized skew-normal distributions. *Statist. Probab. Lett.* **79**, (4), 525–533.
Minozzo, M. & Ferracuti, L. (2012). On the existence of some skew-normal stationary processes. *Chil. J. Stat.* **3**, 157–170.
Nikoloulopoulos, A. K., Joe, H. & Li, H. (2009). Extreme value properties of multivariate t copulas. *Extremes* **12**, (2), 129–148.
Opitz, T. (2013). Extremal t processes: Elliptical domain of attraction and a spectral representation. *J. Multivariate Anal.* **122**, (0), 409 –413.
Padoan, S. A. (2011). Multivariate extreme models based on underlying skew-t and skew-normal distributions. *J. Multivariate Anal.* **102**, (5), 977 –991.
Padoan, S. A., Ribatet, M. & Sisson, S. A. (2010). Likelihood-based inference for max-stable processes. *J. Amer. Statist. Assoc.* **105**, 263–277.
R Core Team. (2014). *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria.
Schlather, M. (2002). Models for stationary max-stable random fields. *Extremes* **5**, 33–44.
Smith, R. L. (1990). Max-stable processes and spatial extremes University of Surrey 1990 technical report.
Varin, C., Reid, N. & Firth, D. (2011). An overview of composite likelihood methods. *Statist. Sinica* **21**, (1), 5–42.
Wood, A. T. A. (1994). Simulation of stationary Gaussian processes in $[0;1]^d$. *J. Comput. Graph. Statist.* **3**, 409–432.
Zhang, H. & El-Shaarawi, A. (2010). On spatial skew-Gaussian processes and applications. *Environmetrics* **21**, (1), 33–47.

Received June 2015, in final form March 2016

Simone A. Padoan, Department of Decision Sciences, Bocconi University, via Roentgen, 1, 20136, Milan, Italy.
E-mail: simone.padoan@unibocconi.it

Appendix A: Proofs

A.1. Proof of Proposition 1

Items (1)–(3) are easily derived following the proof of Propositions (1)–(4) of Arellano-Valle & Genton (2010) and taking into account the next result.

Lemma 1. *Let* $Y = (Y_1^\top, Y_2^\top)^\top \sim \mathcal{T}_d(\mu, \Omega, \kappa, \nu)$*, where* $Y_1 \in \mathbb{R}$ *and* $Y_2 \in \mathbb{R}^{d-1}$ *with the corresponding partition of the parameters* (μ, Ω, ν) *and* $\kappa = (\kappa_1, 0^\top)^\top$ *with* $\kappa_1 \in \mathbb{R}$*. Then,*

$$(Y_1 \mid Y_2 = y_2) \sim \mathcal{T}(\mu_{1\cdot 2}, \Omega_{11\cdot 2}, \kappa_{1\cdot 2}, \nu_{1\cdot 2}), \qquad y_2 \in \mathbb{R}^{d-1}$$

© 2016 Board of the Foundation of the Scandinavian Journal of Statistics.

14679469, 2017, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.12240 by University of New South Wales, Wiley Online Library on [04/08/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

where $\mu_{1\cdot 2} = \mu_1 + \Omega_{12}\Omega_{22}^{-1}(y_2 - \mu_2)$, $\Omega_{1\cdot 2} = \zeta_2\Omega_{11\cdot 2}$, $\zeta_2 = \{\nu + Q_{\Omega_{22}^{-1}}(z_2)\}/(\nu + d_2)$, $z_2 = \omega_2^{-1}(y_2 - \mu_2)/\Omega_2$, $\omega_2 = diag(\Omega_{22})^{1/2}$, $\Omega_{11\cdot 2} = \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}$, $\kappa_{1\cdot 2} = \zeta_2^{-1/2}\kappa$, $\nu_{1\cdot 2} = \nu + d - 1$.

Proof of Lemma 1. The marginal density of Y_2 is equal to

$$f_{Y_2}(y_2) = \int_0^\infty \frac{v^{\nu/2-1}e^{-v}}{(\nu/2)}\phi_{d-1}\left(\frac{y_2-\mu_2}{\sqrt{\frac{\nu}{2v}}};\Omega_{22}\right)\left(\frac{2v}{\nu}\right)^{(d-1)/2}\mathrm{d}v = \psi_{d-1}(y_2;\mu_2,\Omega_{22},\nu),$$

namely, it is a $(d-1)$-dimensional central t pdf. The joint density of Y is equal to

$$\begin{aligned} f_{Y_2}(y_2)f_{Y_1|Y_2=y_2}(y_1) = \psi_{d-1}(y_2;\mu_2,\Omega_{22},\nu)&\int_0^\infty \frac{v^{(\nu+d-1)/2-1}e^{-v}}{\left(\frac{\nu+d-1}{2}\right)} \\ &\phi\left\{(\Omega_{1\cdot 2})^{-1/2}(y_1-\mu_{1\cdot 2})\sqrt{\frac{2v}{\nu+d-1}} - (\Omega_{11\cdot 2})^{-1/2}\kappa_1\right\}\mathrm{d}v \\ &= \int_0^\infty \frac{(\Omega_{11\cdot 2})^{-1/2}v^{\nu/2-1}e^{-v}}{\left(\frac{\nu}{2}\right)}\left(\frac{2v}{\nu}\right)^{d/2}\phi_{d-1}\left(\frac{y_2-\mu_2}{\sqrt{\frac{\nu}{2v}}}\right) \\ &\phi\left\{(\Omega_{11\cdot 2})^{1/2}(y_1-\mu_{1\cdot 2})\sqrt{\frac{2v}{\nu}} - \kappa_1\right\}\mathrm{d}v \\ &= \int_0^\infty \frac{v^{\nu/2-1}e^{-v}}{\left(\frac{\nu}{2}\right)}\phi_d\left\{\begin{pmatrix} y_1-\mu_1-\kappa_1\sqrt{\frac{\nu}{2v}} \\ y_2-\mu_2\end{pmatrix};\sqrt{\frac{\nu}{2v}}\Omega\right\}\mathrm{d}v. \end{aligned}$$

□

A.2. Proof of Proposition 2

Let $Z \sim \mathcal{ST}(\alpha,\tau,\kappa,\nu)$. Then $1 - \Psi(x;\alpha,\tau,\nu) \approx x^{-\nu}\mathscr{L}(x;\alpha,\tau,\nu)$ as $x \to +\infty$, for any $\nu > 1$, where

$$\mathscr{L}(x;\alpha,\tau,\kappa,\nu) = \frac{\{(\nu+1)/2\}\Psi(\alpha\sqrt{\nu+1};\nu+1)}{(\nu/2)\sqrt{\pi}\nu^{3/2}\Psi(\tau/\sqrt{1+\alpha^2};\kappa/\sqrt{1+\alpha^2},\nu)}\left(\frac{1}{x^2}+\frac{1}{\nu}\right)^{-(\nu+1)/2}$$

is a slowly varying function (e.g. de Haan & Ferreira, 2006, Appendix B). From Corollary 1.2.4 in de Haan & Ferreira (2006), it follows that the normalization constants are $a_n = \Psi^{\leftarrow}(1-1/n;\alpha,\tau,\kappa,\nu)$, where $\Psi^{\leftarrow}$ is the inverse function of Ψ and $b_n = 0$, and therefore $a_n = \{n\mathscr{L}(\alpha,\tau,\kappa,\nu)\}^{1/\nu}$, where $\mathscr{L}(\alpha,\tau,\kappa,\nu) \equiv \mathscr{L}(\infty;\alpha,\tau,\kappa,\nu)$. Applying Theorem 1.2.1 in de Haan & Ferreira (2006), we obtain that $M_n/a_n \Rightarrow U$, where U has ν-Fréchet univariate marginal distributions.

Let $Z \sim \mathcal{ST}_d(\bar{\Omega},\alpha,\tau,\kappa,\nu)$. For any $j \in \{1,\ldots,d\}$, consider the partition $Z = (Z_j, Z_{I_j}^\top)^\top$, where $I_j = \{1,\ldots,d\}\backslash j$ and $Z_j = Z_{\{j\}}$ and the respective partition of $(\bar{\Omega},\alpha)$. Define $a_n = (a_{n,1},\ldots,a_{n,d})$, where $a_{n,j} = \{n\mathscr{L}(\alpha_j^*,\tau_j^*,\kappa_j^*,\nu)\}^{1/\nu}$ and $\alpha_j^* = \alpha_{\{j\}}^*$, $\tau_j^* = \tau_{\{j\}}^*$ and $\kappa_j^* = \kappa_{\{j\}}^*$ are the marginal parameters (5) under Proposition 1 (1). From Theorem 6.1.1 and Corollary 6.1.3 in de Haan & Ferreira (2006), $M_n/a_n \Rightarrow U$, where the distribution of U is $G(x) = \exp\{-V(x)\}$ with $V(x) = \lim_{n\to+\infty} n\{1 - \Pr(Z_1 \le a_{n,1}x_1,\ldots,Z_d \le a_{n,d}x_d)\}$ for all $x = (x_1,\ldots,x_d)^\top \in \mathbb{R}_+^d$. Applying the conditional tail dependence function framework of Nikoloulopoulos *et al.* (2009), it follows that

© 2016 Board of the Foundation of the Scandinavian Journal of Statistics.

14679469, 2017, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.12240 by University of New South Wales, Wiley Online Library on [04/08/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$V(x_j, i \in I) = \lim_{n\to\infty} \sum_{j=1}^{d} x_j^{-\nu} \Pr(Z_i \le a_{n,i} x_i, i \in I_j \mid Z_j = a_{n,j} x_j).$$

From the conditional distribution in Proposition 1 (1), we have that

$$\left\{ \left(\frac{Z_i - a_{n,j} x_j}{\{\zeta_{n,j}(1-\omega_{i,j}^2)\}^{1/2}}, i \in I_j \right)^{\top} \mid Z_j = a_{n,j} x_j \right\} \sim \mathcal{ST}_{d-1}\left(\bar{\Omega}_j^{+}, \alpha_j^{+}, \tau_{n,j}, \kappa_{n,j}, \nu+1\right),$$

for $j \in \ldots 1, \ldots, d$, where $\bar{\Omega}_j^{+} = \omega_{I_j I_j \cdot j}^{-1} \Omega_{I_j I_j \cdot j} \omega_{I_j I_j \cdot j}^{-1}$, $\omega_{I_j I_j \cdot j} = \mathrm{diag}(\Omega_{I_j I_j \cdot j})^{1/2}$, $\bar{\Omega}_{I_j I_j \cdot j} = \bar{\Omega}_{I_j I_j} - \bar{\Omega}_{I_j j} \bar{\Omega}_{j I_j}$, $\alpha_j^{+} = \bar{\Omega}_{I_j I_j \cdot j} \alpha_{I_j}$ $\zeta_{n,j} = [\nu + (a_{n,j} x_j)^2]/(\nu+1)$, $\tau_{n,j} = [(\bar{\Omega}_{j I_j} \alpha_{I_j} + \alpha_j) a_{n,j} x_j + \tau]/\zeta_{n,j}^{1/2}$ and $\kappa_{n,j} = \kappa/\zeta_{n,j}^{1/2}$. Now, for any $j \in \{1, \ldots, d\}$ and all $i \in I_j$

$$\frac{a_{n,i} x_i - a_{n,j} x_j}{\left\{\zeta_{n,j}\left(1-\omega_{i,j}^2\right)\right\}^{1/2}} \to \frac{\left(x_i^{+}/x_j^{+} - \omega_{i,j}\right)(\nu+1)^{1/2}}{\{(1-\omega_{i,j})\}^{1/2}} \quad \text{as } n \to +\infty,$$

where $\omega_{i,j}$ is the (i,j)-th element of $\bar{\Omega}$, $x_j^{+} = x_j \mathscr{L}^{1/\nu}(\alpha_j^*, \tau_j^*, \kappa_j^*, \nu)$ and $\tau_{n,j} \to \tau_j^{+} = (\bar{\Omega}_{j I_j} \alpha_{I_j} + \alpha_j)(\nu+1)^{1/2}$, and $\kappa_{n,j} \to 0$ as $n \to +\infty$. As a consequence

$$V(x_j, j \in I) = \sum_{j=1}^{d} x_j^{-\nu} \Psi_{d-1}\left(\left(\sqrt{\frac{\nu+1}{1-\omega_{i,j}^2}} \left(\frac{x_i^{+}}{x_j^{+}} - \omega_{i,j} \right), i \in I_j \right)^{\top}; \bar{\Omega}_j^{+}, \alpha_j^{+}, \tau_j^{+}, \nu+1 \right).$$

A.3. Proof of Proposition 4

Recall that if $Z \sim \mathcal{SN}_2(\bar{\Omega}, \alpha)$, then $Z_j \sim \mathcal{SN}(\alpha_j^*)$ and $Z_j \mid Z_{3-j} \sim \mathcal{SN}(\alpha_{j\cdot 3-j})$ for $j = 1, 2$ (e.g. Azzalini, 2014, Ch. 2, or Proposition 1), where

$$\alpha_j^* = \frac{\alpha_j + \omega\alpha_{3-j}}{\sqrt{1+\alpha_{3-j}^2(1-\omega^2)}}, \quad \alpha_{j\cdot 3-j} = \alpha_j\sqrt{1-\omega^2}.$$

Define $x_j(u) = \Phi^{\leftarrow}(1-u; \alpha_j^*)$, for any $u \in [0,1]$, where $\Phi^{\leftarrow}(\cdot; \alpha_j^*)$ is the inverse of the marginal distribution function $\Phi(\cdot; \alpha_j^*)$, $j = 1, 2$. The asymptotic behaviour of $x_j(u)$ as $u \to 0$ is

$$x_j(u) = \begin{cases} x(u), & \text{if}\, \alpha_j^* \ge 0 \\ x(u)/\bar{\alpha}_j - \{2\log(1/u)\}^{-1/2} \log\left(\sqrt{\pi}\alpha_j^*\right), & \text{if}\, \alpha_j^* < 0 \end{cases} \tag{23}$$

for $j = 1, 2$, where $\bar{\alpha}_j = \{1 + \alpha_j^{*2}\}^{1/2}$ and $x(u) \approx \{2\log(1/u)\}^{1/2} - \{2\log(1/u)\}^{-1/2}\{\log\log(1/u) + \log(2\sqrt{\pi})\}$ (Padoan, 2011). The limiting behaviour of the joint survivor function of the bivariate skew-normal distribution is described by

$$p(u) = \Pr\{Z_1 > x_1(u), Z_2 > x_2(u)\}, \qquad u \to 0. \tag{24}$$

For case (a), when $\alpha_1, \alpha_2 > 0$, then $x_1(u) = x_2(u) = x(u)$, and the joint upper tail (24) behaves as

© 2016 Board of the Foundation of the Scandinavian Journal of Statistics.

14679469, 2017, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.12240 by University of New South Wales, Wiley Online Library on [04/08/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$
\begin{aligned}
p(u) &= \int_{x(u)}^{\infty} \left\{1 - \Phi\left(\frac{y(u) - \omega v}{\sqrt{1-\omega^2}}; \alpha_{1\cdot 2}\right)\right\} \phi(v; \alpha_2^*) \mathrm{d}v \\
&\approx \frac{\sqrt{1-\omega^2}}{x(u)} \int_0^{\infty} \frac{\phi_2(x(u), x(u) + t/x(u); \bar{\Omega}, \alpha)}{x(u)(1-\omega) - \omega t/x(u)} \mathrm{d}t \\
&\approx \frac{e^{-x^2(u)/(1+\omega)}}{\pi(1-\omega)x^2(u)} \left(\int_0^{\infty} e^{-t/(1+\omega)} \mathrm{d}t - \frac{e^{-x^2(u)(\alpha_1+\alpha_2)^2/2}}{\sqrt{2\pi}(\alpha_1+\alpha_2)x(u)} \int_0^{\infty} e^{-t\{1/(1+\omega)+\alpha_2(\alpha_1+\alpha_2)\}} \mathrm{d}t\right) \\
&= \frac{e^{-x^2(u)/(1+\omega)}(1+\omega)}{\pi(1-\omega)x(u)^2} \left(1 - \frac{e^{-x^2(u)(\alpha_1+\alpha_2)^2/2}}{\sqrt{2\pi}(\alpha_1+\alpha_2)\{1+\alpha_2(\alpha_1+\alpha_2)(1+\omega)\}x(u)}\right),
\end{aligned}
\tag{25}
$$

as $u \to 0$. The first approximation is obtained by using $1 - \Phi(x;\alpha) \approx \phi(x;\alpha)/x$ as $x \to +\infty$, when $\alpha > 0$ (Padoan, 2011). The second approximation uses $1 - \Phi(x) \approx \phi(x)/x$ as $x \to +\infty$ (Feller, 1968). Let $X_j = \{-1/\log\Phi(Z_j;\alpha_j^*)\}$, $j = 1,2$. Substituting $x(u)$ into (25), substituting and using the approximation $1-\Pr(X_j > x) \approx 1/x$ as $x \to \infty$, $j = 1,2$, we obtain that (24) with common unit Fréchet margins behaves asymptotically as $\mathscr{L}(x)\ x^{-2/(1+\omega)}$, as $x \to +\infty$, where

$$
\mathscr{L}(x) = \frac{2(1+\omega)(4\pi\log x)^{-\omega/(1+\omega)}}{1-\omega} \left(1 - \frac{(4\pi\log x)^{\{(\alpha_1+\alpha_2)^2-1\}/2}\, x^{-(\alpha_1+\alpha_2)^2}}{(\alpha_1+\alpha_2)\{1+\alpha_2(\alpha_1+\alpha_2)(1+\omega)\}}\right).
\tag{26}
$$

As the second term in the parentheses in (26) is $o(x^{(\alpha_1+\alpha_2)})$, then the quantity inside the parentheses $\to 1$ rapidly as $x \to \infty$, and so $\mathscr{L}(x)$ is well approximated by the first term in (26). When $\alpha_2 < 0$ and $\alpha_1 \geq -\alpha_2/\omega$, then $\alpha_1^*, \alpha_2^* > 0$, and we obtain the same outcome.

For case (b), when $\alpha_2 < 0$ and $-\omega, \alpha_2 \leq \alpha_1 < -\omega^{-1}\alpha_2$, then $\alpha_1^* \geq 0$ and $\alpha_2^* < 0$, and hence $x_1(u) = x(u)$ and $x_2(u) \approx x(u)/\bar{\alpha}_2$ as $u \to 0$. When $\alpha_1 > -\bar{\alpha}_2\alpha_2$, then following a similar derivation to those in (25), we obtain that

$$
p(u) \approx \frac{\bar{\alpha}_2^2(1-\omega^2)(1-\omega\bar{\alpha}_2)^{-1}}{\pi(\bar{\alpha}_2-\omega)x^2(u)} \exp\left[-\frac{x^2(u)}{2}\left\{\frac{1-\omega^2+(\bar{\alpha}_2-\omega)^2}{(1-\omega^2)\bar{\alpha}_2^2}\right\}\right], \quad \text{as } u \to 0.
$$

Similarly, when $\alpha_1 < -\bar{\alpha}_2\alpha_2$, and noting that $\Phi(x) \approx -\phi(-x)/x$ as $x \to -\infty$, then

$$
p(u) \approx \frac{-\bar{\alpha}_2^2\{1-\omega\bar{\alpha}_2+\alpha_2(\alpha_2+\alpha_1\bar{\alpha}_2)(1-\omega^2)\}^{-1}}{\pi(\bar{\alpha}_2-\omega)(1-\omega^2)^{-1}(\alpha_1+\alpha_2/\bar{\alpha}_2)x^3(u)} e^{-\frac{x^2(u)}{2}\left\{\frac{1-\omega^2+(\bar{\alpha}_2-\omega)^2}{(1-\omega^2)\bar{\alpha}_2^2}+\left(\alpha_1+\frac{\alpha_2}{\bar{\alpha}_2}\right)^2\right\}}, \text{ as } \boldsymbol{u \to 0}.
$$

For case (c), when $\alpha_2 < 0$ and $0 < \alpha_1 < -\omega\alpha_2$, then $\alpha_1^*, \alpha_2^* < 0$, and hence $x_1(u) \approx x(u)/\bar{\alpha}_1$ and $x_2(u) \approx x(u)/\bar{\alpha}_2$ as $u \to 0$. Then, as $u \to 0$, we have

$$
\begin{aligned}
p(u) \approx{}& \frac{-\bar{\alpha}_2^{3/2}\bar{\alpha}_1^2(1-\omega^2)(\bar{\alpha}_2-\omega\bar{\alpha}_1)^{-1}(\alpha_1\bar{\alpha}_2+\alpha_2\bar{\alpha}_1)^{-1}}{\pi\{1-\omega\bar{\alpha}_2+\alpha_2(\alpha_2+\alpha_1\bar{\alpha}_2/\bar{\alpha}_1)(1-\omega^2)\}x^3(u)} \\
&\times \exp\left[-\frac{x^2(u)}{2(1-\omega^2)}\left(\frac{\alpha_1^2(1-\omega^2)+1}{\bar{\alpha}_1^2}+\frac{\alpha_2^2(1-\omega^2)+1}{\bar{\alpha}_2^2}+\frac{2(\alpha_1\alpha_2(1-\omega^2)-\omega)}{\bar{\alpha}_1\bar{\alpha}_2}\right)\right] \quad \boldsymbol{u \to 0}.
\end{aligned}
$$

When $\alpha_1, \alpha_2 < 0$ and $\omega_2^{-1}\alpha_2 \leq \alpha_1 < 0$, the same argument holds. Finally, interchanging α_1 with α_2 produces the same results, but substituting α_j and $\bar{\alpha}_j$ with α_{3-j} and $\bar{\alpha}_{3-j}$, respectively, for $j = 1,2$.

© 2016 Board of the Foundation of the Scandinavian Journal of Statistics.

14679469, 2017, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.12240 by University of New South Wales, Wiley Online Library on [04/08/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

A.4. Proof of Theorem 1

Let $Y(s)$ be a skew-normal process with finite-dimensional distribution $\mathcal{SN}_d(\bar{\Omega}, \alpha, \tau)$. For any $j \in I = \{1, \ldots, d\}$, consider the partition $Y = (Y_j, Y_{I_j}^\top)^\top$, where $I_j = I \backslash j$, $Y_j = Y_{\{j\}} = Y(s_j)$ and $Y_{I_j} = (Y_i, i \in I_j)^\top$ and the respective partition of $(\bar{\Omega}, \alpha)$. The exponent function (14) is

$$V(x_j, j \in I) = \mathbb{E}\left[\max_j \left\{\frac{(Y_j^+/x_j)^\xi}{m_j^+}\right\}\right] = \int_{\mathbb{R}^d} \max_j \left\{\frac{(y_j/x_j)^\xi}{m_j^+}, 0\right\} \phi_d(y; \bar{\Omega}; \alpha, \tau) \mathrm{d}y,$$

where $x_j \equiv x(s_j)$, $y_j \equiv y(s_j)$ and $m_j^+ \equiv m^+(s_j)$. Then

$$V(x_j, j \in I) = \sum_{j=1}^d V_j, \quad V_j = \frac{1}{m_j^+} \int_0^\infty \left(\frac{y_j}{x_j}\right)^\nu \int_{-\infty}^{y_j x_{I_j}/x_j} \phi_d(y; \bar{\Omega}; \alpha, \tau) \mathrm{d}y_{I_j} \mathrm{d}y_j, \tag{27}$$

where $x_{I_j} = (x_i, i \in I_j)^\top$ and $y_{I_j} = (y_i, i \in I_j)^\top$. As $Y_j \sim \mathcal{SN}(\alpha_j^*, \tau_j^*)$, where $\alpha_j^* = \alpha_{\{j\}}^*$ and $\tau_j^* = \tau_{\{j\}}^*$ are the marginal parameters derived from Proposition 1(1), then

$$\begin{aligned} m_j^+ = \int_0^\infty y_j^\nu \phi\left(y_j; \alpha_j^*, \tau_j^*\right) \mathrm{d}y_j &= \frac{1}{\Phi\left\{\tau_j^* \left(1 + \alpha_j^{*2}\right)^{-1/2}\right\}} \int_0^\infty y_j^\nu \phi(y_j) \Phi\left(\alpha_j^* y_j + \tau_j^*\right) \mathrm{d}y_j \\ &= \frac{2^{(\nu-2)/2}\{(\nu+1)/2\} \Psi\left(\alpha_j^* \sqrt{\nu+1}; -\tau_j^*, \nu+1\right)}{\sqrt{\pi}\Phi[\tau\{1 + Q_{\bar{\Omega}}(\alpha)\}^{-1/2}]} \end{aligned}$$

by observing that $\tau_j^* \{1 + \alpha_j^{*2}\}^{1/2} = \tau\{1 + Q_{\bar{\Omega}}(\alpha)\}^{-1/2}$.

For $j = 1, \ldots, d$, define $x_j^\circ = x_j (m_j^+)^{1/\nu}$ and $m_j^+ = \bar{m}_j^+ / \Phi[\tau\{1 + Q_{\bar{\Omega}}(\alpha)\}^{-1/2}]$, where $\bar{m}_j^+ = (\pi)^{1/2} 2^{(\nu-2)/2} \{(\nu+1)/2\} \Psi(\alpha_j^* \sqrt{\nu+1}; -\tau_j^*, \nu+1)$. Then, for any $j = 1, \ldots, d$

$$\begin{aligned} V_j &= \frac{1}{m_j^+} \int_0^\infty \left(\frac{y_j}{x_j}\right)^\nu \int_{-\infty}^{y_j x_{I_j}/x_j} \phi_d\left(y; \bar{\Omega}, \alpha, \tau\right) \mathrm{d}y_{I_j} \mathrm{d}y_j \\ &= \frac{1}{\bar{m}_j^+} \int_0^\infty \left(\frac{y_j}{x_j}\right)^\nu \int_{-\infty}^{y_j x_{I_j}/x_j} \phi_d\left(y; \Omega\right) \Phi\left(\alpha^\top y + \tau\right) \mathrm{d}y_{I_j} \mathrm{d}y_j \\ &= \frac{1}{\bar{m}_j^+} \int_0^\infty \left(\frac{y_j}{x_j}\right)^\nu \phi(y_j) \int_{-\infty}^{y_j x_{I_j}/x_j} \phi_{d-1}\left(y_{I_j} - y_j \bar{\Omega}_{j, I_j}; \bar{\Omega}_j^\circ\right) \Phi\left(\alpha^\top y + \tau\right) \mathrm{d}y_{I_j} \mathrm{d}y_j \\ &= \frac{1}{\bar{m}_j^+} \int_0^\infty \left(\frac{y_j}{x_j}\right)^\nu \phi(y_j) \Phi_d\left(y_j^\circ; \Omega_j^{\circ\circ}\right) \mathrm{d}y_j, \end{aligned}$$

where

$$y_j^\circ = \begin{pmatrix} y_j\, \omega_{I_j I_j \cdot j}^{-1} \left(x_{I_j}^\circ / x_j^\circ - \bar{\Omega}_{I_j j}\right) \\ y_j \alpha_j^* + \tau_j^* \end{pmatrix},$$

© 2016 Board of the Foundation of the Scandinavian Journal of Statistics.

14679469, 2017, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.12240 by University of New South Wales, Wiley Online Library on [04/08/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

with $\omega_{I_j I_j \cdot j} = \text{diag}(\bar{\Omega}_{I_j I_j \cdot j})^{1/2}$, $\bar{\Omega}_{I_j I_j \cdot j} = \bar{\Omega}_{I_j I_j} - \bar{\Omega}_{I_j j}\bar{\Omega}_{j I_j}$, $y_j \alpha_j^* + \tau_j^* = \frac{y_j(\alpha_j + \bar{\Omega}_{jj}^{-1}\bar{\Omega}_{jI_j}\alpha_{I_j}) + \tau}{\{1+Q_{\bar{\Omega}_{I_j I_j \cdot j}}(\alpha_{I_j})\}^{1/2}}$ and

$$\Omega_j^{\circ\circ} = \begin{pmatrix} \bar{\Omega}_j^{\circ} & -\frac{\bar{\Omega}_{I_j I_j \cdot j}\,\omega_{I_j I_j \cdot j}^{-1}\alpha_{I_j}}{\left\{1+Q_{\bar{\Omega}_{I_j I_j \cdot j}}\left(\alpha_{I_j}\right)\right\}^{1/2}} \\ -\left(\frac{\bar{\Omega}_{I_j I_j \cdot j}\,\omega_{I_j I_j \cdot j}^{-1}\alpha_{I_j}}{\left\{1+Q_{\bar{\Omega}_{I_j I_j \cdot j}}\left(\alpha_{I_j}\right)\right\}^{1/2}}\right)^{\top} & 1 \end{pmatrix},$$

where $\bar{\Omega}_j^{\circ} = \omega_{I_j I_j \cdot j}^{-1}\,\bar{\Omega}_{I_j I_j \cdot j}\,\omega_{I_j I_j \cdot j}^{-1}$ and $\frac{\Omega_{I_j I_j \cdot j}\,\omega_{I_j I_j \cdot j}^{-1}\alpha_{I_j}}{\left\{1+Q_{\Omega_{I_j I_j \cdot j}}(\alpha_{I_j})\right\}^{1/2}} = \frac{\Omega_j^{\circ}\,\omega_{I_j I_j \cdot j}\,\alpha_{I_j}}{\left\{1+Q_{\bar{\Omega}_j^{\circ}}(\omega_{I_j I_j \cdot j}\alpha_{I_j})\right\}^{1/2}}$.

Applying Dutt's (Dutt, 1973) probability integrals, we obtain

$$\begin{aligned} V_j &= \frac{1}{\bar{m}_j^+}\int_0^{\infty}\left(\frac{y_j}{x_j}\right)^{\nu}\phi(y_j)\Phi_d\left(y_j^{\circ};\Omega_j^{\circ\circ}\right)\mathrm{d}y_j, \\ &= \frac{1}{x_j^{\nu}}\,\frac{\Psi_{d+1}\left(\left(\left(\sqrt{\frac{\nu+1}{1-\omega_{i,j}^2}}\left(\frac{x_i^{\circ}}{x_j^{\circ}}-\omega_{i,j}\right), i\in I_j\right), \alpha_j^*\sqrt{\nu+1}\right)^{\top}; \Omega_j^{\circ\circ}, \left(0, -\tau_j^*\right)^{\top}, \nu+1\right)}{\Psi\left(\alpha_j^*\sqrt{\nu+1}; -\tau_j^*, \nu+1\right)}. \end{aligned}$$

This is recognized as the form of a $(d-1)$-dimensional non-central extended skew-t distribution with $\nu+1$ degrees of freedom (Jamalizadeh *et al.*, 2009), from which V_j can be expressed as

$$V_j = \frac{1}{x_j^{\nu}}\Psi_{d-1}\left(\left(\sqrt{\frac{\nu+1}{1-\omega_{i,j}^2}}\left(\frac{x_i^{\circ}}{x_j^{\circ}}-\omega_{i,j}\right), i\in I_j\right)^{\top}; \bar{\Omega}_j^{\circ}, \alpha_j^{\circ}, \tau_j^{\circ}, \kappa_j^{\circ}, \nu+1\right)$$

for $j = 1, \ldots, d$, where $\alpha_j^{\circ} = \omega_{I_j I_j \cdot j}\,\alpha_{I_j}$, $\tau_j^{\circ} = (\bar{\Omega}_{jI_j}\alpha_{I_j} + \alpha_j)(\nu+1)^{1/2}$ and $\kappa_j^{\circ} = -\left\{1 + Q_{\bar{\Omega}_{I_j I_j \cdot j}}(\alpha_{I_j})\right\}^{-1/2}\tau$. Substituting the expression for V_j into (27) then gives the required exponent function.

© 2016 Board of the Foundation of the Scandinavian Journal of Statistics.

14679469, 2017, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/sjos.12240 by University of New South Wales, Wiley Online Library on [04/08/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License