Proofs and additional examples

The EC is organized as follows: In EC.1, we present the proofs of the main results Theorems 1 - 3. EC.2 serves to illustrate the wide applicability of the tail modeling framework with examples and sufficient conditions for Assumptions 2 - 3. Proofs of the results relating to portfolio credit risk applications, namely Proposition 4 and Theorem 6, are presented in Section EC.3. Proofs of technical results (Lemma 1 - 2, Proposition 3 and Theorems 4 - 5) are given in EC.4. EC.5 presents the proof of efficiency when $\boldsymbol{\kappa} = \boldsymbol{\kappa}_2$ is used in Algorithm 1. EC.6 outlines the verification of Assumption 1(b). EC.7 explores how the IS Algorithm 1 with the choice $\boldsymbol{\kappa} = \boldsymbol{\kappa}_1$ is well-suited for use in stochastic optimisation.

EC.1. Proofs of main results

The following definitions and notational convention are used in the proofs: For $r > 0$ and $\boldsymbol{x} \in \mathbb{R}^d$, let $B_r(\boldsymbol{x}) = \{\boldsymbol{y} \in \mathbb{R}^d : \|\boldsymbol{y} - \boldsymbol{x}\| \leq r\}$ denote the metric ball of radius r around $\boldsymbol{x}$. Unless specified explicitly, $\|\boldsymbol{x}\| = \max_{i=1,\ldots,d} |x_i|$ denotes the ℓ_∞-norm. Let $\mathcal{S}^{d-1} = \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\| = 1\}$ denote the unit sphere and $\mathbb{R}^d_{++} = \{\boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{x} > 0\}$ denote the interior of the positive orthant. For $M > 0$, let $B_M := \{\boldsymbol{x} \in \mathbb{R}^d_+ : \|\boldsymbol{x}\| \leq M\}$. For $A \subseteq \mathbb{R}^d$, let $\mathrm{cl}(A)$ denote its closure, $\mathrm{int}(A)$ denote its interior,

$$\chi_A(\boldsymbol{x}) := \begin{cases} 0, & \text{if } \boldsymbol{x} \in A \\ +\infty, & \text{otherwise,} \end{cases} \qquad \text{and} \qquad [A]^{1+r} := \{\boldsymbol{x} + \boldsymbol{y} \in \mathbb{R}^d_+ : \boldsymbol{x} \in A, \|\boldsymbol{y}\| \leq r\},$$

respectively denote the characteristic function and the set of points in $\mathbb{R}^d_+$ lying within a distance $r \in (0, \infty)$ from A. For $f : \mathbb{R}^d \to \mathbb{R}$, $\alpha \in \mathbb{R}$, $M > 0$, let

$$\mathrm{lev}^+_\alpha(f) = \{\boldsymbol{x} \in \mathbb{R}^d_+ : f(\boldsymbol{x}) \geq \alpha\} \quad \text{and} \quad \Xi_{\alpha,M}(f) = \mathrm{lev}^+_\alpha(f) \cap B_M \cap \mathbb{R}^d_{++}$$

denote the super-level sets of f restricted, respectively, to the positive orthant and to the bounded subset $B_M \cap \mathbb{R}^d_{++}$. Recall that $\boldsymbol{Y} = \boldsymbol{\Lambda}(\boldsymbol{X})$, the component-wise inverse $\boldsymbol{q}(t) = (\Lambda_1^{\leftarrow}(t), \ldots, \Lambda_d^{\leftarrow}(t))$, and $\boldsymbol{q}^* = \lim_{t\to\infty}[\boldsymbol{q}(t)/q_\infty(t)]$, if the limit exists. Throughout the proofs, we suppose $u > x_0$ as specified in Assumption 1(a). Define $L_u : \mathbb{R}^d_{++} \to \mathbb{R}$ and $f_{\mathrm{LD}} : \mathbb{R}^d_+ \to \mathbb{R}$ as,

$$L_u(\boldsymbol{x}) := u^{-1} L(\boldsymbol{q}(t(u)\boldsymbol{x})), \qquad f_{\mathrm{LD}}(\boldsymbol{y}) := L^*(\boldsymbol{q}^* \boldsymbol{y}^{\mathbf{1}/\boldsymbol{\alpha}}), \text{ where} \tag{EC.1a}$$

$$t(u) := \Lambda_{\min}\left(u^{1/\rho}\right) \quad \text{and} \quad q_\infty(t) := \|\boldsymbol{q}(t)\|_\infty, \tag{EC.1b}$$

Write $\mathrm{Diag}(\boldsymbol{a})$ for a diagonal matrix with diagonal entries $a_1, \ldots, a_d$ and $\mathrm{sgn}(\boldsymbol{x})$ for the vector of signs of $\boldsymbol{x}$. To avoid clutter in the expressions, define

$$\boldsymbol{Y}_u := t(u)^{-1}\boldsymbol{Y}, \quad \boldsymbol{\psi}_u := \boldsymbol{\Lambda} \circ \boldsymbol{T}^{-1} \circ \boldsymbol{q}, \quad \text{and} \quad c_\rho(u) := (l/u)^{1/\rho}.$$

where the parameter u in the symbol $\boldsymbol{\psi}_u$ is explicitly indicated to remind the role of u in $\boldsymbol{T}^{-1}$.

Proof of Theorem 1. A sufficient condition (see (Dembo & Zeitouni 1998, Theorem 4.1.11)) to verify the existence of LDP is to show that for all $\boldsymbol{x} \in \mathbb{R}^d_{++}$,

$$-I(\boldsymbol{x}) = \inf_{\delta>0} \limsup_{t\to\infty} \frac{1}{t} \log P\left(\frac{\boldsymbol{Y}}{t} \in B_\delta(\boldsymbol{x})\right) = \inf_{\delta>0} \liminf_{t\to\infty} \frac{1}{t} \log P\left(\frac{\boldsymbol{Y}}{t} \in B_\delta(\boldsymbol{x})\right). \tag{EC.2}$$

Fix any $\varepsilon, M \in (0,\infty)$ and $\boldsymbol{x} \in (0,M)^d$. Since $f_{\boldsymbol{Y}}(\boldsymbol{y}) = p(\boldsymbol{y})\exp(-\varphi(\boldsymbol{y}))$,

$$P(\boldsymbol{Y}/t \in B_\delta(\boldsymbol{x})) = \int_{\boldsymbol{y}/t\in B_\delta(\boldsymbol{x})} p(\boldsymbol{y})\exp(-\varphi(\boldsymbol{y}))d\boldsymbol{y} = t^d \int_{\boldsymbol{z}\in B_\delta(\boldsymbol{x})} p(t\boldsymbol{z})\mathrm{e}^{-\varphi(t\boldsymbol{z})}d\boldsymbol{z}.$$

Recall that the continuous convergences in Assumption 3 imply the following uniform convergences over compact sets not containing the origin (see (Rockafellar & Wets 1998, Theorem 7.14)):

$$n^{-1}\varphi(n\boldsymbol{x}) \xrightarrow{n\to\infty} I(\boldsymbol{x}) \text{ and } n^{-1}\log p(n\boldsymbol{x}) \xrightarrow{n\to\infty} 0. \tag{EC.3}$$

Due to this local uniform convergence and the continuity of I, there exist $\delta_0, t_0 \in (0,\infty)$ such that,

$$\left|\frac{\varphi(t\boldsymbol{z})}{t} - I(\boldsymbol{x})\right| \le \left|\frac{\varphi(t\boldsymbol{z})}{t} - I(\boldsymbol{z})\right| + |I(\boldsymbol{z}) - I(\boldsymbol{x})| \le \varepsilon/2, \text{ for all } \boldsymbol{z} \in B_\delta(\boldsymbol{x})$$

and $\exp(-\varepsilon t/2) \le p(t\boldsymbol{z}) \le \exp(\varepsilon t/2)$, whenever $t > t_0$, $\delta < \delta_0$ and $B_{\delta_0}(\boldsymbol{x})$ does not contain the origin. Thus, given ε, M and $\boldsymbol{x} \in (0,M)^d$, there exist $\delta_0, t_0 \in (0,\infty)$ such that for all $t > t_0$ and $\delta \in (0,\delta_0)$,

$$\exp\left(-t(I(\boldsymbol{x})+\varepsilon)\right) \le f_{\boldsymbol{Y}}(t\boldsymbol{z}) \le \exp\left(-t(I(\boldsymbol{x})-\varepsilon)\right), \text{ uniformly over } \boldsymbol{z} \in B_\delta(\boldsymbol{x}); \tag{EC.4}$$

Then $t^d\mathrm{Vol}(B_\delta(\boldsymbol{x}))\exp(-t(I(\boldsymbol{x})+\epsilon)) \le P(t^{-1}\boldsymbol{Y} \in B_\delta(\boldsymbol{x})) \le t^d\mathrm{Vol}(B_\delta(\boldsymbol{x}))\exp(-t(I(\boldsymbol{x})-\epsilon))$. Since $P(\boldsymbol{Y}/t \in B_\delta(\boldsymbol{x}))$ is increasing in δ and these bounds hold for any $\delta < \delta_0$,

$$-I(\boldsymbol{x}) - \epsilon \le \inf_{\delta>0} \liminf_{t\to\infty} \frac{1}{t}\log P(t^{-1}\boldsymbol{Y} \in B_\delta(\boldsymbol{x})) \le \inf_{\delta>0} \limsup_{t\to\infty} \frac{1}{t}\log P(\boldsymbol{Y}/t \in B_\delta(\boldsymbol{x})) \le -I(\boldsymbol{x}) + \epsilon.$$

Since the choices $\varepsilon, M \in (0,\infty)$ are arbitrary, (EC.2) holds. □

EC.1.1. Proof of Theorem 2

For functions f and g, let $f \wedge g$ (resp. $f \vee g$) denote their point-wise minimum (resp. maximum).

LEMMA EC.1. *Under Assumption 2, we have* $q_\infty(t(u)) = u^{1/\rho}$. *Therefore when Assumption 1 additionally holds, the events* $\{L(\boldsymbol{X}) \ge u\}$ *and* $\{\boldsymbol{Y}_u \in \mathrm{lev}_1^+(L_u)\}$ *coincide.*

Proof. Consider increasing real valued functions f_1, f_2. By the definition of left-continuous inverses, $(f_1 \vee f_2)^{\leftarrow}(y) = \inf\{u : f_1(u) \ge y\} \wedge \inf\{u : f_2(u) \ge y\} = f_1^{\leftarrow}(y) \wedge f_2^{\leftarrow}(y)$. From induction, $(\vee_{i=1}^d f_i)^{\leftarrow} = \wedge_{i=1}^d f_i^{\leftarrow}$, given increasing functions $f_1, \ldots, f_d$. Since $\{\Lambda_i : i = 1,\ldots,d\}$ are continuous, $q_i^{\leftarrow} = \Lambda_i$ (see (de Haan & Ferreira 2010, Excercise 1.1 (a))). Consequently,

$$q_\infty^{\leftarrow}(t) = \min_{i=1,...d} q_i^{\leftarrow}(t) = \Lambda_{\min}(t). \tag{EC.5}$$

Then $q_\infty(\Lambda_{\min}(x)) = x$ for all $x \in (x_0, \infty)$ due to the strict monotonicity in Assumption 2. Since $\boldsymbol{q} = \boldsymbol{\Lambda}^{\leftarrow}$ is injective, $\{\boldsymbol{Y}_u \in A\} = \{\boldsymbol{X} \in \mathbf{q}(t(u)A)\}$, for any measurable A. With $L_u(\boldsymbol{x}) := u^{-1}L(\boldsymbol{q}(t(u)\boldsymbol{x}))$,

$$\{\boldsymbol{q}(t(u)\boldsymbol{y}) : \boldsymbol{y} \in \mathrm{lev}_1^+(L_u)\} = \{\boldsymbol{q}(t(u)\boldsymbol{y}) : L(\boldsymbol{q}(t(u)\boldsymbol{y})) \geq u\} = \{\boldsymbol{x} : L(\boldsymbol{x}) > u\} \cap \mathrm{supp}(\boldsymbol{X}). \qquad \square$$

Lemma EC.2. *If Assumptions 1 and 2 hold and the limit $\boldsymbol{q}^*$ exists, the sequence of functions $\{L_u : u > 0\}$ converge continuously to f_{LD} on $\mathbb{R}^d_{++}$. Consequently, for any $\alpha, \varepsilon, M, K > 0$, there exists u_0 large enough such that for all $u > u_0$,*

$$\Xi_{\alpha,M}(L_u) \subseteq \left[\Xi_{\alpha,M}(f_{\mathrm{LD}})\right]^{1+\varepsilon/2} \quad \text{and} \quad \Xi_{\alpha+\varepsilon,M}(f_{\mathrm{LD}}) \subseteq [\Xi_{\alpha+\varepsilon/2,M}(L_u)]^{1+\varepsilon/K}$$

Proof. **1)** Consider any sequences $\{u_n\} \subset \mathbb{R}_+, \{\boldsymbol{x}_n\} \subset \mathbb{R}^d_+$ satisfying $u_n \uparrow \infty, \boldsymbol{x}_n \to \boldsymbol{x} > \mathbf{0}$.We rewrite,

$$L_{u_n}(\boldsymbol{x}_n) = u_n^{-1} L\left(\boldsymbol{q}(t(u_n)\boldsymbol{x}_n)/\boldsymbol{q}(t(u_n)\mathbf{1}) \cdot \hat{\boldsymbol{q}}(t(u_n))u_n^{1/\rho}\right).$$

Consider any $\delta > 0$ such that $B_\delta(\boldsymbol{x})$ does not include $\mathbf{0}$. As $\Lambda_i \in \mathcal{RV}(\alpha_i)$, we have $q_i \in \mathcal{RV}(1/\alpha_i)$ (see Part 9 of (de Haan & Ferreira 2010, Proposition B.1.9)). Due to the uniform convergence $\lim_{t\to\infty} \boldsymbol{q}(t\boldsymbol{y})/\boldsymbol{q}(t\mathbf{1}) = \boldsymbol{y}^{\mathbf{1}/\boldsymbol{\alpha}}$ over $\boldsymbol{y} \in B_\delta(\boldsymbol{x})$ (see (de Haan & Ferreira 2010, Proposition B.1.4)),

$$\boldsymbol{q}(t(u_n)\boldsymbol{x}_n)/\boldsymbol{q}(t(u_n)\mathbf{1}) \to \boldsymbol{x}^{\mathbf{1}/\boldsymbol{\alpha}}, \text{ and } \hat{\boldsymbol{q}}(t(u_n)) \to \boldsymbol{q}^*,$$

Applying triangle inequality, $\boldsymbol{p}_n = \boldsymbol{q}(t(u)\boldsymbol{x}_u)/\boldsymbol{q}(t(u)\mathbf{1})\hat{\boldsymbol{q}}(t(u)) \to \boldsymbol{x}^{\mathbf{1}/\boldsymbol{\alpha}}\boldsymbol{q}^*$. Continuous convergence

$$L_{u_n}(\boldsymbol{x}_n) := u_n^{-1} L(u_n^{1/\rho_n}\boldsymbol{p}_n) \to L^*(\boldsymbol{q}^*\boldsymbol{x}^{\mathbf{1}/\boldsymbol{\alpha}}) =: f_{\mathrm{LD}}(\boldsymbol{x})$$

follows from Assumption 1(b).

2) We next prove the claims on the set inclusions using the notions of lim sup, lim inf of a sequence of sets defined in the Kuratowski sense (see (Rockafellar & Wets 1998, Chapter 1)). Taking the ambient space $\mathcal{X} = \mathbb{R}^d_{++}$ in (Beer *et al.* 1992, Theorem 3.1), we obtain

$$\limsup_u \Xi_\beta(L_u) \subseteq \Xi_\beta(f_{\mathrm{LD}}), \quad \liminf_u \Xi_{\beta_u}(L_u) \supseteq \Xi_\beta(f_{\mathrm{LD}}) \text{ for some } \beta_u \nearrow \beta, \tag{EC.6}$$

as a consequence of above verified $L_{u_n}(\boldsymbol{x}_n) \to f_{\mathrm{LD}}(\boldsymbol{x})$. Refer the construction in the proof of (Beer *et al.* 1992, Theorem 3.1) for using fixed level β in the first set inclusion and considering increasing sequence β_u in the second set inclusion in (EC.6). Now, setting $\beta = \alpha$ in (EC.6), using the equivalence between $(v)_b$ and $(vii)_b$ in (Salinetti & Wets 1981, Theorem 2.2),

$$\Xi_{\alpha,M}(L_u) \subseteq \left[\Xi_{\alpha,M}(f_{\mathrm{LD}})\right]^{1+\varepsilon/2} \cap \mathbb{R}^d_{++} \subseteq \left[\Xi_{\alpha,M}(f_{\mathrm{LD}})\right]^{1+\varepsilon/2}.$$

Set $\beta = \alpha + \varepsilon$, and let $\beta_u \nearrow \alpha + \varepsilon$ be selected as in (EC.6). Therefore, for all large enough u, $\Xi_{\beta_u,M}(L_u) \subseteq \Xi_{\alpha+\varepsilon/2,M}(L_u)$. From (EC.6), $\Xi_{\alpha+\varepsilon,M}(f_{\mathrm{LD}}) \subseteq \liminf_u \Xi_{\beta_u}(L_u) \subseteq \liminf_u \Xi_{\alpha+\varepsilon/2,M}(L_u)$. Further, from the equivalence between $(vii)_a$ and $(v)_a$ in (Salinetti & Wets 1981, Theorem 2.2),

$$\Xi_{\alpha+\varepsilon,M}(f_{\mathrm{LD}}) \subseteq [\Xi_{\alpha+\varepsilon/2,M}(L_u)]^{1+\varepsilon/K} \cap \mathbb{R}^d_{++} \subseteq [\Xi_{\alpha+\varepsilon/2,M}(L_u)]^{1+\varepsilon/K}. \qquad \square$$

COROLLARY EC.1. *Suppose that Assumptions 1 and 2 hold and the limit $\boldsymbol{q}^*$ exists. Then for any $\alpha, \varepsilon, M > 0$, there exists u_0 large enough such that for all $u > u_0$,*

$$\text{lev}_\alpha^+(L_u) \cap B_M \subseteq \left[\Xi_{\alpha,M}(f_{\text{LD}})\right]^{1+\varepsilon} \quad \textit{and} \quad \Xi_{\alpha+\varepsilon,M}(f_{\text{LD}}) \subseteq \Xi_{\alpha,M}(L_u). \tag{EC.7}$$

Consequently, $\liminf_{n\to\infty} \chi_{\text{lev}_\alpha^+(L_{u_n})}(\boldsymbol{x}_n) \geq \chi_{\text{lev}_\alpha^+(f_{\text{LD}})}(\boldsymbol{x})$, *for any $\boldsymbol{x}_n \to \boldsymbol{x}$ and $u_n \to \infty$.*

Proof. Notice that for $\varepsilon > 0$, $\text{lev}_\alpha^+(L_u) \cap B_M \subseteq [\Xi_{\alpha,M}(L_u)]^{1+\varepsilon/2}$, as a consequence of the definitions at the beginning of Section EC.1. The first set inclusion now follows from the definition of the $[A]^{(1+r)}$ for a set $A \subset \mathbb{R}_+^d$. For the second set inclusion, observe that for $K > 0$ (to be chosen imminently),

$$\Xi_{\alpha+\varepsilon,M}(f_{\text{LD}}) \subseteq [\Xi_{\alpha+\varepsilon/2,M}(L_u)]^{1+\varepsilon/K} \cap \mathbb{R}_{++}^d$$

for all large enough u. Further, for any $\boldsymbol{x}$, $\boldsymbol{y}$,

$$|L_u(\boldsymbol{x}) - L_u(\boldsymbol{y})| \leq |L_u(\boldsymbol{y}) - f_{\text{LD}}(\boldsymbol{y})| + |f_{\text{LD}}(\boldsymbol{y}) - f_{\text{LD}}(\boldsymbol{x})| + |f_{\text{LD}}(\boldsymbol{x}) - L_u(\boldsymbol{x})|.$$

Recall that $L_u \to f_{\text{LD}}$ uniformly on $\mathbb{R}_{++}^d \cap B_{2M}$ (see for example, (Rockafellar & Wets 1998, Theorem 7.14)). Hence the first and third terms above may be made less than $\varepsilon/6$ for a large enough u for any choices of $\boldsymbol{x}, \boldsymbol{y} \in [\Xi_{\alpha+\varepsilon/2,M}(L_u)]^{1+\varepsilon/K} \cap \mathbb{R}_{++}^d$. Next, f_{LD} is uniformly continuous over B_{2M}. Therefore, there exists a $\kappa > 0$, such that for $\boldsymbol{x}, \boldsymbol{y} \in B_{2M}$, whenever $\|\boldsymbol{x} - \boldsymbol{y}\| \leq \kappa$, $|f_{\text{LD}}(\boldsymbol{x}) - f_{\text{LD}}(\boldsymbol{y})| \leq \varepsilon/6$. Consider any $K \geq \varepsilon/\kappa$. For any $\boldsymbol{y} \in [\Xi_{\alpha+\varepsilon/2,M}(L_u)]^{1+\varepsilon/K} \cap \mathbb{R}_{++}^d$, there exists $\boldsymbol{x} \in \Xi_{\alpha+\varepsilon/2,M}(L_u)$ such that $\|\boldsymbol{x} - \boldsymbol{y}\| \leq \kappa$, and consequently,

$$|L_u(\boldsymbol{x}) - L_u(\boldsymbol{y})| \leq \varepsilon/6 + \varepsilon/6 + \varepsilon/6 \leq \varepsilon/2.$$

Therefore, whenever $\boldsymbol{y} \in [\Xi_{\alpha+\varepsilon/2,M}(L_u)]^{1+\varepsilon/K} \cap \mathbb{R}_{++}^d$, $L_u(\boldsymbol{y}) \geq \alpha$. Hence $\Xi_{\alpha+\varepsilon,M}(f_{\text{LD}}) \subseteq \Xi_{\alpha,M}(L_u)$. To verify the conclusion on characteristic functions, we proceed as follows: The bound in the statement is immediate if $\boldsymbol{x} \in \text{lev}_\alpha^+(f_{\text{LD}})$, as $\chi_{\text{lev}_\alpha^+(f_{\text{LD}})}(\boldsymbol{x}) = 0$ in that case. Consider the case where $\boldsymbol{x}_n \to \boldsymbol{x} \notin \text{lev}_\alpha^+(f_{\text{LD}})$. Then for a suitably small $\delta > 0$, $B_\delta(\boldsymbol{x}) \cap [\text{lev}_\alpha^+(f_{\text{LD}})]^{1+\delta} = \varnothing$, because of the continuity of f_{LD} and $\text{lev}_\alpha^+(f_{\text{LD}})$ being a closed set. Fix $M > \|\boldsymbol{x}\| + \delta$. With $u_n \to \infty$, we have

$$\text{lev}_\alpha^+(L_{u_n}) \cap B_M \subseteq \left[\Xi_{\alpha,M}(f_{\text{LD}})\right]^{1+\delta} \subseteq \left[\text{lev}_\alpha^+(f_{\text{LD}}) \cap B_M\right]^{1+\delta}, \quad n > n_0$$

for sufficiently large n_0, due to the first inclusion in (EC.7). For n_1 chosen large enough to ensure $\{\boldsymbol{x}_n\}_{n>n_1} \subseteq B_\delta(\boldsymbol{x})$, we have $\{\boldsymbol{x}_n : n > n_1\} \cap [\text{lev}_\alpha^+(f_{\text{LD}})]^{1+\delta} = \varnothing$. Consequently, $\boldsymbol{x}_n \notin \text{lev}_\alpha^+(L_{u_n}) \cap B_M$, for all $n > n_2 := \max\{n_0, n_1\}$. Since $M > \|\boldsymbol{x}\| + \delta$ and $\|\boldsymbol{x}_n\| \leq \|\boldsymbol{x}\| + \delta$ for $n > n_2$, $\boldsymbol{x}_n \notin \text{lev}_\alpha^+(L_{u_n}) \cap (\mathbb{R}^d \setminus B_M)$ either. Therefore $\boldsymbol{x}_n \notin \text{lev}_\alpha^+(L_{u_n})$ and $\chi_{\text{lev}_\alpha^+(L_{u_n})}(\boldsymbol{x}_n) = \infty$ for $n > n_2$. □

Lemma EC.3. *Suppose that $f,g:\mathbb{R}^d_+\to\mathbb{R}$ are continuous. For any α,δ,M positive,*

$$\inf_{\boldsymbol{x}\in[\Xi_{\alpha,M}(f)]^{1+\varepsilon}} g(\boldsymbol{x}) \geq \inf_{\boldsymbol{x}\in \mathrm{lev}^+_\alpha(f)\cap B_M} g(\boldsymbol{x})-\delta \quad \text{and}$$
$$\inf_{\boldsymbol{x}\in \mathrm{int}(\Xi_{\alpha+\varepsilon,M}(f))} g(\boldsymbol{x}) \leq \inf_{\boldsymbol{x}\in \mathrm{lev}^+_\alpha(f)\cap B_M} g(\boldsymbol{x})+\delta,$$

for all ε suitably small.

Proof. Observe that the sequence of sets $\mathcal{X}_n := \{[\Xi_{\alpha,M}(f)]^{1+1/n}\}_{n\geq 1}$ are uniformly bounded in n and $\cup_{i\geq 1}\mathcal{X}_i$ is relatively compact. Further, $\mathcal{X}_n\searrow \mathrm{lev}^+_\alpha(f)\cap B_M$ in the Kuratowski sense. Therefore, from (Langen 1981, Theorem 2.2 (iii)), for all small enough ϵ,

$$\inf_{\boldsymbol{x}\in[\Xi_{\alpha,M}(f)]^{1+\epsilon}} g(\boldsymbol{x}) \geq \inf_{\boldsymbol{x}\in \mathrm{lev}^+_\alpha\cap B_M} g(\boldsymbol{x})-\delta.$$

In a similar spirit, due to the continuity of $f(\cdot)$, one has $\Xi_{\alpha+1/n,M}(f)\nearrow \Xi_{\alpha,M}(f)$. Then, upon an application of (Langen 1981, Theorem 2.2 (iii)),

$$\inf_{\boldsymbol{x}\in\Xi_{\alpha+\varepsilon,M}(f)} g(\boldsymbol{x}) \leq \inf_{\boldsymbol{x}\in\Xi_{\alpha,M}(f)} g(\boldsymbol{x})+\delta.$$

The statement then follows from the continuity of $g(\cdot)$. □

Proof of Theorem 2. Fix any $\delta, M>0$. Observe that $I(\cdot)$ and $L^*(\cdot)$ are continuous. As a consequence of Lemma EC.3, there exists $\varepsilon>0$ suitably small such that

$$\inf_{\boldsymbol{p}\in[\Xi_{1,M}(f_{\mathrm{LD}})]^{1+\epsilon}} I(\boldsymbol{p}) \geq \inf_{\boldsymbol{p}\in\Xi_{1,M}(f_{\mathrm{LD}})} I(\boldsymbol{p})-\delta, \text{ and} \tag{EC.8a}$$

$$\inf_{\boldsymbol{p}\in \mathrm{int}(\Xi_{1+\varepsilon,M}(f_{\mathrm{LD}}))} I(\boldsymbol{p}) \leq \inf_{\boldsymbol{p}\in \mathrm{lev}^+_1(f_{\mathrm{LD}})\cap B_M} I(\boldsymbol{p})+\delta. \tag{EC.8b}$$

Large deviations upper bound. Due to Lemma EC.1,

$$P\big(L(\boldsymbol{X})\geq u\big) \leq P\big(\boldsymbol{Y}_u\in \mathrm{lev}^+_1(L_u)\cap B_M\big)+P\big(\boldsymbol{Y}_u\in B^c_M\big). \tag{EC.9}$$

From Corollary EC.1, $P\big(\boldsymbol{Y}_u\in \mathrm{lev}^+_1(L_u)\cap B_M\big)\leq P\big(\boldsymbol{Y}_u\in \big[\Xi_{1,M}(f_{\mathrm{LD}})\big]^{1+\varepsilon}\big)$, for all u sufficiently large. Since the expansion set $[A]^{1+r}$ is closed for any $A\subseteq\mathbb{R}^d_+$, the set $[\Xi_{\alpha,M}(f_{\mathrm{LD}})]^{1+\varepsilon}$ is closed. Therefore,

$$\begin{aligned}\limsup_{u\to\infty}\frac{1}{t(u)}\log P\big(\boldsymbol{Y}_u\in \mathrm{lev}^+_1(L_u)\cap B_M\big) &\leq \limsup_{u\to\infty}\frac{1}{t(u)}\log P\big(t(u)^{-1}\boldsymbol{Y}\in\big[\Xi_{1,M}(f_{\mathrm{LD}})\big]^{1+\varepsilon}\big)\\ &\leq -\inf_{\boldsymbol{p}\in[\Xi_{1,M}(f_{\mathrm{LD}})]^{1+\varepsilon}} I(\boldsymbol{p})\\ &\leq -\inf_{\boldsymbol{p}\in \mathrm{lev}^+_1(f_{\mathrm{LD}})\cap B_M} I(\boldsymbol{p})+\delta,\end{aligned}$$

where the second inequality follows from the Tail LDP in Theorem 1 and the third inequality is a consequence of the choice of ε satisfying (EC.8a). Since $\delta>0$ is arbitrary,

$$\limsup_{u\to\infty}\frac{1}{t(u)}\log P\big(\boldsymbol{Y}_u\in \mathrm{lev}^+_1(L_u)\cap B_M\big) \leq -\inf_{\boldsymbol{p}\in \mathrm{lev}^+_1(f_{\mathrm{LD}})\cap B_M} I(\boldsymbol{p}) \leq -I^*,$$

where $I^* := \inf_{\boldsymbol{p}\in \text{lev}_1^+(f_{\text{LD}})} I(\boldsymbol{p})$. A similar application of Theorem 1 results in

$$\limsup_{u\to\infty} \frac{1}{t(u)} \log P\left(\boldsymbol{Y}_u \in B_M^c\right) \leq -\inf_{\boldsymbol{p}\in \text{cl}(B_M^c)} I(\boldsymbol{p}) = -M, \tag{EC.10}$$

where the latter inequality follows from Lemma 2d and the continuity of $I(\cdot)$. Combining these conclusions with that in (EC.9) and (Dembo & Zeitouni 1998, Lemma 1.2.15),

$$\limsup_{u\to\infty} \frac{1}{t(u)} \log P\left(L(\boldsymbol{X}) \geq u\right) \leq -\min\left\{M, I^*\right\}.$$

Since M can be made arbitrarily large, we have $\limsup_{u\to\infty} t(u)^{-1} \log P\left(L(\boldsymbol{X}) \geq u\right) \leq -I^*$.

Large deviations lower bound. We have from Lemma EC.1 and the definition of $\Xi_{\alpha,M}$ that $P(L(\boldsymbol{X}) \geq u) \geq P(\boldsymbol{Y}_u \in \Xi_{1,M}(L_u))$. For ε satisfying (EC.8b), it follows from the second set inclusion in Corollary EC.1 that $P(L(\boldsymbol{X}) \geq u) \geq P(\boldsymbol{Y}_u \in \text{int}\left(\Xi_{1+\varepsilon,M}(f_{\text{LD}})\right))$, for all sufficiently large enough u. Then as an application of the tail LDP in Theorem 1,

$$\liminf_{u\to\infty} \frac{1}{t(u)} \log P(L(\boldsymbol{X}) \geq u) \geq -\inf_{\boldsymbol{p}\in \text{int}\left(\Xi_{1+\varepsilon,M}(f_{\text{LD}})\right)} I(\boldsymbol{p}) \geq -\inf_{\boldsymbol{p}\in \text{lev}_1^+(f_{\text{LD}})\cap B_M} I(\boldsymbol{p}) - \delta,$$

where the latter inequality is a consequence of ε satisfying (EC.8b). Since M, δ are arbitrary,

$$\liminf_{u\to\infty} \frac{1}{t(u)} \log P\left(L(\boldsymbol{X}) \geq u\right) \geq -\inf_{\boldsymbol{p}\in \text{lev}_1^+(f_{\text{LD}})} I(\boldsymbol{p}) = -I^*. \qquad \square$$

REMARK EC.1. In the case where L_u as defined in (EC.1a) is independent of u, it can be seen that for a fixed map $\boldsymbol{Q}: \mathbb{R}_+^d \to \mathbb{R}_d^+$, $\boldsymbol{Q}(\text{lev}_1^+(L_u)) = \text{lev}_1^+(f_{\text{LD}})$. Thus, in such a case, Theorem 2 may be derived as a consequence of the contraction principle, rather than utilise the more elaborate machinery developed in Lemmas EC.2-EC.3. The former holds for example, if (i) $L(n\boldsymbol{x}) = n^\rho L(\boldsymbol{x})$ for all $n, \boldsymbol{x}$ and (ii) $\boldsymbol{\Lambda}(\boldsymbol{x}) = \boldsymbol{x}^{\boldsymbol{\alpha}}$ for some $\alpha > 0$.

EC.1.2. Proof of Proposition 1 and useful bounds on the inverse of $\boldsymbol{T}(\cdot)$.

Proof of Proposition 1: First consider the case $\boldsymbol{\kappa} = \boldsymbol{\kappa}^{(1)}$. Let $k_\rho(u) := \log(u/l)$. It is sufficient to show that the determinant of the Jacobian of the map $\boldsymbol{x} \mapsto \boldsymbol{T}(\boldsymbol{x})$ equals $J(\boldsymbol{x})$. In that case, the density of $\boldsymbol{Z}$, denoted by $f_{\boldsymbol{Z}}(\cdot)$, is $f_{\boldsymbol{Z}}(\boldsymbol{T}(\boldsymbol{x})) = f_{\boldsymbol{X}}(\boldsymbol{x})/J(\boldsymbol{x})$; consequently, the likelihood ratio between the distributions of $\boldsymbol{Z}$ and that of $\boldsymbol{X}$ (or the Radon-Nikodym derivative evaluated at the samples $\boldsymbol{Z}_1 \ldots \boldsymbol{Z}_n$) is given as in Algorithm 1. So the rest of the verification is devoted to checking that $J(\cdot)$ indeed equals the determinant of the Jacobian $\text{Jac}_{\boldsymbol{T}}(\cdot) = \partial \boldsymbol{T}/\partial \boldsymbol{x}$. To this end, define

$\boldsymbol{\psi}(\boldsymbol{x}) = \log|\boldsymbol{x}| + k_\rho(u)\boldsymbol{\kappa}(\boldsymbol{x})$ and observe that $\boldsymbol{T}(\boldsymbol{x}) = \text{Diag}(\text{sgn}(\boldsymbol{x}))\text{e}^{\boldsymbol{\psi}(\boldsymbol{x})}$. Now, following the chain rule for Jacobians,

$$\begin{aligned}
&\text{Jac}(\boldsymbol{x}) = \text{Diag}(\text{sgn}(\boldsymbol{x}))\text{Diag}(\text{e}^{\boldsymbol{\psi}(\boldsymbol{x})})\text{Jac}_\psi(\boldsymbol{x}), \text{ for almost every } \boldsymbol{x}, \text{ and}\\
&\text{Jac}_\psi(\boldsymbol{x}) = \text{Diag}\left(\frac{\text{sgn}(\boldsymbol{x})}{|\boldsymbol{x}|}\right) + \rho^{-1}k_\rho(u)\left[\text{Diag}\left(\frac{\text{sgn}(\boldsymbol{x})}{(1+|\boldsymbol{x}|)\|\log(1+|\boldsymbol{x}|)\|_\infty}\right) - \frac{\log(1+|\boldsymbol{x}|)}{\|\log(1+|\boldsymbol{x}|)\|_\infty^2}(\boldsymbol{e}^*/(1+|\boldsymbol{x}|))^\intercal\right]\\
&\qquad = \text{Diag}\left(\text{sgn}(\boldsymbol{x})\left[\frac{1}{|\boldsymbol{x}|} + \frac{\rho^{-1}k_\rho(u)}{(1+|\boldsymbol{x}|)\|\log(1+|\boldsymbol{x}|)\|_\infty}\right]\right) - \rho^{-1}k_\rho(u)\left[\frac{\log(1+|\boldsymbol{x}|)}{\|\log(1+|\boldsymbol{x}|)\|_\infty^2}(\boldsymbol{e}^*/(1+|\boldsymbol{x}|))^\intercal\right]
\end{aligned}$$

where $\boldsymbol{e}_i^* = \text{sgn}(x_i)$ if $|x_i| = \|\boldsymbol{x}\|_\infty$, and $\boldsymbol{e}_i^* = 0$ otherwise. Notice that for this component, $\|\boldsymbol{x}\|_\infty = |x_i|$. Now, recall that if $M = A + \mathbf{u}\mathbf{v}^\intercal$, then $|M| = (1 + \mathbf{u}^\intercal A^{-1}\mathbf{v})|A|$. Set

$$\begin{aligned}
&\boldsymbol{u} = \log(1+|\boldsymbol{x}|)/\|\log(1+|\boldsymbol{x}|)\|_\infty^2, \quad \boldsymbol{v} = -(k_\rho(u)\boldsymbol{e}^*/(1+|\boldsymbol{x}|))^\intercal, \text{ and}\\
&A = \text{Diag}\left(\text{sgn}(\boldsymbol{x})\left[\frac{1}{|\boldsymbol{x}|} + \frac{k_\rho(u)}{(1+|\boldsymbol{x}|)\|\log(1+|\boldsymbol{x}|)\|_\infty}\right]\right).
\end{aligned}$$

Then, almost everywhere,

$$\begin{aligned}
1 + \boldsymbol{u}^\intercal A^{-1}\boldsymbol{v} &= 1 - \frac{\rho^{-1}k_\rho(u)\|\boldsymbol{x}\|_\infty}{\|1+|\boldsymbol{x}|\|_\infty\|\log(1+|\boldsymbol{x}|)\|_\infty + k_\rho(u)\|\boldsymbol{x}\|_\infty}, \text{ and}\\
|A| &= \prod_{i=1}^d \frac{1}{|x_i|} \times \prod_{i=1}^d \left(1 + \frac{\rho^{-1}k_\rho(u)|x_i|}{(1+|x_i|)\|\log(1+|\boldsymbol{x}|)\|_\infty}\right).
\end{aligned}$$

To complete the verification, observe $|\text{Diag}(\text{e}^{\boldsymbol{\psi}(\boldsymbol{x})})| = \prod_{i=1}^d |x_i| \cdot (u/l)^{\mathbf{1}^\intercal\boldsymbol{\kappa}(\boldsymbol{x})}$.

We next show that $\boldsymbol{T}(\cdot)$ is onto. Fix a $\boldsymbol{y} \in \mathbb{R}_+^d$. Let $I = \{i : y_i = \|\boldsymbol{y}\|_\infty\}$. Define the set,

$$S = \{\boldsymbol{x} : x_i = \|\boldsymbol{y}\|_\infty c_\rho(u) \text{ for all } i \in I,\, x_i \in [0, \|\boldsymbol{y}\|_\infty c_\rho(u)],\, i \notin I\}, \tag{EC.11}$$

where recall $c_\rho(u) := (l/u)^{1/\rho}$. Notice that for all $\boldsymbol{x} \in S$,

$$\boldsymbol{T}_i(\boldsymbol{x}) = \begin{cases} x_i[c_\rho(u)]^{\frac{-\log(1+x_i)}{\log(1+c_\rho(u)\|\boldsymbol{y}\|_\infty)}} & \text{for } i \notin I,\\ \|\boldsymbol{y}\|_\infty & \text{for } i \in I.\end{cases}$$

Restricted to $\boldsymbol{x} \in S$, $\boldsymbol{T}_i(\boldsymbol{x})$ is only a function of x_i for $i \notin I$. Fixing $i \notin I$, see that $\boldsymbol{T}_i(\boldsymbol{x}) < y_i$, if $x_i < y_i c_\rho(u)$; and $\boldsymbol{T}_i(\boldsymbol{x}) = \|\boldsymbol{y}\|_\infty > y_i$ if $x_i = \|\boldsymbol{y}\|_\infty c_\rho(u)$. Since $\boldsymbol{T}_i$ are all continuous maps, by the intermediate value theorem, there exists some $\boldsymbol{x}' \in [\boldsymbol{y}c_\rho(u), \|\boldsymbol{y}\|_\infty c_\rho(u)]^d$ such that $\boldsymbol{T}(\boldsymbol{x}') = \boldsymbol{y}$. The above argument also shows that if $\boldsymbol{x}_1 \neq \boldsymbol{x}'$, $\boldsymbol{T}(\boldsymbol{x}_1) \neq \boldsymbol{T}(\boldsymbol{x}') = \boldsymbol{y}$, that is, $\boldsymbol{T}(\cdot)$ is $1 \leftrightarrow 1$. Since $\boldsymbol{T}(\cdot)$ is symmetric about the origin, one can similarly extend the proof to the case where $\boldsymbol{y} \in \mathbb{R}^d$.

Now suppose that $\boldsymbol{\kappa} = \boldsymbol{\kappa}^{(2)}$. Here, observe that by a direct application of the chain rule,

$$\left[\frac{\partial\boldsymbol{\kappa}(\boldsymbol{x})}{\partial\boldsymbol{x}}\right] = \text{Diag}\left(\frac{\log(u/l)}{\log l}\frac{\text{sgn}(\boldsymbol{x})}{1+|\boldsymbol{x}|}\right).$$

Substituting in the expression for the Jacobian in (9) completes the proof. □

In Lemmas EC.4 and EC.5 below, assume that $\boldsymbol{\kappa} = \boldsymbol{\kappa}^{(1)}$:

LEMMA EC.4. *For any $\boldsymbol{y} \in \mathbb{R}_+^d$ satisfying $\|\boldsymbol{y}\|_\infty \geq 1/c_\rho(u)$,*

$$\boldsymbol{y} c_\rho(u) \leq \boldsymbol{T}^{-1}(\boldsymbol{y}) \leq \min\left\{ \boldsymbol{y}[c_\rho(u)]^{\frac{\log \boldsymbol{y}}{\log \|\boldsymbol{y}\|_\infty}} \vee \mathbf{1}, \|\boldsymbol{y}\|_\infty c_\rho(u) \right\}. \tag{EC.12}$$

Proof. We verify the bounds by exhibiting $\boldsymbol{x}'$ sandwiched component-wise between the left and right hand sides of (EC.12) and satisfying $\boldsymbol{T}(\boldsymbol{x}') = \boldsymbol{y}$. For any $\boldsymbol{y}$ as in the statement, we first set

$$\tilde{x}_i = \begin{cases} y_i \, [c_\rho(u)]^{\frac{\log y_i}{\log \|\boldsymbol{y}\|_\infty}} & \text{if } y_i \geq 1, \\ 1 & \text{otherwise,} \end{cases} \tag{EC.13}$$

for $i = 1, \ldots, d$. See that $\tilde{x}_i \in [1, \|\boldsymbol{y}\|_\infty c_\rho(u)]$. This is because of the following two observations: 1) $\log \tilde{x}_i = (1 + \log c_\rho(u)/\log \|\boldsymbol{y}\|_\infty) \log y_i \geq 0$, when $\|\boldsymbol{y}\|_\infty \geq 1/c_\rho(u) > 1$; and 2) likewise,

$$\log \tilde{x}_i - \log(\|\boldsymbol{y}\|_\infty c_\rho(u)) = \left(1 + \frac{\log c_\rho(u)}{\log \|\boldsymbol{y}\|_\infty}\right)(\log y_i - \log \|\boldsymbol{y}\|_\infty) \leq 0.$$

Set $I = \{i : y_i = \|\boldsymbol{y}\|_\infty\}$. With the set S defined as in (EC.11), we thus have $\tilde{\boldsymbol{x}} = (\tilde{x}_1, \ldots, \tilde{x}_d) \in S$ and $\tilde{x}_i \geq 1$ for all i. Since $\log(1+t)/\log(t)$ is decreasing over $t \geq 1$, we have

$$\boldsymbol{T}_i(\boldsymbol{x}) = x_i [c_\rho(u)]^{-\frac{\log(1+x_i)}{\|\log(1+\boldsymbol{x})\|_\infty}} \geq x_i [c_\rho(u)]^{\frac{-\log x_i}{\log \|\boldsymbol{x}\|_\infty}}, \tag{EC.14}$$

for any $\boldsymbol{x}$ such that $x_i \geq 1, i = 1, \ldots, d$. With $\tilde{\boldsymbol{x}}$ defined via (EC.13), we obtain the following from the bound in (EC.14) and the observation $(\log \tilde{x}_i)/(\log \|\tilde{\boldsymbol{x}}\|_\infty) = (\log y_i)/(\log \|\boldsymbol{y}\|_\infty)$:

$$\boldsymbol{T}_i(\tilde{\boldsymbol{x}}) = \begin{cases} \boldsymbol{T}_i(y_i [c_\rho(u)]^{\frac{\log y_i}{\log \|\boldsymbol{y}\|_\infty}}) & \text{if } y_i \geq 1, \\ \boldsymbol{T}_i(1) & \text{otherwise,} \end{cases} \geq \begin{cases} y_i & \text{if } y_i \geq 1, \\ 1 & \text{otherwise.} \end{cases}$$

For $i \notin I$, the map $\boldsymbol{T}_i(\boldsymbol{x})$, when restricted to $\boldsymbol{x} \in S$, is a function only of x_i and satisfies $\boldsymbol{T}_i(\tilde{\boldsymbol{x}}) \geq y_i$ (regardless of whether $y_i > 1$ or $y_i < 1$). Applying the intermediate value theorem component-wise, we see that there exists some x_i' in the interval $[y_i c_\rho(u), \tilde{x}_i]$ containing y_i such that $T_i(x_i') = y_i$ for all $i \notin I$. Setting $x_i' = \|\boldsymbol{y}\|_\infty c_\rho(u)$ for $i \in I$, we obtain $\boldsymbol{T}(\boldsymbol{x}') = \boldsymbol{y}$. Hence the bounds (EC.12) hold. ☐

LEMMA EC.5. *Suppose that Assumptions 2 - 3 hold, $l(u)$ is slowly varying in u, and $\lim_{u \to \infty} l(u) = \infty$. Then for any $\gamma > 0$, the below convergence holds uniformly over $\boldsymbol{p}$ in compact subsets of $\mathbb{R}_{++}^d$:*

$$\|\psi_u(t(u)\boldsymbol{p})\|_\infty = o(t(u)), \text{ as } u \to \infty.$$

Proof. Recall $\psi_u := \boldsymbol{\Lambda} \circ \boldsymbol{T}^{-1} \circ \boldsymbol{q}$, $c_\rho(u) := (l(u)/u)^{1/\rho}$ and $q_\infty(t(u)) := \|\boldsymbol{q}(t(u))\|_\infty$. Fix any $M > 0, \gamma \in (0, M)$ and $\boldsymbol{p} \in B_M \setminus B_\gamma$. As $\lim_{u \to \infty} l(u) = +\infty$, $\boldsymbol{q} \in \mathcal{RV}$, and $q_\infty(t(u)) = u^{1/\rho}$ (see Lemma EC.1),

$$\|\boldsymbol{q}(t(u)\boldsymbol{p})\|_\infty c_\rho(u) = l(u)^{1/\rho} \frac{\|\boldsymbol{q}(t(u)\boldsymbol{p})\|_\infty}{\|\boldsymbol{q}(t(u))\|_\infty} \to \infty. \tag{EC.15}$$

Then from (EC.12), $\boldsymbol{T}_i^{-1}(\boldsymbol{q}(t(u)\boldsymbol{p})) \leq \|\boldsymbol{q}(t(u)\boldsymbol{p})\|_\infty c_\rho(u) \leq \|\boldsymbol{q}(t(u)M\mathbf{1})\|_\infty c_\rho(u)$, for $i = 1, \ldots, d$, as $\boldsymbol{q}$ is monotone and $\|\boldsymbol{p}\|_\infty \leq M$. With $q_\infty(t(u)) = u^{1/\rho}$,

$$\boldsymbol{q}(t(u)M\mathbf{1}) c_\rho(u) = \frac{\boldsymbol{q}(t(u)M\mathbf{1})}{\boldsymbol{q}(t(u))} \frac{\boldsymbol{q}(t(u))}{q_\infty(t(u))} l(u)^{1/\rho} \leq \frac{\boldsymbol{q}(t(u)M\mathbf{1})}{\boldsymbol{q}(t(u))} l(u)^{1/\rho}.$$

Since $\boldsymbol{q} \in \mathcal{RV}(\mathbf{1}/\boldsymbol{\alpha})$, we have $\boldsymbol{q}(t(u)M\mathbf{1})c_\rho(u) \le M^{\mathbf{1}/\boldsymbol{\alpha}} l(u)^{1/\rho}(\mathbf{1}+o(\mathbf{1}))$. Therefore,

$$\boldsymbol{T}_i^{-1}(\boldsymbol{q}(t(u)\boldsymbol{p})) \le \|\boldsymbol{q}(t(u)M\mathbf{1})\|_\infty c_\rho(u) \le \max_{i=1,\dots,d} M^{1/\alpha_i} l(u)^{1/\rho}(1+o(1)),$$

uniformly over $\boldsymbol{p} \in B_M \setminus B_\gamma(\mathbf{0})$. Recall $\boldsymbol{\Lambda} \in \mathcal{RV}(\boldsymbol{\alpha})$ is monotone. Write $\bar{\alpha} = \max_i \alpha_i$. Then for $\delta > 0$,

$$\|\boldsymbol{\psi}_u(t(u)\boldsymbol{p})\|_\infty = \max_{i=1,\dots,d} \Lambda_i\left(\boldsymbol{T}_i^{-1}(\boldsymbol{q}(t(u)\boldsymbol{p}))\right) \le (1+\delta)M_0 l(u)^{\bar{\alpha}/\rho+\delta},$$

for all u sufficiently large (see (de Haan & Ferreira 2010, Proposition B.1.9 (7))). Here $M_0 > 0$ is a suitably large constant. Since $t(u) = \Lambda_{\min}(u^{1/\rho})$, noting $l(u)^{\bar{\alpha}/\rho+\delta} = o(t(u))$ completes the proof. □

EC.1.3. Proof of Theorem 3 with $\boldsymbol{\kappa} = \boldsymbol{\kappa}_1$

Here, we demonstrate the proof of Theorem 3 in the case where $\boldsymbol{\kappa} = \boldsymbol{\kappa}_1$. The proof for $\boldsymbol{\kappa} = \boldsymbol{\kappa}_2$ is outlined in EC.5. Unless explicitly specified, the only assumption made in the proofs below is that $\boldsymbol{X}$ has a probability density $f_{\boldsymbol{X}}(\cdot)$. As a consequence, the hazard rates

$$\lambda_i(x) = f_{X_i}(x)/\Lambda_i(x), \quad i = 1, \dots, d$$

are well-defined. In the above, $f_{X_i}(\cdot)$ denotes the probability density of the component X_i.

Lemma EC.6. *The second moment $M_{2,u} = \mathbb{E}[\exp(-t(u)F_u(\boldsymbol{Y}_u))]$, where $F_u : \mathbb{R}_+^d \to \mathbb{R} \cup \{+\infty\}$ is,*

$$F_u(\boldsymbol{p}) := a_u(\boldsymbol{p}) + b_u(\boldsymbol{p}) - 2d\frac{\log t(u)}{t(u)} + \chi_{\mathrm{lev}_1^+(L_u)}(\boldsymbol{p}),$$

for $u > 0$, and $a_u : \mathbb{R}_+^d \to \mathbb{R}$, $b_u : \mathbb{R}_+^d \to \mathbb{R}$ are defined as follows:

$$a_u(\boldsymbol{p}) = \frac{1}{t(u)}\left[\log f_{\boldsymbol{Y}}(\boldsymbol{\psi}_u(t(u)\boldsymbol{p}) - \log f_{\boldsymbol{Y}}(t(u)\boldsymbol{p})\right],$$

$$b_u(\boldsymbol{p}) = \frac{1}{t(u)}\left[\sum_{i=1}^d \left[\log \lambda_i(\boldsymbol{T}_i^{-1}(\boldsymbol{q}(t(u)\boldsymbol{p}))) - \log \lambda_i(q_i(t(u)p_i))\right] - \log J(\boldsymbol{T}^{-1}(\boldsymbol{q}(t(u)\boldsymbol{p})))\right].$$

Proof. Since the change of measure is effected by the map $\boldsymbol{Z} = \boldsymbol{T}(\boldsymbol{X})$,

$$M_{2,u} = \mathbb{E}\left[\left(\frac{f_{\boldsymbol{X}}(\boldsymbol{Z})}{f_{\boldsymbol{Z}}(\boldsymbol{Z})}\right)^2 \mathbb{I}(L(\boldsymbol{Z}) \ge u)\right] = \mathbb{E}\left[\left(\frac{f_{\boldsymbol{X}}(\boldsymbol{X})}{f_{\boldsymbol{Z}}(\boldsymbol{X})}\right)\mathbb{I}(L(\boldsymbol{X}) \ge u)\right]$$
$$= \int_{L(\boldsymbol{x}) \ge u} \frac{f_{\boldsymbol{X}}(\boldsymbol{x})}{f_{\boldsymbol{X}}(\boldsymbol{T}^{-1}(\boldsymbol{x}))} J(\boldsymbol{T}^{-1}(\boldsymbol{x})) f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x}. \tag{EC.16}$$

Changing variables from $\boldsymbol{x}$ to $\boldsymbol{y} = \boldsymbol{q}^{-1}(\boldsymbol{x})$, we have $f_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{i=1}^d \lambda_i(x_i) f_{\boldsymbol{Y}}(\boldsymbol{q}^{-1}(\boldsymbol{x}))$, where $\lambda_i(x) = f_{X_i}(x)/\bar{F}_{X_i}(x)$ is the hazard rate of X_i. Thus,

$$M_{2,u} = \int_{L(\boldsymbol{q}(\boldsymbol{y})) \ge u} \prod_{i=1}^d \frac{\lambda_i(q_i(y_i))}{\lambda_i(\boldsymbol{T}_i^{-1}(q_i(y_i)))} \frac{f_{\boldsymbol{Y}}(\boldsymbol{y})}{f_{\boldsymbol{Y}}(\boldsymbol{\psi}_u(\boldsymbol{y}))} J(\boldsymbol{T}^{-1}(\boldsymbol{q}(\boldsymbol{y}))) f_{\boldsymbol{Y}}(\boldsymbol{y}) d\boldsymbol{y}$$
$$= t(u)^d \int_{\boldsymbol{p} \in \mathrm{lev}_1^+(L_u)} \underbrace{\frac{f_{\boldsymbol{Y}}(t(u)\boldsymbol{p})}{f_{\boldsymbol{Y}}(\boldsymbol{\psi}_u(t(u)\boldsymbol{p}))}}_{(a)} \underbrace{\prod_{i=1}^d \frac{\lambda_i(q_i(t(u)p_i))}{\lambda_i(\boldsymbol{T}_i^{-1}(\boldsymbol{q}(t(u)\boldsymbol{p})))} J(\boldsymbol{T}^{-1}(\boldsymbol{q}(t(u)\boldsymbol{p})))}_{(b)} f_{\boldsymbol{Y}}(t(u)\boldsymbol{p}) d\boldsymbol{p},$$

Since $\boldsymbol{Y}_u = t(u)^{-1}\boldsymbol{Y}$, $f_{\boldsymbol{Y}}(t(u)\boldsymbol{p}) = t(u)^d f_{\boldsymbol{Y}_u}(\boldsymbol{p})$. Checking the terms labeled (a) and (b) in the above expression equal $\exp(-t(u)a_u(\boldsymbol{p}))$ and $\exp(-t(u)b_u(\boldsymbol{p}))$, respectively, we obtain

$$M_{2,u} = t(u)^{2d}\mathbb{E}\left[\exp\left\{-t(u)\left[a_u(\boldsymbol{Y}_u) + b_u(\boldsymbol{Y}_u) + \chi_{\text{lev}_1^+(L_u)}(\boldsymbol{Y}_u)\right]\right\}\right]. \qquad \square$$

Thanks to Lemma EC.4 - EC.5, the terms in Lemma EC.6 enjoy the following bounds.

Lemma EC.7. *Suppose that Assumptions 2 - 3 are satisfied and $l(u)$ is slowly varying in u. Then $a_u(\boldsymbol{p}) \geq I(\boldsymbol{p}) + o(1)$, as $u \to \infty$, uniformly over $\boldsymbol{p}$ in compact subsets of $\mathbb{R}^d_{++}$.*

Proof. Under Assumption 3, uniformly over $\hat{\boldsymbol{y}} := \boldsymbol{y}/\|\boldsymbol{y}\|$ on $\mathcal{S}^{d-1} \cap \mathbb{R}^d_+$,

$$-\frac{\log f_{\boldsymbol{Y}}(\|\boldsymbol{y}\|\hat{\boldsymbol{y}})}{\|\boldsymbol{y}\|} \to I(\hat{\boldsymbol{y}}), \text{ as } \|\boldsymbol{y}\| \to \infty \quad (\text{replacing } n \text{ there by } \|\boldsymbol{y}\|)\,. \tag{EC.17}$$

Fix any $M > 0, \gamma \in (0, M)$ and $\varepsilon \in (0,1)$. From the monotonicity of $\boldsymbol{\Lambda}$, the lower bound in (EC.12), and (EC.15), $\|\boldsymbol{\psi}_u(t(u)\boldsymbol{p})\|_\infty := \|\boldsymbol{\Lambda}(\boldsymbol{T}^{-1}(\boldsymbol{q}(t(u)\boldsymbol{p})))\|_\infty \geq \|\boldsymbol{\Lambda}\left(\boldsymbol{q}(t(u)\boldsymbol{p})c_\rho(u)\right)\|_\infty \to \infty$, for any $\boldsymbol{p} \in B_M \setminus B_\gamma$, as $u \to \infty$. Then due to (EC.17) and the upper bound in (EC.12),

$$-\log f_{\boldsymbol{Y}}(\boldsymbol{\psi}(t(u)\boldsymbol{p})) \leq \|\boldsymbol{\psi}(t(u)\boldsymbol{p})\|_\infty\big[\sup_{\hat{\boldsymbol{y}} \in \mathcal{S}^{d-1}\cap\mathbb{R}^d_+} I(\hat{\boldsymbol{y}})) + \varepsilon\big] \leq \varepsilon t(u)(M_1 + \varepsilon)),$$

for all sufficiently large u; here $M_1 := \sup\{I(\hat{\boldsymbol{y}}) : \hat{\boldsymbol{y}} \in \mathcal{S}^{d-1} \cap \mathbb{R}^d_+\}$ is a finite positive constant (due to the regularity properties of I in Lemma 2). Observe that from the upper bound in (EC.4),

$$-\log f_{\boldsymbol{Y}}((t(u)\boldsymbol{p})) \geq t(u)(I(\boldsymbol{p}) - \varepsilon),$$

uniformly over $\boldsymbol{p} \in B_M \setminus B_\gamma$ and all u sufficiently large. Combining the above displayed bounds on $-\log f_{\boldsymbol{Y}}(\cdot)$ terms, we obtain from the definition of $a_u(\cdot)$ that $a_u(\boldsymbol{p}) \geq I(\boldsymbol{p}) - \varepsilon - \varepsilon(M_1 + \varepsilon)$. $\square$

Lemma EC.8. *Suppose that Assumptions 2 and 4 are satisfied and $l(u)$ is slowly varying in u. Then $\liminf_{u\to\infty} b_u(\boldsymbol{p}) \geq 0$, where the convergence is uniform over $\boldsymbol{p} \in \mathbb{R}^d_+$.*

Proof. Recall the definitions of $J(\boldsymbol{x})$ and $\tilde{J}_i(\boldsymbol{x})$ in Table 1. Since $\boldsymbol{\kappa}(\boldsymbol{x})$ in (4) satisfies $\boldsymbol{\kappa}(\boldsymbol{x}) \in [0,1]^d$, we have $\mathbf{1}^\intercal\boldsymbol{\kappa}(\boldsymbol{x}) \leq d$. Next observe that for all $t \geq 0$, $t/[(1+t)\log(1+t)] \leq \mathrm{e}$. Therefore for $\boldsymbol{x} \in \mathbb{R}^d_{++}$,

$$\prod_{i=1}^d \tilde{J}_i(\boldsymbol{x}) \leq \prod_{i=1}^d \left[1 + \frac{\rho^{-1}\log(u/l)}{\log(1+|\boldsymbol{x}|_i)}\frac{|x_i|}{1+|x_i|}\right] \leq \left(1 + \mathrm{e}\rho^{-1}\log(u/l)\right)^d.$$

Moreover, $\max\{\tilde{J}_i(\boldsymbol{x}), \ldots, \tilde{J}_d(\boldsymbol{x})\} \geq 1$. Combining these observations we obtain that

$$J(\boldsymbol{x}) \leq \left[1 + \mathrm{e}\rho^{-1}\log(u/l)\right]^d (u/l)^d, \quad \boldsymbol{x} \in \mathbb{R}^d_{++}. \tag{EC.18}$$

To bound the terms involving hazard rates $\lambda_i(\cdot)$, we proceed as follows: Due to Assumption 4, $\lambda_i(\cdot)$ is eventually monotone for any i. From Lemma EC.4, if λ_i is eventually decreasing, $\lambda_i(\boldsymbol{T}_i^{-1}(\boldsymbol{q}(t(u)\boldsymbol{p})) > \lambda_i(q_i(t(u)p_i))$. If λ_i is eventually increasing, the bound $\boldsymbol{T}^{-1}(\boldsymbol{q}(t(u)\boldsymbol{p})) \geq \boldsymbol{q}(t(u)\boldsymbol{p})c_\rho(u)$ from (EC.12) implies $\lambda_i(\boldsymbol{T}_i^{-1}(q_i(t(u)p_i))) \geq \lambda_i(q_i(t(u)p_i)c_\rho(u))$. In either case,

$$\log\lambda_i\big(\boldsymbol{T}_i^{-1}(\boldsymbol{q}(t(u)\boldsymbol{p}))\big) - \log\lambda_i\big(q_i(t(u)p_i)\big) \geq \log\lambda_i\big(q_i(t(u)p_i)c_\rho(u)\big) - \log\lambda_i\big(q_i(t(u)p_i)\big). \quad \text{(EC.19)}$$

Since $\Lambda_i \in \mathcal{RV}(\alpha_i)$ and $\lambda_i = \Lambda_i'$ is monotone, $\lambda_i \in \mathcal{RV}(\alpha_i - 1)$ (see (de Haan & Ferreira 2010, Proposition B.1.9 (11))). Given $\varepsilon > 0$, an application of Potter's bounds (de Haan & Ferreira 2010, Proposition B.1.9 (7)) yields $\lambda_i(q_i(t(u)p_i)c_\rho(u))/\lambda_i(q_i(t(u)p_i)) \geq c_\rho(u)^{\alpha_i - 1 + \epsilon}$, for all u sufficiently large. Then from (EC.19), $\log\lambda_i\big(\boldsymbol{T}_i^{-1}(\boldsymbol{q}(t(u)\boldsymbol{p}))\big) - \log\lambda_i\big(q_i(t(u)p_i)\big) \geq (\alpha_i - 1 + \varepsilon)\log c_\rho(u)$, for $i = 1, \ldots, d$. Since $c_\rho(u) := (l/u)^{1/\rho}$, we obtain the following by combining this bound with (EC.18):

$$\inf_{\boldsymbol{p}\in\mathbb{R}_+^d} b_u(\boldsymbol{p}) \geq -\rho^{-1}\frac{\log(u/l)}{t(u)}\sum_{i=1}^{d}(\alpha_i - 1 + \varepsilon) - d\frac{\log\left(1 + e\rho^{-1}\log(u/l)\right)}{t(u)} - d\frac{\log(u/l)}{t(u)} \to 0,$$

where the convergence follows from noting $t(u) := \Lambda_{\min}(u^{1/\rho})$ and $\log(u/l) = o(t(u))$. □

Lemma EC.9. *Suppose that Assumptions 1 and 2 are satisfied. Then for all sufficiently large u, $\mathrm{lev}_1^+(L_u) \subseteq \mathbb{R}_+^d \setminus B_\gamma$, for some $\gamma > 0$.*

Proof. Recall the definition $f_{\mathrm{LD}}(\boldsymbol{y}) := L^*(\boldsymbol{q}^*\boldsymbol{y}^{\mathbf{1}/\boldsymbol{\alpha}})$. The function $f_{\mathrm{LD}}(\cdot)$ is therefore continuous and bounded on the unit sphere $\mathcal{S}^{d-1} \cap \mathbb{R}_+^d$. Since $q_i^* = 0$ if $\alpha_i > \min_j \alpha_j$, we have for all $c > 0$,

$$f_{\mathrm{LD}}(c\boldsymbol{y}) = L^*(\boldsymbol{q}^*(c\boldsymbol{y})^{\mathbf{1}/\boldsymbol{\alpha}}) = L^*(c^{1/\alpha_*}\boldsymbol{q}^*\boldsymbol{y}^{\mathbf{1}/\boldsymbol{\alpha}}) = c^{\rho/\alpha_*}L^*(\boldsymbol{q}^*\boldsymbol{y}^{\mathbf{1}/\boldsymbol{\alpha}}) = c^{\rho/\alpha_*}f_{\mathrm{LD}}(\boldsymbol{y}) > 0,$$

from the homogeneity of L^*. Combining these observations, we obtain that for any $\gamma > 0$, $\sup_{\boldsymbol{y}\in B_\gamma} f_{\mathrm{LD}}(\boldsymbol{y}) < \gamma^{\rho/\alpha_*}M_1$, where $M_1 > \max_{\boldsymbol{y}\in B_1} f_{\mathrm{LD}}(\boldsymbol{y})$. Choosing $\gamma < (2M_1)^{-\alpha_*/\rho}$ ensures $\sup_{\boldsymbol{y}\in B_\gamma} f_{\mathrm{LD}}(\boldsymbol{y}) < 1/2$. Thus $[\Xi_{1,\gamma}(f_{\mathrm{LD}})]^{1+\varepsilon} \cap B_\gamma = \emptyset$, if $\varepsilon > 0$ is chosen suitably small. Then from the inclusion $\mathrm{lev}_1^+(L_u) \cap B_\gamma \subseteq [\Xi_{1,\gamma}(f_{\mathrm{LD}})]^{1+\varepsilon}$ from Corollary EC.1, we have $\mathrm{lev}_1^+(L_u) \cap B_\gamma = \emptyset$, for all u sufficiently large. □

Recall from Lemma EC.6 that the second moment $M_{2,u} = E\left[\exp\left(-t(u)F_u(\boldsymbol{Y}_u)\right)\right]$.

Lemma EC.10. *Suppose that Assumptions 1 - 4 are satisfied and $l(u)$ is taken to be slowly varying in u. Then there exists u_1 sufficiently large such that for all $u > u_1$, $\inf_{\boldsymbol{p}\in\mathbb{R}_+^d} F_u(\boldsymbol{p}) \geq 0$.*

Checking this non-negativity in Lemma EC.10, while is executed along similar lines as in the proofs of Lemma EC.7 - EC.8, is technically more involved. Its proof is therefore given later in Section EC.4, which is devoted to technical results that are repetitive in terms of the key ideas involved. We now prove the key variance reduction result, namely, Theorem 3.

Proof of Theorem 3. From Lemma EC.6, we have $M_{2,u} = E\left[\exp\{-t(u)F_u(\boldsymbol{Y}_u)\}\right]$. Define the function $F:\mathbb{R}_+^d \to \mathbb{R}\cup\{+\infty\}$ as,

$$F(\boldsymbol{p}) := I(\boldsymbol{p}) + \chi_{\mathrm{lev}_1^+(f_{\mathrm{LD}})}(\boldsymbol{p}).$$

Since $\mathrm{lev}_1^+(f_{\mathrm{LD}})$ is closed and $I(\cdot)$ is continuous, $F(\cdot)$ is lower semi-continuous. Consider sequences $\{u_n\}\subseteq\mathbb{R}_+, \{\boldsymbol{p}_n\}\subseteq\mathbb{R}^d$ satisfying $u_n\to\infty$ and $\boldsymbol{p}_n\to\boldsymbol{p}$. Due to Lemma EC.9, there exists $\gamma, n_0>0$ such that $\mathrm{lev}_1^+(L_{u_n})\cap B_\gamma=\varnothing$, for all $n>n_0$. Suppose $\boldsymbol{p}\notin B_{\gamma/2}$. Then from the uniform convergences of $a_u(\cdot), b_u(\cdot)$ in Lemma EC.7 - EC.8 and that of $\chi_{\mathrm{lev}_1^+(L_u)}(\cdot)$ in Corollary EC.1,

$$\liminf_{n\to\infty} F_{u_n}(\boldsymbol{p}_n) := \liminf_{n\to\infty} a_{u_n}(\boldsymbol{p}_n) + b_{u_n}(\boldsymbol{p}_n) - 2d\frac{\log t(u_n)}{t(u_n)} + \chi_{\mathrm{lev}_1^+(L_{u_n})}(\boldsymbol{p}_n) \geq F(\boldsymbol{p}). \tag{EC.20}$$

On the other hand, if $\boldsymbol{p}\in B_{\gamma/2}$, we have $\{\boldsymbol{p}_n: n\geq n_1\}\subseteq B_\gamma$ for some $n_1>n_0$. Since $\mathrm{lev}_1^+(L_{u_n})\cap B_\gamma = \varnothing$ for all $n>n_1$, we obtain $\inf_{n\geq n_1} F_{u_n}(\boldsymbol{p}_n)=\infty$. Thus, regardless of the membership of $\boldsymbol{p}$ (in the ball $B_{\gamma/2}$), (EC.20) holds. From Lemma EC.10, we deduce that there exists $n_2>n_1$ satisfying that the family $\{F_{u_n}: n\geq n_2\}$ comprises non-negative valued functions. Recall from Theorem 1 that the sequence $\{\boldsymbol{Y}_{u_n}: n\geq 1\}$ satisfies LDP with rate function $I(\cdot)$. Then due to a general version of Varadhan's integral lemma (see (Varadhan 1988, Theorem 2.2)), we obtain from (EC.20) that

$$\limsup_{n\to\infty}\frac{1}{t(u_n)}\log E\left[\exp\left\{-t(u_n)F_{u_n}(\boldsymbol{Y}_{u_n})\right\}\right] \leq -\inf_{\boldsymbol{p}\in\mathbb{R}^d}\{F(\boldsymbol{p})+I(\boldsymbol{p})\} = -2\inf_{\boldsymbol{p}\in\mathrm{lev}_1^+(f_{\mathrm{LD}})} I(\boldsymbol{p}).$$

Since $M_{2,u}=E\left[\exp\{-t(u)F_u(\boldsymbol{Y}_u)\}\right]$, combining this conclusion with the bounds $M_{2,u}\geq p_u^2$ and $\liminf_{u\to\infty}[t(u)^{-1}\log p_u^2]\geq -2I^*$ from Theorem 2, we obtain (23). □

EC.1.4. Proofs of Propositions 2-3 and Corollaries 1-2

Proof of Proposition 2: Recall that $\boldsymbol{Y}_u=\boldsymbol{Y}/t(u)$. Observe that $\boldsymbol{\Lambda}(\boldsymbol{Z}_u^*)/t(u)$ has the distribution of $\boldsymbol{Y}_u$ given $L_u(\boldsymbol{Y}_u)\geq 1$. Now, notice that for any $\boldsymbol{p}$, $\delta>0$,

$$P(\boldsymbol{Y}_u\in B_\delta(\boldsymbol{p})\mid\{L_u(\boldsymbol{Y}_u)\geq 1\}) = \frac{P(\boldsymbol{Y}_u\in B_\delta(\boldsymbol{p})\cap\{L_u(\boldsymbol{Y}_u)\geq 1\})}{P(L_u(\boldsymbol{Y}_u)\geq 1)} \tag{EC.21}$$

The numerator in (EC.21) can be upper bounded invoking the LDP for $\boldsymbol{Y}_u$:

$$\inf_{\delta>0}\limsup_{u\to\infty}\frac{1}{t(u)}\log P(\boldsymbol{Y}_u\in B_\delta(\boldsymbol{p})\cap\{L_u(\boldsymbol{Y}_u)\geq 1\}) \leq -[I(\boldsymbol{p})+\chi_1(\boldsymbol{p})] \tag{EC.22}$$

To see this, notice that if $\boldsymbol{p}\in\{f_{\mathrm{LD}}(\boldsymbol{p})\geq 1\}$, (EC.22) follows from the continuity of $I(\cdot)$ and the LDP upper bound for $\boldsymbol{Y}_u$. Now suppose, $\boldsymbol{p}\notin\{f_{\mathrm{LD}}(\boldsymbol{p})\geq 1\}$. From the continuous convergence of L_u to f_{LD}, following the proof of Corollary EC.1, there exist δ_0, u_0, such that $B_\delta(\boldsymbol{p})\cap\{L_u(\boldsymbol{Y}_u)\geq 1\} = \emptyset\ \forall\delta<\delta_0$ and $u>u_0$. Thus, for all $\delta<\delta_0, u>u_0$, the probability in (EC.22) evaluates to 0. The bound in (EC.22) now follows. Using the tail asymptotic (Thm. 2) in the denominator of (EC.21),

$$\inf_{\delta>0}\limsup_{u\to\infty}\frac{1}{t(u)}\log P(\boldsymbol{Y}_u\in B_\delta(\boldsymbol{p})\mid\{L_u(\boldsymbol{Y}_u)\geq 1\}) \leq -[I(\boldsymbol{p})-I^*+\chi_1(\boldsymbol{p})]$$

Fix an arbitrary $\varepsilon > 0$ and let $M > 0$ sufficiently large so that $B_1(\boldsymbol{p}) \subset B_M$. Recall that as a consequence of Corollary EC.1, $\Xi_{1+\varepsilon,M}(L_u) \subseteq \Xi_{1,M}(f_{\mathrm{LD}}) \subseteq \mathrm{lev}_1^+(L_u)$ for all u large enough. Then,

$$P(\boldsymbol{Y}_u \in B_\delta(\boldsymbol{p}) \cap \{L_u(\boldsymbol{Y}_u) \geq 1\}) \geq P(\boldsymbol{Y}_u \in B_\delta(\boldsymbol{p}) \cap \{\boldsymbol{Y}_u \in \mathrm{Int}(\Xi_{1+\varepsilon,M}(f_{\mathrm{LD}}))\}).$$

Apply the LDP lower bound for $\boldsymbol{Y}_u$ to the left hand side above, and use the continuity of $I(\cdot)$ (refer to the proof of Theorem 1)

$$\inf_{\delta>0} \liminf_{u\to\infty} \frac{1}{t(u)} \log P(\boldsymbol{Y}_u \in B_\delta(\boldsymbol{p}) \cap \{L_u(\boldsymbol{Y}_u) \geq 1\}) \geq -[I(\boldsymbol{p}) + \varepsilon + \chi_{\mathrm{lev}^+_{1+\varepsilon}(f_{\mathrm{LD}})}(\boldsymbol{p})] \tag{EC.23}$$

From Theorem 2, observe that

$$\lim_{u\to\infty} \frac{1}{t(u)} \log P(L_u(\boldsymbol{Y}_u) \geq 1) = -I^*.$$

Noting that $\varepsilon > 0$ in (EC.23) is arbitrary,

$$\inf_{\delta>0} \liminf_{u\to\infty} \frac{1}{t(u)} \log P(\boldsymbol{Y}_u \in B_\delta(\boldsymbol{p}) \mid \{L_u(\boldsymbol{Y}_u) \geq 1\}) \geq -[I(\boldsymbol{p}) - I^* + \chi_1(\boldsymbol{p})]$$

The LDP required in Proposition 2 follows as a consequence (Dembo & Zeitouni 1998, Theorem 4.1.11). To see the last statement observe that as $u \to \infty$,

$$\frac{\boldsymbol{Z}_u^*}{\boldsymbol{q}(t(u))} = \frac{\boldsymbol{q}\left(t(u)\frac{\boldsymbol{\Lambda}(\boldsymbol{Z}_u^*)}{t(u)}\right)}{\boldsymbol{q}(t(u))} \quad \text{and} \quad \frac{\boldsymbol{q}(t(u)\boldsymbol{p}_u)}{\boldsymbol{q}(t(u))} \to \boldsymbol{p}^{1/\boldsymbol{\alpha}} \text{ whenever } \boldsymbol{p}_u \to \boldsymbol{p}.$$

An application of the approximate contraction principle (Dembo & Zeitouni 1998, Theorem 4.2.23) to the sequences $\{\boldsymbol{\Lambda}(\boldsymbol{Z}_u^*)/t(u) : u > 0\}$ shows that $\{\boldsymbol{Z}_u^*/\boldsymbol{q}(t(u)) : u > 0\}$ satisfies an LDP with rate function $I_1(\boldsymbol{p}) = I(\boldsymbol{p}^{\boldsymbol{\alpha}}) - I^* + \chi_1(\boldsymbol{p}^{\boldsymbol{\alpha}})$. The last statement now follows as a consequence of the definition of the LDP (see Dembo & Zeitouni 1998, Section 1.2). □

Proof of Proposition 3: Let $\boldsymbol{q}$, $\boldsymbol{\Lambda}$ and $t(u)$ be defined as before. Under the assumptions of the proposition, uniformly over $\boldsymbol{p}$ in compact sets of $\mathbb{R}_+^d \setminus \{0\}$,

$$\begin{aligned} \frac{\boldsymbol{T}_u \circ \boldsymbol{q}(t(u)\boldsymbol{p})}{\boldsymbol{q}(t(u))} &= [\boldsymbol{p}s^{\alpha_*/\rho}]^{1/\boldsymbol{\alpha}}[1 + o(1)] \implies \\ \boldsymbol{T}_u \circ \boldsymbol{q}(t(u)\boldsymbol{p}) &= \boldsymbol{q}(t(u))([\boldsymbol{p}s^{\alpha_*/\rho}]^{1/\boldsymbol{\alpha}} + o(1)) \implies \\ \frac{\boldsymbol{\Lambda} \circ \boldsymbol{T}_u \circ \boldsymbol{q}(t(u)\boldsymbol{p})}{t(u)} &= \boldsymbol{p}s^{\alpha_*/\rho}[1 + o(1)] \text{ (apply } \boldsymbol{\Lambda} \in \mathcal{RV}(\boldsymbol{\alpha}) \text{ to both the sides of the above).} \end{aligned}$$

For convenience, denote $[t(u)]^{-1}\boldsymbol{\Lambda} \circ \boldsymbol{T}_u \circ \boldsymbol{q}(t(u)\boldsymbol{p}) = \boldsymbol{\phi}_u(\boldsymbol{p})$. Observe that $\boldsymbol{\phi}_u(\cdot)$ converges uniformly over compact sets to $s^{\alpha_*/\rho}\mathrm{Id}$, where Id is the identify function. Then, from the approximate contraction principle (see (Dembo & Zeitouni 1998, Theorem 4.2.23)), and the homogeneity of $I(\cdot)$, $\boldsymbol{\phi}_u(\boldsymbol{Y}_u)$ satisfies an LDP with rate function $I_1(\boldsymbol{x}) = s^{-\alpha_*/\rho}I(\boldsymbol{x})$. Observe next that $[t(u)]^{-1}\boldsymbol{\Lambda}(\boldsymbol{Z}_u)$ has the distribution of $\boldsymbol{\phi}_u(\boldsymbol{Y}_u)$ given $L_u(\boldsymbol{\phi}_u(\boldsymbol{Y}_u)) \geq 1$. Then,

$$P(\boldsymbol{\phi}_u(\boldsymbol{Y}_u) \in B_\delta(\boldsymbol{p}) \mid \{L_u(\boldsymbol{\phi}_u(\boldsymbol{Y}_u)) \geq 1\}) = \frac{P(\boldsymbol{\phi}_u(\boldsymbol{Y}_u) \in B_\delta(\boldsymbol{p}) \cap \mathrm{lev}_1^+(L_u))}{P(\boldsymbol{\phi}_u(\boldsymbol{Y}_u) \in \mathrm{lev}_1^+(L_u))} \tag{EC.24}$$

The limit of the numerator of (EC.24) can be evaluated by invoking the LDP of $\boldsymbol{\phi}_u(\boldsymbol{Y}_u)$ and proceeding as in the proof of Proposition 2, replacing $I(\cdot)$ there by $s^{-\alpha_*/\rho}I(\cdot)$:

$$\begin{aligned}\inf_{\delta>0}\limsup_{u\to\infty}\frac{1}{t(u)}\log P(\boldsymbol{\phi}_u(\boldsymbol{Y}_u)\in B_\delta(\boldsymbol{p})\cap \mathrm{lev}_1^+(L_u)) &= \inf_{\delta>0}\liminf_{u\to\infty}\frac{1}{t(u)}\log P(\boldsymbol{\phi}_u(\boldsymbol{Y}_u)\in B_\delta(\boldsymbol{p})\cap \mathrm{lev}_1^+(L_u))\\ &= -s^{-\alpha_*/\rho}I(\boldsymbol{p})+\chi_1(\boldsymbol{p})\\ &= -s^{-\alpha_*/\rho}[I(\boldsymbol{p})+\chi_1(\boldsymbol{p})]\end{aligned}$$

where the last statement follows since $\chi_1(\boldsymbol{p})$ equals either 0 or $+\infty$. The denominator can be evaluated by noting that

$$\lim_{u\to\infty}\frac{1}{t(u)}\log P(L_u(\boldsymbol{\phi}_u(\boldsymbol{Y}_u))\geq 1) = -\inf_{\boldsymbol{p}\in\mathrm{lev}_1^+(f_{\mathrm{LD}})} s^{-\alpha_*/\rho}I(\boldsymbol{p}) = -s^{-\alpha_*/\rho}I^*.$$

Combining everything together and observing that $s^{-\alpha_*/\rho}t(u)=t(u/s)(1+o(1))$, $\{\boldsymbol{\Lambda}(\boldsymbol{Z}_u)/t(u): u>0\}$ satisfies LDP with rate $I-I^*+\chi_1$ and speed $t(u/s)$ as a consequence of (Dembo & Zeitouni 1998, Theorem 4.1.11). Finally, to verify condition (ii), observe that from the above calculation, $\log P(L(\boldsymbol{T}_u(\boldsymbol{X})\geq u) = -t(u/s)[I^*+o(1)] = [1+o(1)]\log P(L(\boldsymbol{X})\geq u/s)$. □

Proof of Corollary 1: Consider $\{\boldsymbol{p}_{u_n}\}_{n\geq 1}\subset\mathbb{R}_+^d$ such that $\boldsymbol{p}_{u_n}\to\boldsymbol{p}\in\mathbb{R}_+^d\setminus\{\boldsymbol{0}\}$ as $n\to\infty$.

1) Suppose $\boldsymbol{p}>\boldsymbol{0}$. Since with $\boldsymbol{q}\in\mathcal{RV}(1/\boldsymbol{\alpha})$, from the definition of $\boldsymbol{\kappa}(\cdot)$ in (4), with $(u_n/l_n)\to s$, $\lim_{n\to\infty}\boldsymbol{\kappa}(\boldsymbol{q}((t(u_n)\boldsymbol{p}_{u_n}))=\alpha_*/(\rho\boldsymbol{\alpha})$;. which proves the first part of the corollary.

2) When $\boldsymbol{p}>\boldsymbol{0}$ from the definition of $\boldsymbol{T}$ in (4),

$$\boldsymbol{T}_{u_n}\big(\boldsymbol{q}(t(u_n)\boldsymbol{p}_n)\big)=\boldsymbol{q}(t(u_n)\boldsymbol{p}_n)(s+o(1))^{\frac{\alpha_*}{\rho\boldsymbol{\alpha}}}(1+o(1))=\boldsymbol{q}(t(u_n))(s^{\alpha_*/\boldsymbol{\alpha}\rho}\boldsymbol{p}^{1/\boldsymbol{\alpha}})(1+o(1)) \text{ as } n\to\infty.$$

Suppose $\boldsymbol{p}=(p_1,\ldots,p_d)$ is such that the subset of indices $I:=\{i:p_i=0\}$ is non-empty. Since $\boldsymbol{p}\neq\boldsymbol{0}$, I is a strict subset of $\{1,\ldots,d\}$. Since $p_i=0$ for $i\in I$, we have $q_i(t(u_n)p_{i,u_n})=o(q_i(t(u_n)))$. Then, $\boldsymbol{T}_{u_n,i}\big(\boldsymbol{q}(t(u_n))\boldsymbol{p}_{u_n}^{1/\boldsymbol{\alpha}}\big)=o(q_i(t(u_n)))$ for all $i\in I$. Then, the sequence of functions

$$\frac{\boldsymbol{T}_{u_n}(\boldsymbol{q}(t(u_n)\boldsymbol{p}_{u_n})}{\boldsymbol{q}(t(u_n))}\to\boldsymbol{p}^{1/\boldsymbol{\alpha}}s^{\alpha_*/\boldsymbol{\alpha}\rho}=[\boldsymbol{T}_s^*(\boldsymbol{p})]^{1/\boldsymbol{\alpha}}.$$

From (Rockafellar & Wets 1998, Theorem 7.14) and the above continuous, uniformly over compact subsets of $\mathbb{R}_+^d\setminus\{0\}$, $\boldsymbol{T}_u(\boldsymbol{q}(t(u)\boldsymbol{p}))=\boldsymbol{q}(t(u)\boldsymbol{p})s^{\alpha_*/\boldsymbol{\alpha}\boldsymbol{\rho}}$. Thus, the statement in (20) holds uniformly over compact $A\subseteq\mathbb{R}_+^d\setminus\{0\}\}$ and $\{\boldsymbol{T}_u^{(1)}\}_{u>0}$ is therefore rate function preserving. □

Proof of Corollary 2: Suppose that $\boldsymbol{p}_u\to\boldsymbol{p}>0$. Observe that

$$\begin{aligned}\boldsymbol{T}^{(2)}(\boldsymbol{q}(t(u)\boldsymbol{p}_u)) &= \boldsymbol{q}(t(u)\boldsymbol{p}_u)\,(s+o(1))^{\frac{\log(1+\boldsymbol{q}(t(u)\boldsymbol{p}_u))}{\log u-\log(s+o(1))}}\\ &= \boldsymbol{q}(t(u))\boldsymbol{p}^{\boldsymbol{1}/\boldsymbol{\alpha}}[s+o(1)]^{\frac{\log(1+u^{\alpha_*/\boldsymbol{\alpha}\boldsymbol{\rho}})}{\log u}}(1+o(1)) \text{ since } \boldsymbol{q}\in\mathcal{RV}(1/\boldsymbol{\alpha})\\ &= \boldsymbol{q}(t(u))\boldsymbol{p}^{\boldsymbol{1}/\boldsymbol{\alpha}}[s^{\alpha_*/\boldsymbol{\alpha}\rho}+o(1)]\end{aligned}\tag{EC.25}$$

The case where the set $\{i:p_i=0\}\neq\emptyset$ can the handled similar to the proof of Corollary 1. □

Table EC.1 Examples of some marginal distributions satisfying Assumption 2 and their right tail parameter α. Larger the parameter α, the lighter the respective tail is.

Distribution families	Tail parameter α
Exponential, Erlang, Gumbel, Logistic	1
Gamma, Chi-squared, phase-type	1
Gaussian, Chi, mixtures of Gaussians, Rayleigh	2
Weibull with shape parameter k	k
Generalized-gamma with shape parameter k	k

Table EC.2 Examples of some heavy-tailed marginal distributions satisfying Assumption 5 and the respective right tail parameter α. Larger the parameter α, relatively lighter is the respective tail.

Distribution families	Tail parameter γ
Lognormal	2
Generalized Pareto, Student's t, Regularly varying	1
Log-Laplace, Frechet, Lomax, Log-logistic, Cauchy	1

EC.2. Sufficient conditions and examples for the tail models considered

This section serves to illustrate the variety of multivariate distribution families which come under the tail modeling framework considered and to provide sufficient conditions. In Tables EC.1 - EC.2 below, we provide some examples of uni-variate distribution families which satisfy either the marginal assumptions in Assumption 2, or its heavier-tailed counterpart in Assumption 5.

The notion of multivariate regular variation turns out to be convenient in understanding the condition in Assumption 3 in terms of the probability density of $\boldsymbol{X}$. We say that a function $f:\mathbb{R}_+^d \to \mathbb{R}_+$ is *multivariate regularly varying* if for any sequence $\boldsymbol{x}_n$ of $\mathbb{R}_+^d$ satisfying $\boldsymbol{x}_n \to \boldsymbol{x} \neq \boldsymbol{0}$,

$$\lim_{n\to\infty} n^{-1} f(\boldsymbol{h}(n)\boldsymbol{x}_n) = f^*(\boldsymbol{x}), \tag{EC.26}$$

for some limiting $f^*:\mathbb{R}_+^d \to (0,\infty)$ and a component-wise increasing $\boldsymbol{h}(t) = (h_1(t),\dots,h_d(t))$ satisfying $h_i \in \mathcal{RV}(1/\rho_i),\ \rho_i > 0, i = 1,\dots,d$. It follows from (EC.26) that $f^*(\cdot)$ satisfies $f^*(s^{1/\boldsymbol{\rho}}\boldsymbol{x}) = sf^*(\boldsymbol{x})$. In the above, the notation $s^{1/\boldsymbol{\rho}}$ is to be interpreted as the vector $s^{1/\boldsymbol{\rho}} = (s^{1/\rho_1},\dots,s^{1/\rho_d})$. When referring to (EC.26), we write $f \in \mathcal{MRV}$, or more specifically, $f \in \mathcal{MRV}(f^*,\boldsymbol{h})$ if there is a need to explicitly specify the scaling functions $\boldsymbol{h}(\cdot)$ and the respective limit function $f^*(\cdot)$.

For instance, the function $f:\mathbb{R}_+^2 \to \mathbb{R}_+$ defined as in, $f(\boldsymbol{x}) = x_1^{2.5}(1-\exp(-x_2)) + x_2^{0.5}$, satisfies $f \in \mathcal{MRV}(f^*,\boldsymbol{h})$ with $f^*(\boldsymbol{x}) = x_1^{2.5} + x_2^{0.5}$ and $\boldsymbol{h}(t) = (t^{1/2.5}, t^2)$. See Table EC.3 below for some useful examples which arise in the context of tail modeling and Resnick (2007) for a detailed treatment. The following result on $\mathcal{MRV}$ functions is required in the subsequent proofs. Let $\mathrm{Id}(\boldsymbol{x}) = \boldsymbol{x}$.

Lemma EC.11. *Suppose $\boldsymbol{g}:\mathbb{R}_+^d \to \mathbb{R}_+^d$ is such that $g_i \in \mathcal{RV}(\beta_i)$ is monotone, for $i \in \{1,\dots,d\}$. For $s>0$, define $\tilde{\boldsymbol{g}}_s(t) := \boldsymbol{g}(st)$ and $\boldsymbol{v}_{s,\boldsymbol{\beta}}(\boldsymbol{x}) := s\boldsymbol{x}^{1/\boldsymbol{\beta}}$. Then for $f:\mathbb{R}_+^d \to \mathbb{R}$, we have $f\circ \boldsymbol{g} \in \mathcal{MRV}(u^*,\mathrm{Id})$, so long as $f \in \mathcal{MRV}(u^* \circ \boldsymbol{v}_{s,\boldsymbol{\beta}}, \tilde{\boldsymbol{g}}_s)$ for some $s>0$.*

Proof. Consider any $M > 0$. Since $g_i \in \mathcal{RV}(\beta_i), i = 1, \ldots, d$, and are monotone, we have $g_i(tx)/g_i(t) \to x^{\beta_i}$, uniformly over $x \in [0, M]$. Consequently for $s > 0$ and $\boldsymbol{x}_n \to \boldsymbol{x} \in [0, M]^d$, we obtain $\boldsymbol{g}(sn \cdot s^{-1}\boldsymbol{x}_n)/\boldsymbol{g}(sn) = (s^{-1}\boldsymbol{x})^{\boldsymbol{\beta}}(1 + o(1))$. Therefore,

$$n^{-1}(f \circ \boldsymbol{g})(n\boldsymbol{x}_n) = n^{-1} f\left(\frac{\boldsymbol{g}(ns \cdot s^{-1}\boldsymbol{x}_n)}{\boldsymbol{g}(ns)}\boldsymbol{g}(ns)\right) = \frac{f\left((s^{-1}\boldsymbol{x})^{\boldsymbol{\beta}}(1 + o(1))\boldsymbol{g}(ns)\right)}{n} = u^*(\boldsymbol{x})(1 + o(1)),$$

where the last equality follows from $f \in \mathcal{MRV}(u^* \circ \boldsymbol{v}_{s,\boldsymbol{\beta}}, \tilde{\boldsymbol{g}}_s)$. □.

Suppose that the support of $\boldsymbol{X}$ contains $\mathbb{R}^d_+$. Propositions EC.1 - EC.2 below give sufficient conditions under which Assumption 3 is satisfied.

Proposition EC.1 **(**Sufficient conditions on the density of $\boldsymbol{X}$**)**. *Suppose that the marginal distributions of $\boldsymbol{X}$ satisfy either Assumptions 2 and 4, and the density of $\boldsymbol{X}$ when written in the form,*

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \exp(-\psi(\boldsymbol{x})), \qquad \textit{for } \boldsymbol{x} \in \mathbb{R}^+_d, \tag{EC.27}$$

satisfies $\psi \in \mathcal{MRV}(\psi^*, \boldsymbol{h})$. Then $\boldsymbol{X}$ satisfies Assumptions 3. In particular, the hazard functions $\boldsymbol{\Lambda} = (\Lambda_1, \ldots, \Lambda_d)$ in Assumption 2 and the limiting function $I(\cdot)$ in Assumption 3 are related to $\boldsymbol{h}$ and ψ^* as follows: there exists $\boldsymbol{c} \in \mathbb{R}^d_{++}$ such that

$$I(\boldsymbol{x}) = \psi^*(\boldsymbol{c}\boldsymbol{x}^{\mathbf{1}/\boldsymbol{\alpha}}) \qquad \textit{and} \qquad \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{q}(\boldsymbol{x})(\boldsymbol{c}^{-1} + o(1)),$$

as $\|\boldsymbol{x}\| \to \infty$, and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d) \in \mathbb{R}^d_{++}$ is such that $h_i \in \mathcal{RV}(1/\alpha_i), i = 1, \ldots, d$.

Proof. Since $\boldsymbol{Y} = \boldsymbol{\Lambda}(\boldsymbol{X})$ and $\boldsymbol{X} = \boldsymbol{q}(\boldsymbol{Y})$, we have (12) satisfied with $\varphi(\boldsymbol{y}) = \psi(\boldsymbol{q}(\boldsymbol{y}))$ and $p(\boldsymbol{y}) = \prod_i [\lambda_i(q_i(y_i))]^{-1}$, as a consequence of change of variables. Observe that $\Lambda_i(t) = \int_{-\infty}^t \lambda_i(x)dx$, for monotone λ_i. Therefore using (de Haan & Ferreira 2010, Proposition B.1.9(11)), $t\lambda_i(t)/\Lambda_i(t) \to \alpha_i$ and $\lambda_i \in \mathcal{RV}(\alpha_i - 1)$. This implies $\log \lambda_i \circ q_i \in \mathcal{RV}(0)$. Since the support of $\boldsymbol{X}$ contains $\mathbb{R}^d_+$, whenever $\boldsymbol{y}_n \to \boldsymbol{y} \in \mathbb{R}^+_d$, $n^{-\epsilon} \log p(n\boldsymbol{y}_n) = o(1)$ for all $\epsilon > 0$. Therefore Assumption 3 holds if $\varphi \in \mathcal{MRV}(I, \mathrm{Id})$.

To see the latter, recall that $\varphi(\boldsymbol{y}) = \psi(\boldsymbol{q}(\boldsymbol{y}))$. Substituting $\boldsymbol{g}(\boldsymbol{y}) = \boldsymbol{q}(\boldsymbol{y})$ and $\boldsymbol{\beta} = \mathbf{1}/\boldsymbol{\alpha}$ in Lemma EC.11, a sufficient condition for $\varphi \in \mathcal{MRV}(I, \mathrm{Id})$ is that $\psi \in \mathcal{MRV}(I \circ \boldsymbol{v}_{s,\mathbf{1}/\boldsymbol{\alpha}}, \boldsymbol{q}_s)$, for some $s > 0$. Under the proposition assumptions, $\psi \in \mathcal{MRV}(\psi^*, \boldsymbol{h})$. Equating parameters, $I(\boldsymbol{x}) = \psi^*(\boldsymbol{c}\boldsymbol{x}^{\mathbf{1}/\boldsymbol{\alpha}})$, and $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{q}(\boldsymbol{x})(\boldsymbol{c}^{-1} + o(1))$, where $\boldsymbol{c} = s^{-\mathbf{1}/\boldsymbol{\alpha}} \in \mathbb{R}^d_{++}$. □

To identify sufficient conditions in the presence of heavier tailed distributions, let $\mathcal{L}$ denote the collection of indices of the components $(X_1, \ldots, X_d)$ which satisfy the lighter tailed assumption in Assumption 2. For $i \notin \mathcal{L}$, we have the respective X_i satisfying the heavier tailed assumption in Assumption 5. Let $\boldsymbol{Z} = (Z_1, \ldots, Z_d)$ be defined as follows:

$$Z_i = \begin{cases} \log(1 + X_i) & \text{if } X_i > 0 \text{ and } i \notin \mathcal{L}, \\ X_i & \text{otherwise.} \end{cases} \tag{EC.28}$$

PROPOSITION EC.2 (Sufficient conditions in the presence of heavier tails). *Suppose that the marginal distributions of $\boldsymbol{X}$ satisfy Assumptions 4 and 5. Let the probability density of $\boldsymbol{Z} = (Z_1, \ldots, Z_d)$ in (EC.28), when written in the form,*

$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = \exp(-\hat{\psi}(\boldsymbol{z})), \qquad \textit{for } \boldsymbol{z} > 0,$$

satisfy $\hat{\psi} \in \mathcal{MRV}(\psi^*, \boldsymbol{h})$. Then $\boldsymbol{X}$ satisfies Assumption 3. In particular, hazard function $\boldsymbol{\Lambda}$ and the limiting $I(\cdot)$ in Assumption 3 are related to $\boldsymbol{h}$ and ψ^* as follows: there exists $\boldsymbol{c} \in \mathbb{R}^d_{++}$ such that

$$I(\boldsymbol{x}) = \psi^*(\boldsymbol{c}\boldsymbol{x}^{1/\boldsymbol{\alpha}}) \qquad \textit{and} \qquad h_i(x_i) = \begin{cases} q_i(x_i)(c_i^{-1} + o(1)) & \textit{if } i \in \mathcal{L}, \\ \log(q_i(x_i))(c_i^{-1} + o(1)) & \textit{otherwise}, \end{cases}$$

as $\|\boldsymbol{x}\| \to \infty$, and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d) \in \mathbb{R}^d_{++}$ is such that $h_i \in \mathcal{RV}(1/\alpha_i), i = 1, \ldots, d$.

Proof of Proposition EC.2. For $i \in \{1, \ldots, d\}$, let $\tilde{\Lambda}_i$ and $\tilde{\lambda}_i$ denote the hazard function and hazard rate of Z_i, respectively. Let $\tilde{q}_i := \tilde{\Lambda}_i^{\leftarrow}$. To rewrite the density of $\boldsymbol{Y}$ in terms of that of $\boldsymbol{Z}$, see that $\tilde{\Lambda}_i(z) = \Lambda_i(\exp(z) - 1)$, when $i \notin \mathcal{L}$. Using a change of variables,

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \frac{1}{\prod_{i=1}^d \tilde{\lambda}_i(\tilde{q}_i(y_i))} \exp\left(-\hat{\psi}(\tilde{\boldsymbol{q}}(\boldsymbol{y}))\right).$$

Recall that under the assumptions of the proposition, $\tilde{\Lambda}_i \in \mathcal{RV}(\alpha_i)$ and the support of $\boldsymbol{Z}$ contains $\mathbb{R}^+_d$. Since $\tilde{\Lambda}_i(x) \sim \Lambda_i(\exp(x))$ as $x \to \infty$, we obtain the desired conclusion following the steps in the proof of Proposition EC.1 with $p(\boldsymbol{y}) = 1/\prod_{i=1}^d \tilde{\lambda}_i(\tilde{q}_i(y_i))$ and $\varphi(\boldsymbol{y}) = \hat{\psi}(\tilde{\boldsymbol{q}}(\boldsymbol{y}))$. □

REMARK EC.2. If the density of $\boldsymbol{X}$ is written in the form (EC.27) in the positive orthant, then the respective density for $\boldsymbol{Z}$ in the positive orthant is given by $f_{\boldsymbol{Z}}(\boldsymbol{z}) = \exp(-\hat{\psi}(\boldsymbol{z}))$, where

$$\hat{\psi}(\boldsymbol{z}) := \psi \circ \boldsymbol{E}(\boldsymbol{z}) - \mathbf{1}_{\mathcal{H}}^\intercal \boldsymbol{z},$$

with the map $\boldsymbol{E} : (x_1, \ldots, x_d) \mapsto \big(E_1(x_1), \ldots, E_d(x_d)\big)$ and the vector $\mathbf{1}_{\mathcal{H}} \in \mathbb{R}^d_+$ defined as follows:

$$E_i(x_i) := \begin{cases} x_i & \text{if } i \in \mathcal{L}, \\ \exp(x_i) - 1 & \text{if } i \notin \mathcal{L}, \end{cases} \qquad \text{and} \qquad \mathbf{1}_{\mathcal{H}} = \begin{cases} 0 & \text{if } i \in \mathcal{L}, \\ 1 & \text{if } i \notin \mathcal{L}. \end{cases} \tag{EC.29}$$

Then one can restate the condition in Proposition EC.2, directly in terms of density of $\boldsymbol{X}$, as follows: If $\hat{\psi}(\boldsymbol{z}) = \psi \circ \boldsymbol{E}(\boldsymbol{z}) - \mathbf{1}_{\mathcal{H}}^\intercal \boldsymbol{z} \in \mathcal{MRV}(\psi^*, \boldsymbol{h})$, then the conclusion in Proposition EC.2 holds.

EXAMPLE EC.1 (MULTIVARIATE t DISTRIBUTION). Suppose $\boldsymbol{X}$ is distributed according to multivariate t distribution with density, $f_{\boldsymbol{X}}(\boldsymbol{x}) = c_\rho \exp(-\psi(\boldsymbol{x}))$, where

$$\psi(\boldsymbol{x}) = \frac{\rho + d}{2} \log\left(1 + \frac{\boldsymbol{x}^\intercal \Sigma^{-1} \boldsymbol{x}}{\rho}\right),$$

ρ is a suitable positive integer and c_ρ is the respective normalizing constant. Since the marginals of a multivariate t distribution are heavy-tailed, we have $\boldsymbol{E}(\boldsymbol{x}) = \exp(\boldsymbol{x}) - \mathbf{1}$. With $\boldsymbol{Z} = \log(1+\boldsymbol{X})$, $f_{\boldsymbol{Z}}(\boldsymbol{z}) = \mathrm{e}^{-\|\boldsymbol{z}\|_1} f_{\boldsymbol{X}}(\mathrm{e}^{\boldsymbol{z}} - 1)$, due to change of variables. Then in this case,

$$\hat{\psi}(\boldsymbol{z}) = -\|\boldsymbol{z}\|_1 + \frac{\rho+d}{2}\log\left(1 + \frac{(\mathrm{e}^{\boldsymbol{z}}-1)\Sigma^{-1}(\mathrm{e}^{\boldsymbol{z}}-1)}{\rho}\right).$$

Thus $\hat{\psi} \in \mathcal{MRV}(\psi^*, \boldsymbol{h})$, where $\psi^*(\boldsymbol{z}) := (\rho+d)\|\boldsymbol{z}\|_\infty - \|\boldsymbol{z}\|_1$ and $\boldsymbol{h}(t) = t\mathbf{1}$. □

Example 4 in Section 3 and Example EC.1 above serve as illustrations for the sequence of steps involved in verifying memberships of commonly used multivariate distributions and copula models in the considered tail modeling framework. Table EC.3 below is intended to serve as a reference for identifying the limiting function $I(\cdot)$ in Assumption 3 (or) the respective function $\psi^*(\cdot)$ which arise in the characterizations in Propositions EC.1 - EC.2. Thanks to standardization, the limiting function $I(\cdot)$ is unique despite $\psi^*(\cdot)$ depending on the specific scaling function $\boldsymbol{h}(\cdot)$ employed.

EC.3. Proofs of results on application to credit risk

Let $[J]$ denote the set $\{1,\dots,J\}$. For any $\varepsilon > 0$, $j \in [J], I \subseteq [J]$, and $\boldsymbol{x} \in \mathbb{R}^d_+$, we define,

$$C_{\boldsymbol{x},\varepsilon} := \{i \in [m] : W_{t(i)}(\boldsymbol{x}, \boldsymbol{v}_i) > \gamma(1-\varepsilon)\} \qquad s_\varepsilon(\boldsymbol{x}) := \frac{1}{m}\sum_{i \in C_{\boldsymbol{x},\varepsilon}} e_i, \tag{EC.30}$$

$$e_m(I) := \frac{1}{m}\sum_{i \in [m]: t(i) \in I} e_i, \qquad e_\infty(I) := \sum_{j \in I} \bar{c}_j,$$

$$\underline{w}_j(\boldsymbol{x}) := \min_{i \in [m]: t(i) = j} W_{t(i)}(\boldsymbol{x}, \boldsymbol{v}_i), \text{ and} \qquad e^*_m = \max_{I \notin \mathcal{J}_m} e_m(I). \tag{EC.31}$$

LEMMA EC.12. *$P(\mathcal{E}_m \mid \boldsymbol{X}) \le \exp(-0.5 m\gamma\varepsilon_m e_0^{-1}(q\bar{e}_m - s_{\varepsilon_m}(\boldsymbol{X}))^+ + \exp(-0.5\gamma\varepsilon_m))$, almost surely, where $s_{\varepsilon_m}(\cdot)$ is as defined in* (EC.30).

Proof. For any $\boldsymbol{x} \in \mathbb{R}^d_+$, let $P_{\boldsymbol{x}}(\cdot)$ denote the conditional law of the default variables $(Y_1,\dots,Y_m)$ given $\boldsymbol{X} = \boldsymbol{x}$; let $E_{\boldsymbol{x}}[\cdot]$ denote the associated expectation. For any $\lambda > 0$, we obtain from Markov's inequality, $P_{\boldsymbol{x}}(L_m > q\bar{e}_m) \le \exp(-m\lambda q\bar{e}_m + \log E_{\boldsymbol{x}}[\exp(m\lambda L_m)])$, due to the independence of the default variables $Y_1,\dots,Y_m$ given $\boldsymbol{X}$. Letting $\psi_m(\lambda,\boldsymbol{x}) := \frac{1}{m}\sum_{i=1}^m \log E_{\boldsymbol{x}}[\exp(\lambda e_i Y_i)]$ and $g_m(\boldsymbol{x}) = \sup_{\lambda>0}\{\lambda q\bar{e}_m - \psi_m(\lambda,\boldsymbol{x})\}$, we obtain

$$P_{\boldsymbol{x}}(L_m > q\bar{e}_m) \le \exp(-m g_m(\boldsymbol{x})), \tag{EC.32}$$

as a consequence. With $P_{\boldsymbol{x}}(Y_i = 1)$ given as in (27), we have

$$\begin{aligned}\psi_m(\lambda,\boldsymbol{x}) &= \frac{1}{m}\sum_{i=1}^m \log\left(1 + \frac{\exp(\lambda e_i) - 1}{1 + \exp(\gamma - W_{t(i)}(\boldsymbol{x}, v_i))}\right)\\ &\le \frac{1}{m}\sum_{i \in C_{\boldsymbol{x},\varepsilon}} \log(1 + \exp(\lambda e_i) - 1) + \frac{1}{m}\sum_{i \notin C_{\boldsymbol{x},\varepsilon}} \log\left(1 + \frac{\exp(\lambda e_0) - 1}{1 + \exp(\gamma - \gamma(1-\varepsilon))}\right),\end{aligned}$$

Table EC.3 Some commonly used density families which satisfy Assumptions 2 - 3, along with their limiting functions $I(\cdot)$ and $\psi^*(\cdot)$ (where applicable). Certain constants are written as c or $\mathbf{c}$ (if $\boldsymbol{c}\in\mathbb{R}^d$) to minimize clutter.

Density families	Limiting function $\psi^*(\boldsymbol{x})$	Respective copula family	Limiting $I(\boldsymbol{x})$ in Assumption 3b
Elliptical densities given by $\boldsymbol{X}\overset{d}{=}\boldsymbol{\mu}+\mathcal{R}\Sigma^{1/2}\boldsymbol{U}$:			
1) Multivariate normal: mean $=\boldsymbol{\mu}$, covariance Σ	$\boldsymbol{x}^\intercal\Sigma^{-1}\boldsymbol{x}$	Gaussian copula	$(\boldsymbol{x}^{1/2})^\intercal R^{-1}\boldsymbol{x}^{1/2}$, with $R=$ correlation matrix
2) Multivariate $t-$distribution: $\mathcal{R}\sim F_{d,\rho}$	$(\rho+d)\lVert\boldsymbol{x}\rVert_\infty-\lVert\boldsymbol{x}\rVert_1$	Student-t copula	$(1+\frac{d}{\rho})\lVert\boldsymbol{x}\rVert_\infty-\frac{1}{\rho}\lVert\boldsymbol{x}\rVert_1$
3) $\mathcal{R}$ is light-tailed with p.d.f. $f_\mathcal{R}(r)=\exp(-g(r))$, for $g\in\mathcal{RV}(k)$, $k>0$	$(\boldsymbol{x}^\intercal\Sigma^{-1}\boldsymbol{x})^{k/2}$	Elliptical copula family	$((\boldsymbol{x}^{1/k})^\intercal R^{-1}\boldsymbol{x}^{1/k})^{k/2}$, with $R=$ correlation matrix
4) $\mathcal{R}$ is heavy-tailed with p.d.f. $f_\mathcal{R}(r)=\exp(-g(r))$, for $g\circ\exp\in\mathcal{RV}(k)$, $k>1$	$\lVert\boldsymbol{x}\rVert_\infty^k$	Elliptical copula family	$\lVert\boldsymbol{x}\rVert_\infty$
Exponential family with p.d.f. $f_{\boldsymbol{X}}(\boldsymbol{x})\propto g(\boldsymbol{x})\exp\big(\boldsymbol{\eta}^\intercal\boldsymbol{T}(\boldsymbol{x})\big)$			
Minimal, light-tailed: $\boldsymbol{\eta}^\intercal\boldsymbol{T}\in\mathcal{MRV}(T_\eta^*,n\mathbf{1})$ for some $T_\eta^*(\cdot)$, $g\in\mathcal{RV}$	$T_\eta^*(\boldsymbol{x})$	-	$T_\eta^*(\mathbf{c}\boldsymbol{x}^{1/k})$ with k as in $f\in\mathcal{RV}(k)$ for $f(n)=\boldsymbol{\eta}^\intercal\boldsymbol{T}(n\mathbf{1})$
Generalized linear models: $\boldsymbol{\xi}=(\boldsymbol{X},\boldsymbol{Y})$ with $f_{\boldsymbol{Y}\mid\boldsymbol{X}}(\boldsymbol{y}\mid\boldsymbol{x})\propto b(\boldsymbol{y})\exp\big(\boldsymbol{\ell}^{-1}(\boldsymbol{\beta}^T\boldsymbol{x})^\intercal T_{\boldsymbol{y}}(\boldsymbol{y})\big)$ $\boldsymbol{X}$ with p.d.f in exponential family			
$b\in\mathcal{MRV}$, $f_{\boldsymbol{x}}(\boldsymbol{x})=\mathrm{e}^{-\psi(\boldsymbol{x})}$ $\psi+u\in\mathcal{MRV}(\psi^*,(h,r))$ $u(\boldsymbol{x},\boldsymbol{y})=\boldsymbol{\ell}^{-1}(\boldsymbol{\beta}^T\boldsymbol{x})^\intercal T_{\boldsymbol{y}}(\boldsymbol{y})$, $h_i\in\mathcal{RV}(\alpha_i)$, $r_i\in\mathcal{RV}(\beta_i)$	$\psi^*(\boldsymbol{x},\boldsymbol{y})$	-	$\psi^*\circ\boldsymbol{\pi}_1^{-1}$, with $\boldsymbol{\pi}_1(\boldsymbol{x})=(\boldsymbol{x}^{\boldsymbol{\alpha}},\boldsymbol{y}^{\boldsymbol{\beta}})$
Logconcave densities with p.d.f. $f_{\boldsymbol{X}}(\boldsymbol{x})=\exp(-\psi(\boldsymbol{x}))$			
convex $\psi\in\mathcal{MRV}$	$\psi^*(\cdot)$ limit with scaling $\boldsymbol{h}(t)=\boldsymbol{\Lambda}^{-1}(t)$		$c\psi^*\circ\boldsymbol{\pi}^{-1}$
Archimedian copula family with $C(\boldsymbol{u})=\ \phi^{-1}(\phi(u_1)+\ldots+\phi(u_d))$			
-	-	Gumbel $\phi(u)=(-\log u)^\theta$	$\lVert\boldsymbol{x}\rVert_\theta$
-	-	Clayton, $\phi(u)=\frac{(t^{-\theta}-1)}{\theta}$	$(1+\theta d)\lVert\boldsymbol{x}\rVert_\infty-\theta\lVert\boldsymbol{x}\rVert_1$
-	-	Independence	$\lVert\boldsymbol{x}\rVert_1$
Mixtures of K normal variables with covariances $\Sigma_1,\ldots,\Sigma_k$	$\min_{i=1}^K\boldsymbol{x}^\intercal\Sigma_k^{-1}\boldsymbol{x}$	-	$(\boldsymbol{x}^{1/2})^\intercal R_{k^*}^{-1}\boldsymbol{x}^{1/2}$, where k^* minimizes $\{(\boldsymbol{x}^{1/2})^\intercal R_k^{-1}\boldsymbol{x}^{1/2}:k\le K\}$

μ,Σ are the location, scale parameters, $\boldsymbol{U}$ is uniformly distributed on the unit sphere in $\mathbb{R}^d$ and is independent of $\mathcal{R}$; includes the special cases of factor models if $\Sigma=\Gamma^\intercal\Gamma+\sigma^2\mathbf{I}_d$ for some factor matrix $\Gamma\in\mathbb{R}^{k\times d}$ with $k<d$, and graphical models where $\mathcal{R}$ is Gaussian and the inverse covariance matrix is sparse.

where we have used that $W_{t(i)}(\boldsymbol{x}, v_i) \leq \gamma(1-\varepsilon)$ for every $i \notin C_{\boldsymbol{x},\varepsilon}$ and that $e_i \leq (0, e_0]$. Then,

$$\psi_m(\lambda, \boldsymbol{x}) \leq \lambda s_\varepsilon(\boldsymbol{x}) + \log\left(1 + \exp(\lambda e_0 - \gamma\varepsilon)\right),$$

from the definition of $s_\varepsilon(\boldsymbol{x})$. If $q\bar{e}_m > s_\varepsilon(\boldsymbol{x})$, we obtain a lower bound for $g_m(\boldsymbol{x})$ by considering the specific value $\lambda = \lambda_m^*(\boldsymbol{x}) := 0.5e_0^{-1}\gamma\varepsilon$ as below:

$$\begin{aligned} g_m(\boldsymbol{x}) &\geq \lambda_m^*(\boldsymbol{x}) q\bar{e}_m - \lambda_m^*(\boldsymbol{x}) s_\varepsilon(\boldsymbol{x}) - \log\left(1 + \exp(\lambda_m^*(\boldsymbol{x}) e_0 - \gamma\varepsilon)\right) \\ &\geq 0.5e_0^{-1}\gamma\varepsilon\left(q\bar{e}_m - s_\varepsilon(\boldsymbol{x})\right) - \exp(-0.5\gamma\varepsilon). \end{aligned}$$

If $q\bar{e}_m \leq s_\varepsilon(\boldsymbol{x})$, a trivial bound $g_m(\boldsymbol{x}) \geq 0$ is obtained by letting $\lambda = \lambda_m^*(\boldsymbol{x}) = 0$. Thus, $g_m(\boldsymbol{x}) \geq 0.5e_0^{-1}\gamma\varepsilon\left(q\bar{e}_m - s_\varepsilon(\boldsymbol{x})\right)^+ - \exp(-0.5\gamma\varepsilon_m)$. Combining this with (EC.32) yields the desired result. □

Recall from the definitions that $\mathcal{J}_m := \{I \subseteq [J] : e_m(I) \geq q\bar{e}_m\}$ and $\mathcal{J} := \{I \subseteq [J] : e_\infty(I) \geq q\bar{e}\}$.

Lemma EC.13. *There exists a positive integer m_0 such that $\mathcal{J}_m = \mathcal{J}$ for all $m > m_0$. Consequently,*

$$\inf_{m>m_0}\left(q\bar{e}_m - e_m^*\right) > 0 \quad \text{and} \quad \inf_{m>m_0, I\in\mathcal{J}_m}\left(e_m(I) - q\bar{e}_m\right) > 0.$$

Proof. Notice that under the model assumptions stated in Section 6, $\bar{e}_m \to \bar{e}$ and $q\bar{e} \notin \{e_\infty(I) : I \subseteq [J]\}$. The latter implies that there exists some $\delta_1 > 0$ such that

$$\min_{I\subseteq[J]} |e_\infty(I) - q\bar{e}| > \delta_1. \tag{EC.33}$$

Further, for all $j \in [J]$, $m^{-1}\sum_{i:t(i)=j} e_i \to \bar{c}_j$. Consider any $\delta \in (0, \delta_1/2)$. Due to these convergences, there exists m_0 suitably large such that for all $m > m_0$,

$$e_m(I) \in (e_\infty(I) - \delta, e_\infty(I) + \delta), \text{ and } \bar{e}_m \in (\bar{e} - \delta, \bar{e} + \delta). \tag{EC.34}$$

uniformly over $I \subseteq [J]$. Since $q \in (0,1)$, the above bounds imply that $e_\infty(I) \geq q\bar{e} - 2\delta$ for any $I \in \mathcal{J}_m$. With $\delta < \delta_1/2$, $e_\infty(I) > q\bar{e} - \delta_1$, or equivalently, $e_\infty(I) \geq q\bar{e}$ (due to (EC.33)). Therefore $\mathcal{J}_m \subseteq \mathcal{J}$ for all $m > m_0$. Similarly if $I \in \mathcal{J}$ is such that $e_\infty(I) \geq q\bar{e}$, we automatically have $e_\infty(I) \geq q\bar{e} + \delta_1$ (due to (EC.33)). Similarly from (EC.34), $e_m(I) \geq q\bar{e}_m + \delta_1 - 2\delta \geq q\bar{e}_m$. Therefore $\mathcal{J} \subseteq \mathcal{J}_m$ for all $m > m_0$. Combining the two inclusions result in the desired conclusion that $\mathcal{J} = \mathcal{J}_m$.

With e_m^* defined as in (EC.31) and $\mathcal{J}_m = \mathcal{J}$ for all $m > m_0$, we have $e_m^* = \max_{I\notin\mathcal{J}} e_m(I)$ for all $m > m_0$. Again for $\delta \in (0, \delta_1/2)$, we obtain from (EC.34) that $e_m^* = \max_{I\notin\mathcal{J}} e_m(I) < \max_{I\notin\mathcal{J}} e_\infty(I) + \delta_1/2$ for all $m > m_0$. However $\max_{I\notin\mathcal{J}} e_\infty(I) < q\bar{e} - \delta_1$ due to (EC.33).Therefore $e_m^* < q\bar{e} - \delta_1/2$. Since $q\bar{e} < q\bar{e}_m + \delta$, we arrive at the conclusion that $q\bar{e}_m - e_m^* > \delta_1/2 - \delta$, for all $m > m_0$. Recalling that $\delta < \delta_1/2$, the first inequality in the lemma statement stands verified. Observing that $e_m(I) > q\bar{e}_m$ for any $I \in \mathcal{J}_m$, the second inequality follows from completely analogous arguments. □

Proof of Theorem 6: We treat the terms in the bound, $P(\mathcal{E}_m) \le P(\mathcal{A}_m) + P(\mathcal{E}_m \setminus \mathcal{A}_m)$, separately.

Step 1) To obtain an upper bound for $P(\mathcal{A}_m)$: Define $L^*_{\text{CR}}(\boldsymbol{x}) := \max_{I \in \mathcal{J}} \min_{k \in I} W^*_k(\boldsymbol{q}^* \boldsymbol{x}^{\mathbf{1}/\boldsymbol{\alpha}})$. For any sequence $\{\boldsymbol{x}_n\} \subset \mathbb{R}^d_+$ satisfying $\boldsymbol{x}_n \to \boldsymbol{x} \neq 0$, we first show that $n^{-\rho} L_{\text{CR}}(n\boldsymbol{x}_n) \to L^*_{\text{CR}}(\boldsymbol{x})$. To see this, recall the definition of $L_{\text{CR}}(\boldsymbol{x})$ in (29). From the continuous convergence of $W_i(\cdot, \boldsymbol{v})$ in Assumption 7, $\sup_{I \subset [J], t(i) \in I} |n^{-\rho} W_i(n\boldsymbol{x}_n, \boldsymbol{v}_i) - W^*_{t(i)}(\boldsymbol{x})| \to 0$ as $n \to \infty$. Consequently, for any $I \subset [J]$, $\min_{i:t(i) \in I} n^{-\rho} W_i(n\boldsymbol{x}_n, \boldsymbol{v}_i) \to \min_{j \in I} W^*_j(\boldsymbol{x})$. Since $\mathcal{J}_m = \mathcal{J}$ for all $m > m_0$ (see Lemma EC.13),

$$n^{-\rho} L_{\text{CR}}(n\boldsymbol{x}_n) \to L^*_{\text{CR}}(\boldsymbol{x}), \tag{EC.35}$$

for any $\boldsymbol{x}_n \to \boldsymbol{x} \neq \mathbf{0}$. Since $\mathcal{A}_m := \{L_{\text{CR}}(\boldsymbol{X}) > c(1-\varepsilon_m)m^\eta\}$ and $\varepsilon_m \searrow 0$, this continuous convergence implies that one can apply the asymptotics from Theorem 2 to evaluate $P(\mathcal{A}_m)$ as below:

$$\lim_{m\to\infty} \frac{P(\mathcal{A}_m)}{\Lambda_{\min}(c^{1/\rho} m^{\eta/\rho})} = \lim_{m\to\infty} \frac{\log P(L_{\text{CR}}(\boldsymbol{X}) \ge cm^\eta)}{\Lambda_{\min}(c^{1/\rho} m^{\eta/\rho})} = -I_{\text{CR}}, \tag{EC.36}$$

where I_{CR} is defined as in (31).

Step 2) To show $P(\mathcal{E}_m \setminus \mathcal{A}_m) = o(P(\mathcal{A}_m))$: Recall that $\mathcal{A}_m = \{L_{\text{CR}}(\boldsymbol{X}) > c(1-\varepsilon_m)m^\eta\}$. For any $\boldsymbol{x} \in \mathbb{R}^d_+$ such that $L_{\text{CR}}(\boldsymbol{x}) \le c(1-\varepsilon_m)m^\eta$, we have from the definition of $L_{\text{CR}}(\cdot)$ that,

$$\max_{J \in \mathcal{J}_m} \min_{j \in J} \underline{w}_j(\boldsymbol{x}) \le c(1-\varepsilon_m)m^\eta.$$

For such $\boldsymbol{x}$, the collection $\underline{J} := \{j \in [J] : \underline{w}_j(\boldsymbol{x}) > c(1-\varepsilon)m^\eta\}$ is not a member of $\mathcal{J}_m$. Hence,

$$s_{\varepsilon_m}(\boldsymbol{x}) = \frac{1}{m} \sum_{i \in C_{\boldsymbol{x},\varepsilon_m}} e_i \le \frac{1}{m} \sum_{j \in \underline{J}} \sum_{\{i : t(i) = j\}} e_i \le e^*_m,$$

where e^*_m is defined as in (EC.31). Thus we have from Lemma EC.12 that

$$P(\mathcal{E}_m \setminus \mathcal{A}_m) \le \exp\Big(-0.5 m\gamma\varepsilon_m e_0^{-1} (q\bar{e}_m - e^*_m)^+ + \exp(-0.5\gamma\varepsilon_m) \Big). \tag{EC.37}$$

Recall that $\gamma = cm^\eta(1+o(1))$ and ε_m is chosen such that $m^{\eta(1-\alpha_*/\rho)+1}\varepsilon_m \to \infty$. Therefore,

$$\lim_{m\to\infty} \frac{m\gamma\varepsilon_m}{\Lambda_{\min}(m^{\eta/\rho})} = \lim_{m\to\infty} \frac{m^{1+\eta}\varepsilon_m}{m^{\alpha_*\eta/\rho(1+o(1))}} = \infty.$$

Since $\inf_{m > m_0}(q\bar{e}_m - e^*_m) > 0$ for m_0 large (see Lemma EC.13(b)), we obtain from (EC.37),

$$\limsup_{m\to\infty} \frac{\log P(\mathcal{E}_m \setminus \mathcal{A}_m)}{\Lambda_{\min}(m^{\eta/\rho})} \le \limsup_{m\to\infty} \frac{m\exp(-0.5\gamma\varepsilon_m) - 0.5m\gamma\varepsilon_m e_0^{-1} (q\bar{e}_m - e^*_m)^+}{\Lambda_{\min}(m^{\eta/\rho})} = -\infty. \tag{EC.38}$$

Combining (EC.38) and (EC.36), we arrive at $P(\mathcal{E}_m \setminus \mathcal{A}_m) = o(P(\mathcal{A}_m))$.

Step 3) To obtain a matching lower bound: Choose δ_m such that $\delta_m \searrow 0$ and $\delta_m\gamma \to \infty$. Consider any fixed $\boldsymbol{x} \in B_m := \{L_{\text{CR}}(\boldsymbol{x}) \ge \gamma(1+\delta_m)\}$. Notice that there exists a set $I \in \mathcal{J}_m$ satisfying

$$\min_{i:t(i) \in I} W_{t(i)}(\boldsymbol{x}, v_i) > \gamma(1+\delta_m).$$

For the chosen $\boldsymbol{x}$ and the resulting index set I, we have for all $i \in I$,

$$P(Y_i = 1 \mid \boldsymbol{X} = \boldsymbol{x}) = \frac{1}{1+\exp(\gamma - W_{t(i)}(\boldsymbol{x}, \boldsymbol{v}_i))} \geq \frac{1}{1+\exp(-\gamma\delta_m)} \geq 1-\delta,$$

where $\delta > 0$ is suitably small. The resulting conditional loss for the chosen $\boldsymbol{x}$ is given by,

$$E[L_m \mid \boldsymbol{X} = \boldsymbol{x}] = \frac{1}{m}\sum_{i=1}^{m} e_i P(Y_i = 1 \mid \boldsymbol{X} = \boldsymbol{x}) \geq \frac{1}{m}\sum_{i:t(i)\in I} e_i P(Y_i = 1 \mid \boldsymbol{X} = \boldsymbol{x}) \geq (1-\delta) e_m(I).$$

Since $I \in \mathcal{J}_m$, $\inf_{m>m_0}(e_m(I) - q\bar{e}_m) > 0$ as a consequence of Lemma EC.13. As a result, $E[L_m \mid \boldsymbol{X} = x] \geq q\bar{e}_m + \kappa$, for some $\kappa > 0$. Given a realization of $\boldsymbol{X}$, $L_m = \sum_{i=1}^m e_i Y_i$ is the sum of m independent random variables. So for any $\varepsilon > 0$ and $\boldsymbol{x} \in B_m$,

$$P(L_m \geq q\bar{e}_m \mid \boldsymbol{X} = \boldsymbol{x}) \geq P(L_m \geq E[L_m \mid \boldsymbol{X} = \boldsymbol{x}] - \kappa \mid \boldsymbol{X} = \boldsymbol{x}) \geq 1 - \varepsilon,$$

for all m sufficiently large due to concentration properties of independent sums (see, for example, Cantelli's inequality). Therefore,

$$P(\mathcal{E}_m \cap B_m) = \int_{\boldsymbol{x}\in B_m} P(\mathcal{E}_m \mid \boldsymbol{X} = \boldsymbol{x}) dF(\boldsymbol{x}) \geq \inf_{\boldsymbol{x}\in B_m} P(\mathcal{E}_m \mid \boldsymbol{X} = \boldsymbol{x}) P(B_m) \geq (1-\varepsilon) P(B_m).$$

To conclude the proof, notice that $P(\mathcal{E}_m \cap B_m) \leq P(\mathcal{E}_m) \leq P(\mathcal{A}_m) + o(P(\mathcal{A}_m))$, where $\mathcal{A}_m := \{L_{\text{CR}}(\boldsymbol{x}) \geq cm^\eta(1-\varepsilon_m)\}$ and define $B_m := \{\boldsymbol{x} : L_{\text{CR}}(\boldsymbol{x}) \geq \gamma(1+\delta_m)\}$. Since $\varepsilon_m, \delta_m \to 0$ and $\gamma = cm^\eta(1+o(1))$, $\log P(B_m) = \log P(\mathcal{A}_m)(1+o(1))$ as $m \to \infty$. Combining this with the observation in (EC.36), we obtain the following from $\Lambda_{\min} \in \mathcal{RV}(\alpha_*)$:

$$\log P(\mathcal{E}_m) = -\Lambda_{\min}(c^{1/\rho} m^{\eta/\rho})(I_{\text{CR}} + o(1)) = c^{\alpha_*/\rho}\Lambda_{\min}(m^{\eta/\rho})(I_{\text{CR}} + o(1)). \qquad \square$$

Proof of Proposition 4. First, we write the second moment of the IS estimator as (see (Glasserman *et al.* 2008, Appendix, Pg. 1) for details on how to arrive at the first expression below),

$$\begin{aligned} M_{2,m} &= \mathbb{E}\left[\exp\{-mL_m\lambda_m(\boldsymbol{X}) + m\psi_m(\boldsymbol{X}, \lambda_m(\boldsymbol{X}))\} \frac{f_{\boldsymbol{X}}(\boldsymbol{X})}{f_{\boldsymbol{Z}}(\boldsymbol{X})}\mathbb{I}(\mathcal{E}_m)\right] \\ &\leq \underbrace{\mathbb{E}\left[\frac{f_{\boldsymbol{X}}(\boldsymbol{X})}{f_{\boldsymbol{Z}}(\boldsymbol{X})}\mathbb{I}(\mathcal{A}_m)\right]}_{I_{1,m}} + \underbrace{\mathbb{E}\left[\exp\{-mq\lambda_m(\boldsymbol{X})\bar{e}_m + m\psi_m(\boldsymbol{X}, \lambda_m(\boldsymbol{X}))\} \frac{f_{\boldsymbol{X}}(\boldsymbol{X})}{f_{\boldsymbol{Z}}(\boldsymbol{X})}\mathbb{I}(\mathcal{E}_m \setminus \mathcal{A}_m)\right]}_{I_{2,m}}, \end{aligned}$$

where we use the non-positivity of the term in the exponent and drop $\mathbb{I}(\mathcal{E}_m)$ in $\mathbb{I}(\mathcal{E}_m \cap \mathcal{A}_m)$ to arrive at the term labelled $I_{1,m}$. Note that $I_{1,m}$ equals the second moment of the IS estimator in Algorithm 1 when used for the problem of estimating $P(\mathcal{A}_m)$. Here $\mathcal{A}_m = \{L_{\text{CR}}(\boldsymbol{X}) > u_m\}$ and $u_m := cm^\eta(1-\varepsilon_m) \to \infty$ as $m \to \infty$. Recall the specific choice $u = cm^\eta$ employed in Algorithm 2; in addition, l is taken to be slowly varying in m. Due to the continuous convergence in (EC.35) and these choices of l and u, all the requirements in Theorem 3 stand satisfied under the assumptions stated in Proposition 4. Therefore,

$$\lim_{m\to\infty} \frac{\log I_{1,m}}{\log P(\mathcal{A}_m)^2} = 1, \tag{EC.39}$$

due to Theorem 3. To bound $I_{2,m}$, we obtain the following from the steps leading to (EC.37):

$$I_{2,m} \le \exp\left\{-m\gamma\varepsilon_m e_0^{-1}\left(q\bar{e}_m - e_m^*\right)^+ + \exp(-0.5\gamma\varepsilon_m)\right\} \mathbb{E}\left[\frac{f_{\boldsymbol{X}}(\boldsymbol{X})}{f_{\boldsymbol{X}}(\boldsymbol{T}^{-1}(\boldsymbol{X}))} J(\boldsymbol{T}^{-1}(\boldsymbol{X}))\mathbb{I}(\mathcal{E}_m \setminus \mathcal{A}_m)\right]. \tag{EC.40}$$

Since $u = cm^\eta(1+o(1))$ and l is slowly varying in u, we have $(u/l)^{1/\rho} \le (1+\delta)m^{\eta/\rho(1+\delta)}$ as a consequence of Potter's bounds (de Haan & Ferreira 2010, Proposition B.1.9(7)). Then from the uniform bound for the Jacobian $J(\cdot)$ in (EC.18), we have $J(\boldsymbol{T}^{-1}(\boldsymbol{x})) \le m^K$, for some suitably large constant $K \in (0,\infty)$. Since $f_{\boldsymbol{X}}(\cdot)$ is bounded below on compact subsets of $\mathbb{R}_+^d$, we have $\frac{f_{\boldsymbol{X}}(\boldsymbol{x})}{f_{\boldsymbol{X}}(\boldsymbol{T}^{-1}(\boldsymbol{x}))} \le M$, for some M large enough. Combining these observations with (EC.40),

$$\begin{aligned}\limsup_{m\to\infty} \frac{\log I_{2,m}}{\Lambda_{\min}(m^{\eta/\rho})} &\le \frac{m\exp(-0.5\gamma\varepsilon_m) - m\gamma\varepsilon_m e_0^{-1}\left(q\bar{e}_m - e_m^*\right)^+ + \log M + K\log m + \log P(\mathcal{E}_m\setminus\mathcal{A}_m)}{\Lambda_{\min}(m^{\eta/\rho})} \\ &\le \frac{2m\exp(-0.5\gamma\varepsilon_m) - 2m\gamma\varepsilon_m e_0^{-1}\left(q\bar{e}_m - e_m^*\right)^+ + \log M + K\log m}{\Lambda_{\min}(m^{\eta/\rho})} = -\infty,\end{aligned}$$

where the last inequality follows from (EC.38). Together with (EC.39) and the asymptotic for $\log P(\mathcal{E}_m)$ in the statement of Theorem 6, we thus arrive at

$$\liminf_{m\to\infty} \frac{\log M_{2,m}}{\log P(\mathcal{E}_m)^2} \ge 1.$$

Since the second moment $M_{2,m}$ also satisfies $M_{2,m} \ge P(\mathcal{E}_m)^2$, we obtain the desired limit. □

EC.4. Proofs of Technical Results

Proof of Lemma 1. Since $Y_i := \Lambda_i(X_i)$ and $\Lambda_i(x_i) := \log(1 - F(x_i))$, we obtain $P(Y_i \ge y) = P(\Lambda_i(X_i) \ge y) = P(F_i(X_i) \ge 1 - \mathrm{e}^{-y}) = \mathrm{e}^{-y}$. □

Proof of Lemma 2. a) Observe that $I(\cdot)$ is the limit of continuously converging functions $n^{-1}\varphi(n\boldsymbol{x})$. Therefore, from (Rockafellar & Wets 1998, Theorem 7.14), $I(\cdot)$ is continuous. Parts **b), c), d)** of the lemma statement follow directly as a consequence of Theorem 1 and (de Valk 2016, Proposition 3). □

EC.4.1. Proof of Lemma EC.10

Step 1- Develop an asymptotic upper bound for $\boldsymbol{\psi}_u(t(u)\boldsymbol{p})$: Define the function $G_u(\boldsymbol{x}) = u^{-1}L(\boldsymbol{x}u^{1/\rho})$. Then, under Assumption 1, $\text{lev}_1^+(G_u) \cap B_M \subset [\text{lev}_1^+(L^*) \cap B_M]^{1+\varepsilon}$ for all large enough u. Notice that from the continuity and ρ-homogeneity of the limit L^*, the 1-level set of L^* is disjoint from $B_{\delta_1-\varepsilon}(\mathbf{0})$, for some small enough δ_1,ε. Thus, for all large enough u, $G_u(\boldsymbol{x})\geq 1 \implies \|\boldsymbol{x}\|_\infty \geq \delta_1$ for some $\delta_1>0$. Write

$$L_u(\boldsymbol{p}) = u^{-1}L\left(\frac{\boldsymbol{q}(t(u)\boldsymbol{p})}{u^{1/\rho}}u^{1/\rho}\right) = G_u\left(\frac{\boldsymbol{q}(t(u)\boldsymbol{p})}{u^{1/\rho}}\right).$$

Then for all large enough u, $\text{lev}_1^+(L_u)\subseteq\{\boldsymbol{p}:\|\boldsymbol{q}(t(u)\boldsymbol{p})\|_\infty \geq \delta_1 u^{1/\rho}\}$. Since $(u/l)=o(u)$, for all large enough u, $\|\boldsymbol{q}(t(u)\boldsymbol{p})\|_\infty > 1/c_\rho(u)$. Recall that from Lemma EC.4, whenever $\|\boldsymbol{y}\|_\infty \geq 1/c_\rho(u)$, $\boldsymbol{T}^{-1}(\boldsymbol{y})\leq \boldsymbol{y}[c_\rho(u)]^{\frac{\log \boldsymbol{y}}{\log\|\boldsymbol{y}\|_\infty}}\vee\mathbf{1}$. Therefore for all $u>u_0$ where u_0 is sufficiently large, $\boldsymbol{T}^{-1}(\boldsymbol{q}((t(u)\boldsymbol{p}))\leq \exp(\log\boldsymbol{q}(t(u)\boldsymbol{p})r_u(\boldsymbol{p}))\vee\mathbf{1}$, where

$$r_u(\boldsymbol{p}) = \left(1+\frac{\log c_\rho(u)}{\log\|\boldsymbol{q}(t(u)\boldsymbol{p})\|_\infty}\right)\in(0,1),\ \text{since } \|\boldsymbol{q}(t(u)\boldsymbol{p})\|_\infty > 1/c_\rho(u).$$

From the monotonicity of $\boldsymbol{\Lambda}$, it follows that $\boldsymbol{\psi}_u(t(u)\boldsymbol{p})\leq\boldsymbol{\Lambda}\left(\mathrm{e}^{\log\boldsymbol{q}(t(u)\boldsymbol{p})r_u(\boldsymbol{p})}\right)\vee\boldsymbol{\Lambda}(\mathbf{1})$.

Step 2 - Derive partial upper bounds: To this end, for $k\in[d]$, define the set

$$E_{k,u} = \left\{\boldsymbol{p}:\Lambda_k\left(\mathrm{e}^{\log q_k(t(u)\boldsymbol{p})r_u(\boldsymbol{p})}\right)\geq\Lambda_j\left(\mathrm{e}^{\log q_j(t(u)\boldsymbol{p})r_u(\boldsymbol{p})}\right),\forall j\neq k\right\}\cap\text{lev}_1^+(L_u).$$

This is the set of all $\boldsymbol{p}$, such that the kth component of $\boldsymbol{\Lambda}(\mathrm{e}^{\log\boldsymbol{q}(t(u)\boldsymbol{p})r_u(\boldsymbol{p})})$ is the largest, and therefore achieves the maximum in $\|\boldsymbol{\Lambda}\left(\mathrm{e}^{\log\boldsymbol{q}(t(u)\boldsymbol{p})r_u(\boldsymbol{p})}\right)\|_\infty$. Therefore, $\cup_k E_{k,u}=\text{lev}_1^+(L_u)$ (since the maximum in the $\|\cdot\|_\infty$ norm is achieved for some $k\in[d]$). For $\boldsymbol{p}\in E_{k,u}$, $\|\boldsymbol{\psi}_u(t(u)\boldsymbol{p})\|_\infty\leq\Lambda_k\left(\mathrm{e}^{\log q_k(t(u)\boldsymbol{p})r_u(\boldsymbol{p})}\right)\vee\max_k\Lambda_k(1)$. By definition, over $E_{k,u}$, for all $j\in[d]$,

$$\Lambda_k\left(\mathrm{e}^{\log q_k(t(u)p_k)r_u(\boldsymbol{p})}\right)\geq\Lambda_j\left(\mathrm{e}^{\log q_j(t(u)p_j)r_u(\boldsymbol{p})}\right)\tag{EC.41}$$

Step 3 - Establish a lower bound for $\mathrm{e}^{\log q_k(t(u)p_k)r_u(\boldsymbol{p})}$: Given $\boldsymbol{p}$, let $i(\boldsymbol{p})$ be the index which achieves the maximum in $\boldsymbol{q}(t(u)\boldsymbol{p})$ (if there are multiple, select one arbitrarily). Then, we have that $q_{i(\boldsymbol{p})}(t(u)p_{i(\boldsymbol{p})})\geq\delta_1u^{1/\rho}$. Recall that $c_\rho(u)=(l/u)^{1/\rho}$. Therefore,

$$q_{i(\boldsymbol{p})}(t(u)p_{i(\boldsymbol{p})})c_\rho^{\frac{\log q_{i(\boldsymbol{p})}(t(u)p_{i(\boldsymbol{p})})}{\log\|\boldsymbol{q}(t(u)\boldsymbol{p})\|_\infty}}(u) = q_{i(\boldsymbol{p})}(t(u)p_{i(\boldsymbol{p})})c_\rho(u)\geq\delta_1 l^{1/\rho}.$$

Let κ be such that $\kappa\max_j\alpha_j<\min_j\alpha_j$. Then, notice that for all k

$$\frac{\min_j\alpha_j}{\alpha_k}-\kappa>0.\tag{EC.42}$$

Since $\boldsymbol{\Lambda}\in\mathcal{RV}(\boldsymbol{\alpha})$, an application of (de Haan & Ferreira 2010, Proposition B.1.9 (1)) shows that for all $\kappa>0$, there exists an x_j such that for all $x>x_j$, $\Lambda_k^{-1}(\Lambda_j(x))\geq x^{\alpha_j/\alpha_k-\kappa}$. Further since $l\to\infty$ as $u\to\infty$, there exists a $u_{2,k}$, such that for all $u>u_{2,k}$, for κ selected as in (EC.42), $\delta_1 l^{1/\rho}>\max_j x_j$. Therefore, for all $u>u_{2,k}$, $\Lambda_k^{-1}(\Lambda_j(\delta_1 l^{1/\rho}))\geq(\delta_1 l^{1/\rho})^{\alpha_j/\alpha_k-\kappa}$ for all $j\in[d]$. Now, for all $\boldsymbol{p}\in E_{k,u}$, for all $u>u_0\vee u_1\vee u_{2,k}$,

$$\begin{aligned}\left(\mathrm{e}^{\log q_k(t(u)\boldsymbol{p})r_u(\boldsymbol{p})}\right)&\geq\Lambda_k^{-1}(\Lambda_{i(\boldsymbol{p})}(q_{i(\boldsymbol{p})}(t(u)p_{i(\boldsymbol{p})})))\ \text{since } \boldsymbol{p}\in E_{k,u}\\&\geq\Lambda_k^{-1}(\Lambda_{i(\boldsymbol{p})}(\delta_1 l^{1/\rho}))\geq(\delta_1 l^{1/\rho})^{\frac{\min_j\alpha_j}{\alpha_k}-\kappa}\ \text{by choice } u>\max\{u_1,u_{2,k}\}.\end{aligned}\tag{EC.43}$$

Notice that since $l \to \infty$ as $u \to \infty$, the RHS above can be made arbitrarily large by appropriate choice of u.

Step 4 - Evaluate and combine partial bounds: Fix $\delta > 0$. Observe that due to the monotonicity of Λ_i, (EC.41) implies that for all $\boldsymbol{p} \in E_{k,u}$, for all j, $\Lambda_j^{-1}\left(\Lambda_k\left(\mathrm{e}^{\log q_k(t(u)p_k))r_u(\boldsymbol{p})}\right)\right) \geq \exp(\log q_j(t(u)p_j)r_u(\boldsymbol{p}))$. Next, observe that from an application of (de Haan & Ferreira 2010, Proposition B.1.9 (1)), $\Lambda_j^{-1}(\Lambda_k(x)) \leq x^{\frac{\alpha_k(1+\delta)}{\alpha_j(1-\delta)}}$ whenever $x \geq x_{j,k}$. Now choose u large enough (say $u > u_{3,k}$) so that $(\delta_1 l^{1/\rho})^{\frac{\min_j \alpha_j}{\alpha_k} - \kappa} > \max_j x_{j,k}$. Therefore, for $u > u_0 \vee u_1 \vee u_{2,k} \vee u_{3,k}$, $\inf_{\boldsymbol{p} \in E_{k,u}} \exp(\log(q_k(t(u)p_k)r_u(\boldsymbol{p})) \geq \max_j x_{j,k}$. Then,

$$\exp\left(\frac{\alpha_k(1+\delta)}{\alpha_j(1-\delta)} \log q_k(t(u)p_k)r_u(\boldsymbol{p})\right) \geq \Lambda_j^{-1}\left(\Lambda_k\left(\mathrm{e}^{\log q_k(t(u)p_k))r_u(\boldsymbol{p})}\right)\right), \text{ for all } j \in [d].$$

Thus, for all j,

$$\exp\left(\log q_k(t(u)p_k)r_u(\boldsymbol{p}) - \frac{\alpha_j(1-\delta)}{\alpha_k(1+\delta)} \log q_j(t(u)p_j)r_u(\boldsymbol{p})\right) \geq 1.$$

With $r_u(\boldsymbol{p}) > 0$, this requires that for all j,

$$\frac{\log q_k(t(u)p_k)}{\log q_j(t(u)p_j)} \geq \frac{\alpha_j(1-\delta)}{\alpha_k(1+\delta)}.$$

Upon selecting the j which achieves the maximum in $\|\boldsymbol{q}(t(u)\boldsymbol{p})\|_\infty$, notice that

$$\frac{\log q_k(t(u)p_k)}{\log \|\boldsymbol{q}(t(u)\boldsymbol{p})\|_\infty} \geq \frac{\min \alpha_j(1-\delta)}{\alpha_k(1+\delta)} \text{ and therefore, since } c_\rho(u) = o(1),$$

$$q_k(t(u)p_k)[c_\rho(u)]^{\frac{\log q_k(t(u)p_k)}{\log \|\boldsymbol{q}(t(u)\boldsymbol{p})\|_\infty}} \leq q_k(t(u)p_k)[c_\rho(u)]^{\frac{\min_j \alpha_j(1-\delta)}{\alpha_k(1+\delta)}} \leq q_k(t(u)p_k)\varepsilon^{1/\alpha_k} \text{ where } \varepsilon \text{ is arbitrary.}$$

Therefore, $\Lambda_k(\mathrm{e}^{q_k(t(u)p_k)r_u(\boldsymbol{p})}) \leq \Lambda_k(q_k(t(u)p_k)\varepsilon^{1/\alpha_k}) \leq \varepsilon t(u)p_k$. Now, whenever $u > u_k$, for $\boldsymbol{p} \in E_{k,u}$, $\|\boldsymbol{\psi}_u(t(u)\boldsymbol{p})\|_\infty \leq \varepsilon t(u)p_k \leq \varepsilon t(u)\|\boldsymbol{p}\|_\infty$, uniformly over $\boldsymbol{p} \in E_{k,u}$. By symmetry, for $u > \max_k(u_0 \vee u_1 \vee u_{2,k} \vee u_{3,k})$, uniformly over $\boldsymbol{p} \in \mathrm{lev}_1^+(L_u)$,

$$\|\boldsymbol{\psi}_u(t(u)\boldsymbol{p})\|_\infty \leq \varepsilon t(u)\|\boldsymbol{p}\|_\infty \tag{EC.44}$$

Step 5 - Establish non-negativity: With the bound on $\|\boldsymbol{\psi}_u(t(u)\boldsymbol{p})\|_\infty$ established, from (EC.17) and Lemma EC.8 respectively, for $\varepsilon > 0$ for all large enough u,

$$a_u(\boldsymbol{p}) \geq t(u)\|\boldsymbol{p}\|_\infty(1-\varepsilon), \quad \frac{\log t(u)}{t(u)} \leq \varepsilon \quad \text{and} \quad b_u(\boldsymbol{p}) \geq -\varepsilon t(u).$$

From Lemma EC.9, $\boldsymbol{p} \in \mathrm{lev}_1^+(L_u) \implies \|\boldsymbol{p}\|_\infty > \gamma$. Thus, $F_u(\boldsymbol{p}) \geq 0$ over $\boldsymbol{p} \in \mathrm{lev}_1^+(L_u)$ for all large enough u. Finally, notice that $\chi_{\mathrm{lev}_1^+(L_u)}(\boldsymbol{p}) = \infty$, for $\boldsymbol{p} \notin \mathrm{lev}_1^+(L_u)$. This completes the proof. □

EC.4.2. Proof of log-efficiency in the presence of heavier tails

Recall that $\mathcal{L}$ is the collection of indices of components $(X_1,\dots,X_d)$ which satisfy the lighter tailed assumption in Assumption 2. Then $\mathcal{H} := \{1,\dots,d\}\setminus\mathcal{L}$ denotes the heavy-tailed components.

Proof of Theorem 4. Under Assumption 5, $\bar\Lambda_i \in \mathcal{RV}(\alpha_i)$ for $i\in\mathcal{H}$. Its respective inverse is,

$$\bar q_i := \bar\Lambda_i^{\leftarrow} = \log q_i \in \mathcal{RV}(1/\alpha_i),$$

due to (de Haan & Ferreira 2010, Proposition B.1.9(9)). Let $\bar{\boldsymbol{q}}(\boldsymbol{y}) := (\bar q_1(y_1),\dots q_d(y_d))$. Define the following counterparts to quantities $L_u, f_{\mathrm{LD}}, t(u)$ and $q_\infty(t)$ used in the proof of Theorem 2:

$$\bar L_u(\boldsymbol{x}) := \frac{\log L(\mathrm{e}^{\bar{\boldsymbol{q}}(t(u)\boldsymbol{x})})}{\log u} \quad \text{and} \quad \bar f_{\mathrm{LD}}(\boldsymbol{x}) := \bar L^*(\boldsymbol{q}^*\boldsymbol{x}^{\mathbf{1}/\boldsymbol{\alpha}}), \tag{EC.45a}$$

$$\text{where } t(u) := \min_i \bar\Lambda_i(\log u) = \Lambda_{\min}(u), \quad \text{and} \quad \bar q_\infty(t) := \max_{i=1,\dots,d} \bar q_i(t). \tag{EC.45b}$$

Since $\bar q_\infty^{\leftarrow} = \min_i \bar\Lambda_i$, we have $\bar q_\infty(t(u)) = \log u$ (see Lemma EC.1 and (EC.5)). Letting $\boldsymbol{Y}_u := t(u)^{-1}\boldsymbol{Y} = t(u)^{-1}\boldsymbol{\Lambda}(\boldsymbol{X})$, we have the following equivalence of events,

$$\{L(\boldsymbol{X}) > u\} = \left\{\boldsymbol{Y}_u \in \mathrm{lev}_1^+(\bar L_u)\right\},$$

from the definition of $\bar L_u$ and injectivity of $\bar{\boldsymbol{q}} = \bar{\boldsymbol{\Lambda}}^{\leftarrow}$.

As before, we proceed by showing continuous convergence of L_u to $\bar f_{\mathrm{LD}}$, as $u\to\infty$. For this purpose, consider sequences $\{u_n\}_{n\ge1}\subset\mathbb{R}_+$, $\{\boldsymbol{x}_n\}\subset\mathbb{R}^d_{++}$ such that $u_n\to\infty, \boldsymbol{x}_n\to\boldsymbol{x}>\mathbf{0}$. Since $\bar q_\infty(t(u)) = \log u$, $\bar q_i\in\mathcal{RV}(1/\alpha_i)$ for $i\in\mathcal{H}$, and $\bar q_i\in\mathcal{RV}(0), \hat q_i* = 0$ for $i\in\mathcal{L}$,

$$\frac{\bar{\boldsymbol{q}}(t(u_n)\boldsymbol{x}_n)}{\log u_n} = \frac{\bar{\boldsymbol{q}}(t(u_n)\boldsymbol{x}_n)}{\bar{\boldsymbol{q}}(t(u_n)\mathbf{1})}\hat{\boldsymbol{q}}(t(u_n)) \to \boldsymbol{x}^{\mathbf{1}/\boldsymbol{\alpha}}\boldsymbol{q}^* \tag{EC.46}$$

from (24) and (de Haan & Ferreira 2010, Proposition B.1.9(4)), as $n\to\infty$. Consequently,

$$\bar L_{u_n}(\boldsymbol{x}_n) = \frac{\log L\left(\exp\{\bar{\boldsymbol{q}}(t(u_n)\boldsymbol{x}_n)\}\right)}{\log u_n} = \frac{\log L\left(\exp\{\boldsymbol{q}^*\boldsymbol{x}^{\mathbf{1}/\boldsymbol{\alpha}}\log u_n(1+o(1))\}\right)}{\log u_n},$$

uniformly over $\boldsymbol{x}$ in compact subsets. Letting $\boldsymbol{e}(n,\boldsymbol{x}) := \mathrm{e}^{n\boldsymbol{x}}/\|\mathrm{e}^{n\boldsymbol{x}}\|_\infty$ be the unit vector, we have

$$L(\mathrm{e}^{n\boldsymbol{x}_n}) = L(\|\mathrm{e}^{n\boldsymbol{x}_n}\|_\infty \boldsymbol{e}(n,\boldsymbol{x}_n)) = \|\mathrm{e}^{n\boldsymbol{x}}\|_\infty^\rho L^*(\boldsymbol{e}(n,\boldsymbol{x}))(1+o(1)) = L^*(\mathrm{e}^{n\boldsymbol{x}})(1+o(1)),$$

where the second equality follows from the compact convergence of $L(\cdot)$ in Assumption 1 and the last equality is due to the homogeneity $L^*(c\boldsymbol{x}) = c^\rho L^*(\boldsymbol{x})$. Then from Assumption 6,

$$\bar L_{u_n}(\boldsymbol{x}_n) = \frac{\log L^*\left(\exp\{\boldsymbol{q}^*\boldsymbol{x}^{\mathbf{1}/\boldsymbol{\alpha}}\log u_n\}\right)(1+o(1))}{\log u_n} \to \bar L^*\left(\boldsymbol{q}^*\boldsymbol{x}^{\mathbf{1}/\boldsymbol{\alpha}}\right) =: \bar f_{\mathrm{LD}}(\boldsymbol{x}).$$

Then as a consequence of the continuous convergence $L_u \to \bar{f}_{\mathrm{LD}}$ above, we obtain the following from exactly the same reasoning in the proofs of Lemma EC.2 and Corollary EC.1: given $\varepsilon, M > 0$, there exists u_0 large enough such that for all $u > u_0$,

$$\mathrm{lev}_1^+(\bar{L}_u) \cap B_M \subseteq \left[\Xi_{1,M}(\bar{f}_{\mathrm{LD}})\right]^{1+\varepsilon}, \text{ and } \inf_{n: u_n > u} \chi_{\mathrm{lev}_1^+(\bar{L}_{u_n})}(\boldsymbol{x}_n) \geq \chi_{\mathrm{lev}_1^+(\bar{f}_{\mathrm{LD}})}(\boldsymbol{x}), \tag{EC.47}$$

for any $\boldsymbol{x}_n \to \boldsymbol{x}$ and $u_n \to \infty$. Thanks to these set inclusions, the conclusion in Theorem 4 follows by repeating the steps in the proof of Theorem 2 with L_u, f_{LD} replaced by $\bar{L}_u, \bar{f}_{\mathrm{LD}}$. □

To analyse the variance of the IS estimator, we have the following Lemma.

LEMMA EC.14. *Suppose that Assumption 5 holds, the parameter l in (4) is taken to be slowly varying in u, and $\rho = 1$ in (4). Then uniformly over compact subsets of $\mathbb{R}_+^d \setminus \{\boldsymbol{0}\}$,*

$$\frac{\boldsymbol{\psi}_u(t(u)\boldsymbol{p})}{t(u)} = \boldsymbol{p}\boldsymbol{1}_{\mathcal{H}} \left(1 - \frac{1}{\|\boldsymbol{q}^* \boldsymbol{p}^{\boldsymbol{1}/\boldsymbol{\alpha}}\|_\infty}\right)^{\boldsymbol{\alpha}} (1 + o(1)), \text{ as } u \to \infty, \tag{EC.48}$$

where the vector $\boldsymbol{1}_{\mathcal{H}}$ is the indicator vector (for the heavy-tailed components) defined as in (EC.29).

Proof. With $\rho = 1$, we have $c_\rho(u) = (l(u)/u)$. For $\boldsymbol{x} \in \mathbb{R}_+^d$, $\boldsymbol{T}(\boldsymbol{x}) \leq \tilde{\boldsymbol{T}}(\boldsymbol{x}) := (1 + \boldsymbol{x})[c_\rho(u)]^{-\boldsymbol{\kappa}(\boldsymbol{x})}$, component-wise. Note that $\tilde{\boldsymbol{T}}^{-1} \circ \tilde{\boldsymbol{T}}(\boldsymbol{x}) = \boldsymbol{x}$ for $\boldsymbol{x} \in \mathbb{R}_+^d$ when we take $\tilde{\boldsymbol{T}}^{-1}(\boldsymbol{x}) = \boldsymbol{x}[c_\rho(u)]^{\boldsymbol{\kappa}(\boldsymbol{x}-1)} - 1$. This yields $\boldsymbol{T}^{-1}(\boldsymbol{x}) \geq (\boldsymbol{x}[c_\rho(u)]^{\boldsymbol{\kappa}(\boldsymbol{x}-1)} - \boldsymbol{1})^+$. Combining this with the upper bound in (EC.12), we arrive at the following: For any $\boldsymbol{p} \in \mathbb{R}_+^d, \delta > 0$ there exists u_0 large enough such that for all $u > u_0$,

$$[\exp\{\bar{\boldsymbol{q}}(t(u)\boldsymbol{p}) r_u(\boldsymbol{p})\} - 1]^+ \;\leq\; \boldsymbol{T}^{-1}(\boldsymbol{q}(t(u)\boldsymbol{p})) \;\leq\; \exp\{\bar{\boldsymbol{q}}(t(u)\boldsymbol{p}) r_u(\boldsymbol{p})\} \qquad \boldsymbol{p} \in B_\delta(\boldsymbol{p}),$$
$$\text{where } r_u(\boldsymbol{p}) := 1 + \frac{\log c_\rho(u)}{\|\bar{\boldsymbol{q}}(t(u)\boldsymbol{p})\|_\infty} = 1 - \frac{1 + o(1)}{\|\boldsymbol{p}^{\boldsymbol{1}/\boldsymbol{\alpha}} \boldsymbol{q}^*\|_\infty} \qquad \text{[due to (EC.46)]}.$$

Since $\boldsymbol{\psi}_u := \boldsymbol{\Lambda} \circ \boldsymbol{T}^{-1} \circ \boldsymbol{q}$ and $\boldsymbol{\Lambda}$ is increasing component wise,

$$\boldsymbol{\Lambda}\big([\exp\{\bar{\boldsymbol{q}}(t(u)\boldsymbol{p}) r_u(\boldsymbol{p})\} - \boldsymbol{1}]^+\big) \;\leq\; \boldsymbol{\psi}_u(t(u)\boldsymbol{p}) \;\leq\; \boldsymbol{\Lambda}\big(\exp\{\bar{\boldsymbol{q}}(t(u)\boldsymbol{p}) r_u(\boldsymbol{p})\}\big). \tag{EC.49}$$

As $\Lambda_i \circ \exp = \bar{\Lambda}_i \in \mathcal{RV}(\alpha_i)$ for $i \in \mathcal{H}$ and $\bar{q}_i := \bar{\Lambda}_i^{\leftarrow}$, the term

$$\Lambda_i\big(\exp\{\bar{q}_i(t(u)p_i) r_u(p_i)\}\big) = r_u(p_i)^{\alpha_i} \bar{\Lambda}_i \circ \bar{q}_i(t(u)p_i)(1 + o(1)) = r_u^{\alpha_i}(p_i) t(u) p_i (1 + o(1)), \quad \text{for } i \in \mathcal{H},$$

On the other hand when $i \in \mathcal{L}$, we have $\Lambda_i \in \mathcal{RV}(\alpha_i)$ and $q_i \in \mathcal{RV}(1/\alpha_i)$. In this case,

$$\Lambda_i\big(\exp\{\bar{q}_i(t(u)p_i) r_u(p_i)\}\big) = \Lambda_i\left(q_i(t(u)\boldsymbol{p})^{r_u(p_i)}\right) = O(t(u)^{1 - \|\boldsymbol{p}^{\boldsymbol{1}/\boldsymbol{\alpha}} \boldsymbol{q}^*\|_\infty^{-1} + o(1)}). \quad \text{for } i \in \mathcal{L}$$

Due to the above deduction that $r_u(\boldsymbol{p}) = 1 - \|\boldsymbol{p}^{\boldsymbol{1}/\boldsymbol{\alpha}} \boldsymbol{q}^*\|_\infty^{-1}(1 + o(1))$. Since the above convergences uniformly over compact subsets, (EC.49) results in,

$$\lim_{u \to \infty} t(u)^{-1} \boldsymbol{\psi}_{u,i}(t(u)\boldsymbol{p}) = \begin{cases} p_i \left(1 - \|\boldsymbol{p}^{\boldsymbol{1}/\boldsymbol{\alpha}} \boldsymbol{q}^*\|_\infty^{-1}\right)^{\alpha_i} & \text{for } i \in \mathcal{H}, \\ 0 & \text{for } i \in \mathcal{L}. \end{cases} \qquad \square$$

Proof of Theorem 5: Following the reasoning in the proof of Theorem 5, notice that the second moment of the IS estimator may be written as $M_{2,u} = t^{2d}(u)\mathbb{E}\left[\exp\left\{-t(u)\bar{F}_u(\bar{\boldsymbol{Y}}_u)\right\}\right]$, where

$$\bar{F}_u(\boldsymbol{p}) = a_u(\boldsymbol{p}) + b_u(\boldsymbol{p}) + \chi_{\mathrm{lev}_1^+(\bar{L}_u)}(\boldsymbol{p}). \tag{EC.50}$$

Here, $a_u(\boldsymbol{p})$ and $b_u(\boldsymbol{p})$ are as defined in Lemma EC.6. Following Lemma EC.14 and the proof of Lemma EC.7, we obtain $a_u(\boldsymbol{p}) \geq I(\boldsymbol{p}) - I\left(\boldsymbol{p}\mathbf{1}_{\mathcal{H}}\left(1 - 1/\|\boldsymbol{q}^*\boldsymbol{p}^{\mathbf{1}/\boldsymbol{\alpha}}\|\right)^{\boldsymbol{\alpha}}\right) + o(1)$. uniformly over compact subsets of $\mathbb{R}^d_{++}$. We have from Assumption 5 that for $i \in \mathcal{H}$, $\Lambda \circ \exp \in \mathcal{RV}(\alpha_i)$, for some $\alpha_i \geq 1$. Therefore, $\Lambda_i \in \mathcal{RV}(0)$ whenever $i \notin \mathcal{L}$. Now, since $\lambda_i(\cdot)$ are monotone, (de Haan & Ferreira 2010, Proposition B.1.9 (7)) implies that $\lambda_i \in \mathcal{RV}(\gamma_i - 1)$. Here, $\gamma_i = 0$ if $i \in \mathcal{H}$. Following the steps in the proof of Lemma EC.7, bound the product (for large enough u) in (b) as

$$\prod_{i=1}^{d} \frac{\lambda_i(q_i(t(u)p_i))}{\lambda_i(\boldsymbol{T}_i^{-1}(\boldsymbol{q}(t(u)\boldsymbol{p})))} J(\boldsymbol{T}_i^{-1}(\boldsymbol{q}(t(u)\boldsymbol{p}))) \leq \exp\left(\log(u/l)\left[\sum_{i=1}^{d} \gamma_i \boldsymbol{\kappa}_i(\boldsymbol{q}(t(u)\boldsymbol{p})) + o(1)\right]\right) \tag{EC.51}$$

Whenever $\mathcal{L} \neq [d]$ and $i \in \mathcal{L}$, it is easy to see that $\bar{q}_i^* = 0$. For all such i, $\boldsymbol{\kappa}_i(\boldsymbol{q}(t(u)\boldsymbol{p})) = o(1)$. Further, for all $i \in \mathcal{H}$, $\gamma_i = 0$. Finally, observe that with $\bar{\Lambda}_i \in \mathcal{RV}(\alpha_i)$ for $i \in \mathcal{H}$, we have $\log u = O(t(u))$. Now, (EC.51) suggests $b_u(\boldsymbol{p}) \geq -t(u)\varepsilon$ for all large enough u. Noting the convergences in (EC.47), and repeating the arguments from the proof of Theorem 3, replacing $F_u(\boldsymbol{p})$ there by $\bar{F}_u(\boldsymbol{p})$,

$$\limsup_{u\to\infty} [\Lambda_{\min}(u)]^{-1} \log M_{2,u} \leq -\inf_{\boldsymbol{p}\in \mathrm{lev}_1^+(\bar{f}_{\mathrm{LD}})} 2I(\boldsymbol{p}) + I\left(\boldsymbol{p}\mathbf{1}_{\mathcal{H}}\left(1 - 1/\|\boldsymbol{q}^* \cdot \boldsymbol{p}^{\mathbf{1}/\boldsymbol{\alpha}}\|_\infty\right)^{\boldsymbol{\alpha}}\right) + 2\varepsilon.$$

Due to the homogeneity of $I(\cdot)$ (see Lemma 2(b)), it can be seen that the above infimum occurs at the boundary, $\|\boldsymbol{q}^* \cdot \boldsymbol{p}^{\mathbf{1}/\boldsymbol{\alpha}}\|_\infty = 1$, and therefore, $\limsup_{u\to\infty} \frac{1}{\Lambda_{\min}(u)} \log M_{2,u} \leq -2I^* + 2\varepsilon$. □

Verification of Remark 1: For any $f, g \in \mathcal{RV}(p)$ that are eventually strictly increasing and satisfying $\lim_{x\to\infty} f(x)/g(x) = c \in (0,\infty)$, we first show that $\lim_{x\to\infty} g^{\leftarrow}(x)/f^{\leftarrow}(x) = c^{1/p}$. For this purpose, observe $\lim_{t\to\infty} g^{\leftarrow}(tx)/g^{\leftarrow}(t) = x^{1/p}$ uniformly over $x \in (c/2, 2c)$ as $t \to \infty$. Setting $t = g(f^{\leftarrow}(x))$, we have $t \to \infty$ and $f(f^{\leftarrow}(x))/g(f^{\leftarrow}(x)) \to c$ as $x \to \infty$. Therefore,

$$g^{\leftarrow}(x) = g^{\leftarrow}\left(g(f^{\leftarrow}(x)) \cdot \frac{f(f^{\leftarrow}(x))}{g(f^{\leftarrow}(x))}\right) \sim c^{1/p} f^{\leftarrow}(x).$$

This verifies the claim $g^{\leftarrow}(x)/f^{\leftarrow}(x) \to c^{1/p}$. To see (17) as a consequence, fix any $i \in \{1,\ldots,d\}$ such that q_i^* exists and $q_i^*(x) > 0$. Setting $f = q_i$ and $g(\cdot) = \|\boldsymbol{q}(\cdot)\|_\infty$, we have $f^{\leftarrow} = \Lambda_i, g^{\leftarrow} = \Lambda_{\min}$ (see (EC.5)). Since $\Lambda_{\min} \in \mathcal{RV}(\alpha_*)$, $q_i^* = (\lim_{x\to\infty} \Lambda_{\min}(x)/\Lambda_i(x))^{1/\alpha_*}$. If $q_i^* = 0$, the conclusion is immediate from the differing rates of growths of the numerator and the denominator. Finally to verify the sufficient condition on the derivative, consider any sequence $\{x_n\} \subset \mathbb{R}$ increasing to infinity. Since $|r_i'(x)| \leq Mx^{-(1+\varepsilon)}$ for suitable constants $M, \varepsilon > 0$,

$$|r_i(x_{m+n}) - r_i(x_m)| \leq \int_{x_m}^{x_{m+n}} |r_i'(x)| dx \leq \varepsilon^{-1} M x_m^{-\varepsilon},$$

for all sufficiently large m. Therefore the sequence $\{r_i(x_n) : n \geq 1\}$ is Cauchy and is convergent.

EC.5. Proof Theorem 3 with $\boldsymbol{\kappa} = \boldsymbol{\kappa}_2$

To avoid complicating notation, we omit the dependence of $\boldsymbol{T}_u^{(2)}$ on u and in the subsequent proof, simply use $\boldsymbol{T}$ instead. Observe that $\boldsymbol{T}(\boldsymbol{x}) = (T_{1,c}(x_1), \ldots, T_{1,c}(x_d))$ for a 1-1 onto function $T_{1,c}$ (defined imminently) and therefore itself 1-1 and onto. Define the function $\boldsymbol{\psi}_{u,1} = \boldsymbol{\Lambda} \circ \boldsymbol{T}^{-1} \circ \boldsymbol{q}$. To proceed, we check that the conditions required in the proof of Theorem 3 hold. As in that case, first, we bound $\boldsymbol{T}^{-1}$ from above. Observe that $\boldsymbol{T}(\boldsymbol{x}) = (T_{1,c}(x_1), \ldots, T_{1,c}(x_d))$, where

$$T_{1,c}(y) = y\,[u/l]^{\frac{\log(1+y)}{\log l}} \geq \begin{cases} y^{\log u/\log l} & \text{whenever } y \geq 1 \\ y & \text{otherwise} \end{cases}$$

Therefore

$$T_{1,c}^{-1}(y) \leq \begin{cases} y^{\log l/\log u} & \text{whenever } y \geq 1 \\ y & \text{otherwise.} \end{cases} \tag{EC.52}$$

Denote $\boldsymbol{\psi}_{u,1} = \boldsymbol{\Lambda} \circ \boldsymbol{T}^{-1} \circ \boldsymbol{q}$. The bound on $\boldsymbol{T}^{-1}$ established, we now proceed to check the technical conditions required for log-efficiency. This amounts to verifying (EC.44) with $\boldsymbol{\psi}_u$ replaced by $\boldsymbol{\psi}_{u,1}$, and bounding the Jacobian determinant of $\boldsymbol{T}$. With these bounds established, the rest of the proof is similar to case when $\boldsymbol{\kappa} = \boldsymbol{\kappa}_1$.

Step 1: Verifying condition (EC.44): Observe that given (EC.52),

$$\Lambda_i(\boldsymbol{T}_i^{-1}(\boldsymbol{q}(t(u)\boldsymbol{p}))) \leq \Lambda_i\left(q_i(t(u)p_i)^{\log l/\log u}\right) \vee \Lambda_i(1)$$

With l being slowly varying in u, $\|\boldsymbol{\psi}_{u,1}(t(u)\boldsymbol{p})\|_\infty \leq \|\boldsymbol{\Lambda}[(\boldsymbol{q}(t(u)\boldsymbol{p}))]^{o(1)}\|_\infty$ and therefore, one has the bound $\|\boldsymbol{\psi}_{u,1}(t(u)\boldsymbol{p})\|_\infty \leq \varepsilon t(u)\|\boldsymbol{p}\|_\infty$ for all u large enough. This further implies that uniformly over compact subsets of $\mathbb{R}_+^d$, $\|\boldsymbol{\psi}_{u,1}(t(u)\boldsymbol{p})\|_\infty = o(t(u))$, and establishes an equivalent of Lemma EC.5, but with $\boldsymbol{\kappa}^{(1)}$ replaced by $\boldsymbol{\kappa}^{(2)}$ in the definition of $\boldsymbol{T}$.

Step 2: Bounding the Jacobian determinant: We establish that $\log J(\boldsymbol{q}(t(u)\boldsymbol{p})) = o(t(u))$. To this end, recall that for $\boldsymbol{x} \in \mathbb{R}_+^d$ (the case where $\boldsymbol{x}$ is allowed to be in $\mathbb{R}^d$ can be handled similarly),

$$J(\boldsymbol{x}) = \left(\frac{u}{l}\right)^{\mathbf{1}^\intercal \boldsymbol{\kappa}^{(2)}(\boldsymbol{x})} \prod_{i=1}^d \left(1 + \frac{\log(u/l)}{\log l}\frac{x_i}{1+x_i}\right) = \prod_{i=1}^d \left[(u/l)^{\frac{\log(1+x_i)}{\log l}} + \frac{T_{1,c}(x_i)}{1+x_i}\frac{\log(u/l)}{\log l}\right].$$

Therefore,

$$\log J(\boldsymbol{T}^{-1}(\boldsymbol{q}(t(u)\boldsymbol{p}))) = \sum_{i=1}^d \log\left[(u/l)^{\frac{\log(1+T_{1,c}^{-1}(q_i(t(u)p_i)))}{\log l}} + \frac{q_i(t(u)p_i)}{(1+T_{1,c}^{-1}(q_i(t(u)p_i))))}\frac{\log(u/l)}{\log l}\right]$$

The rate of increase of the right hand side above is determined by the larger of

$$\sum_{i=1}^d \log\left[(u/l)^{\frac{\log(1+T_{1,c}^{-1}(q_i(t(u)p_i)))}{\log l}}\right] \text{ and } \log\left[\frac{q_i(t(u)p_i)}{(1+T_{1,c}^{-1}(q_i(t(u)p_i))))}\frac{\log(u/l)}{\log l}\right]. \tag{EC.53}$$

Since l is slowly varying in u, the bound in (EC.52) yields

$$\limsup_{u\to\infty} \frac{1}{t(u)} \sum_{i=1}^{d} \log\left[(u/l)^{\frac{\log(1+T_{1,c}^{-1}(q_i(t(u)p_i)))}{\log l}} \right] \leq \limsup_{u\to\infty} \frac{\log(u/l)}{t(u)\log l} \|[\boldsymbol{q}(t(u)\boldsymbol{p})]^{o(1)}\|_1 = 0$$

Similarly, the second term of (EC.53) may be bounded as

$$\limsup_{u\to\infty} \frac{1}{t(u)} \log\left[\frac{q_i(t(u)p_i)}{(1+T_{1,c}^{-1}(q_i(t(u)p_i)))} \frac{\log(u/l)}{\log l} \right] = 0, \text{ and consequently,}$$

$$\log J_1(\boldsymbol{T}^{-1}(\boldsymbol{q}(t(u)\boldsymbol{p}))) = o(t(u)) \qquad \text{(see for e.g. (Dembo \& Zeitouni 1998, Lemma 1.2.14)).}$$

Step 3: Combine the bounds: Note the expression for the second moment of the IS estimator with $\boldsymbol{T}^{(2)}$ instead of $\boldsymbol{T}^{(1)}$ is given by replacing $\boldsymbol{\psi}_u$ by $\boldsymbol{\psi}_{u,1}$ and substituting the appropriate Jacobian as given in Table 1. From Steps 1 and 2, the consequences of Lemmas EC.7- EC.10 continue to hold, and the rest of the proof follows from the proof of Theorem 3. □

EC.6. Verifying Assumption 1(b)

Recall that a random vector $\boldsymbol{Y}$ is said to be multivariate regularly varying with index ρ if for any set A not containing the origin,

$$tP\left[\frac{\boldsymbol{Y}}{t^{\rho}} \in A\right] \to \mu(A), \tag{EC.54}$$

for some non-zero radon measure μ. An equivalent formulation (Resnick 2007, Theorem 6.1) is that for some probability measures $\mu_{1/\rho}$ on the line and M on the sphere,

$$tP\left[\left(\|\boldsymbol{Y}\|, \frac{\boldsymbol{Y}}{\|\boldsymbol{Y}\|}\right) \in A\right] \to (\mu_{1/\rho} \times M)(A), \tag{EC.55}$$

for any set A not containing the origin. Here $\mu_{1/\rho}$ is taken to be $\mu_\rho(c,\infty] = c^{-1/\rho}$, without loss of generality, for any $c > 0$. To verify Assumption 1(b), we develop the following characterisation of Assumption 1(b) based on multivariate regular variation: Let $\mathcal{S}_+^{d-1} := \{\boldsymbol{x} \in \mathbb{R}_+^d : \|\boldsymbol{x}\| = 1\}$ denote the intersection of unit sphere and the positive orthant.

Proposition EC.3. *Let R be a random variable satisfying $P(R \leq r) = 1 - 1/r,\ r \geq 1$ and $\boldsymbol{\Theta}$ be uniformly distributed on $\mathcal{S}_+^{d-1}$, independently of R. Then $L(\cdot)$ satisfies Assumption 1(b) with $\rho \in (0,\infty)$ if and only if the random vector $L(R\boldsymbol{\Theta}) \cdot \boldsymbol{\Theta}$ is a multivariate regularly varying random vector with index ρ.*

Verifying regular variation of a random vector is well-studied and continues to be a topic of active research (see Einmahl *et al.* 2021, and references therein). As a consequence of Proposition EC.3, one can obtain independent samples of $L(R\boldsymbol{\Theta})\boldsymbol{\Theta}$ and use the statistical test developed in (Einmahl *et al.* 2021) to verify if $L(\cdot)$ satisfies Assumption 1b merely from oracle queries to the evaluations of $L(\cdot)$. The rest of this section sketches the proof of Proposition EC.3.

Remark EC.3. Suppose that $L(\cdot)$ has an approximation $\tilde{L}(\cdot)$ which satisfies either (i) $|L(n\boldsymbol{\theta}) - \tilde{L}(n\boldsymbol{\theta})| = o(L(n\boldsymbol{\theta}))$, or more generally, (ii) $\tilde{L}(n\boldsymbol{\theta}) = c(\boldsymbol{\theta})L(n\boldsymbol{\theta})(1+o(1))$, uniformly over $\boldsymbol{\theta}$ on $\mathcal{S}_+^{d-1}$ and for some continuous $c(\cdot)$. Then it is sufficient to verify Assumption 1(b) for the approximate functional $\tilde{L}$. Such a verification is useful if, for example, $\tilde{L}(\cdot)$ is either available in explicit form (or) if its evaluations are computationally less expensive than those of $L(\cdot)$.

Proof for Proposition EC.3: a) First suppose that $L(\cdot)$ satisfies Assumption 1(b). Recall that as a consequence of (Resnick 2007, Theorem 6.1, Lemma 6.2), to verify the multivariate regularly varying property, it is sufficient to verify that for all $\boldsymbol{x} > 0$,

$$nP\left(\frac{L(R\boldsymbol{\Theta})\cdot\boldsymbol{\Theta}}{n^\rho} \in [\mathbf{0}, \boldsymbol{x}]^c\right) \to \mu[\mathbf{0}, \boldsymbol{x}]^c,$$

for some measure μ. Let $R_n = R/n$. Using the independence of $(R, \boldsymbol{\Theta})$, the probability above equals

$$k_d \int_{\boldsymbol{\theta}\in S_{d-1}^{\infty}} P\left(\frac{L(nR_n\boldsymbol{\theta})}{n^\rho}\boldsymbol{\theta} \in [\mathbf{0}, \boldsymbol{x}]^c\right) d\boldsymbol{\theta} = k_d \int_{\boldsymbol{\theta}\in S_{d-1}^{\infty}} P(R_n \in S_{n,\boldsymbol{\theta},\boldsymbol{x}}) d\boldsymbol{\theta}, \tag{EC.56}$$

where k_d equals the 1/(volume of $\mathcal{S}_+^{d-1}$) and the set $S_{n,\boldsymbol{\theta},\boldsymbol{x}} = \{r : \boldsymbol{\theta}L(nr\boldsymbol{\theta})/n^\rho \in [0, \boldsymbol{x}]^c\}$. Notice that owing to the convergence of $L(nr\boldsymbol{\theta})/n^\rho$ to $r^\rho L^*(\boldsymbol{\theta})$, the set $S_{n,\boldsymbol{\theta},\boldsymbol{x}}$ converges to $S^*_{\boldsymbol{\theta},\boldsymbol{x}} = \{r : r^\rho L^*(\boldsymbol{\theta})\cdot\boldsymbol{\theta} \in [\mathbf{0}, \boldsymbol{x}]^c\}$ in the Painlevé-Kuratowski sense (Rockafellar & Wets 1998, Section 4.B). Consequently, $nP(R_n \in S_{n,\boldsymbol{\theta},\boldsymbol{x}}) \to \mu_0(S^*_{\boldsymbol{\theta},\boldsymbol{x}})$ where the density of $\mu_0(dr) = r^{-2}dr$ on the line (note that the measure μ_0 is a Radon measure, and not a probability measure). Plugging this back into the integral in (EC.56), it can be seen that whenever $L(\cdot)$ satisfies Assumption 1(b), $L(R\boldsymbol{\Theta})\cdot\boldsymbol{\Theta}$ is multivariate regularly varying (with limiting measure $\mu[\mathbf{0}, \boldsymbol{x}]^c = k_d \int_{\boldsymbol{\theta}\in S_{d-1}^{\infty}} \mu_0(S_{\boldsymbol{\theta},\boldsymbol{x}})d\boldsymbol{\theta}$).

b) Now, suppose that $L(R\boldsymbol{\Theta})\cdot\boldsymbol{\Theta}$ is multivariate regularly varying with index ρ. Let $L_{\boldsymbol{\theta},n}(r) = L(nr\boldsymbol{\theta})/n^\rho$. Then, as a consequence of representation (EC.55), with $\boldsymbol{\Theta}$ and R_n being independent, for any $c > 0$ (the limits below being taken in an appropriate sense),

$$nP(R_n \in \text{lev}_c^+(L_{\boldsymbol{\theta},n})) = nP\left(\frac{L(nR_n\boldsymbol{\Theta})}{n^\rho} \in [c, \infty),\ \boldsymbol{\Theta} \in d\boldsymbol{\theta}\right) \to c^{-1/\rho}m(\boldsymbol{\theta})d\boldsymbol{\theta}, \tag{EC.57}$$

for some $m : \mathcal{S}_+^{d-1} \to \mathbb{R}_+$. Define the function $L^*(\boldsymbol{x}) = (\|\boldsymbol{x}\|/m(\boldsymbol{x}/\|\boldsymbol{x}\|))^\rho$ and let $L^*_{\boldsymbol{\theta}}(r) = L^*(r\boldsymbol{\theta})$. Observe then that $\{r : L^*_{\boldsymbol{\theta}}(r) \geq c\} = [c^{-1/\rho}m(\boldsymbol{\theta}), \infty)$. Now, (EC.57) implies that for every $\boldsymbol{\theta}$ and $c > 0$

$$nP(R_n \in \text{lev}_c^+(L_{n,\boldsymbol{\theta}})) \to c^{1/\rho}m(\boldsymbol{\theta})d\boldsymbol{\theta} \implies \text{ess-inf lev}_c^+(L_{n,\boldsymbol{\theta}}) \to \text{ess-inf lev}_c^+(L^*_{\boldsymbol{\theta}}).$$

Since the above holds for every $c > 0$, the level sets themselves must converge, that is for every $c > 0$, $\text{lev}_c^+(L_{n,\boldsymbol{\theta}}) \to \text{lev}_c^+(L^*_{\boldsymbol{\theta}})$. Further, observe that

$$\text{lev}_c^+(L_n) = \bigcup_{\boldsymbol{\theta}\in S_\infty^{d-1}} \{(r,\boldsymbol{\theta}) : r \in \text{lev}_c^+(L_{n,\boldsymbol{\theta}})\} \quad \text{and} \quad \text{lev}_c^+(L^*_{\boldsymbol{\theta}}) = \bigcup_{\boldsymbol{\theta}\in S_\infty^{d-1}} \{(r,\boldsymbol{\theta}) : r \in \text{lev}_c(L^*)\}.$$

Now, use the uniformity of convergence in (EC.57) over $\boldsymbol{\theta}$ (refer to (Resnick 2007, Theorem 6.4)) to observe that for all $c > 0$, $\text{lev}_c^+(L_n) \to \text{lev}_c^+(L^*)$. An application of (Rockafellar & Wets 1998, Proposition 7.7) implies that since L_n converges epigraphically to L^*, it converges uniformly on compact subsets. Finally, from (Rockafellar & Wets 1998, Theorem 7.14) uniform convergence implies the continuous convergence in Assumption 1(b). □

EC.7. Application to CVaR Minimisation

Suppose that $\ell(\boldsymbol{X}, \boldsymbol{\theta})$ denotes the loss associated with a decision choice $\boldsymbol{\theta}$ under a random vector $\boldsymbol{X}$. Let $v_\beta(\boldsymbol{\theta})$ denote the $(1-\beta)$-th quantile of $\ell(\boldsymbol{X}, \boldsymbol{\theta})$. Then its CVaR at the tail-level $\beta \in (0,1)$ is

$$C_\beta(\boldsymbol{\theta}) := E\left[\ell(\boldsymbol{X}, \boldsymbol{\theta}) \mid \ell(\boldsymbol{X}, \boldsymbol{\theta}) \geq v_\beta(\boldsymbol{\theta})\right],$$

Minimizing CVaR $C_\beta(\boldsymbol{\theta})$ over a compact set Θ enjoys the following variational representation (see Rockafellar & Uryasev 2000),

$$\inf_{u \in \mathbb{R}, \boldsymbol{\theta} \in \Theta} \left[u + \beta^{-1} E\left(\ell(\boldsymbol{X}, \boldsymbol{\theta}) - u\right)^+\right] = \inf_{u \in \mathbb{R}, \boldsymbol{\theta} \in \Theta} f(u, \boldsymbol{\theta}), \tag{EC.58}$$

If $\ell(\boldsymbol{X}, \cdot)$ is convex, then $f(\cdot)$ is convex. As a result, CVaR minimization has become the most prominent vehicle in a number of applications for arriving at decisions with low tail risks. To perform the minimization without expending exorbitant computational effort in problems with small β, one can consider minimising the following IS weighted Sample Average Approximation:

$$\hat{f}_{is,n}(u, \boldsymbol{\theta}) = \left[u + \frac{1}{n\beta} \sum_{i=1}^{n} (\ell(\boldsymbol{Z}_{u,i}, \boldsymbol{\theta}) - u)^+ \mathcal{L}_i\right], \tag{EC.59}$$

where, as before, $\mathcal{L}$ is defined to be the likelihood between $\boldsymbol{X}$ and $\boldsymbol{Z}_u = \boldsymbol{T}_u(\boldsymbol{X}) := \boldsymbol{X}[u/l]^{\boldsymbol{\kappa}(\boldsymbol{X})}$:

$$\mathcal{L} = \frac{f_{\boldsymbol{X}}\left(\boldsymbol{X}[u/l]^{\boldsymbol{\kappa}(\boldsymbol{X})}\right)}{f_{\boldsymbol{X}}(\boldsymbol{X})} J(\boldsymbol{X}) \text{ where } J(\boldsymbol{x}) = \text{Det}[\partial \boldsymbol{T}_u(\boldsymbol{x})/\partial \boldsymbol{x}] \text{ is as defined in (9).} \tag{EC.60}$$

Notice that since $\mathcal{L}$ depends on u through the factor $[u/l]^{\boldsymbol{\kappa}(\boldsymbol{X})}$. Therefore even if $f(\cdot)$ as defined in (EC.58) were convex in $(u, \boldsymbol{\theta})$, the IS weighted objective (EC.59) need not be convex. In turn, such lack of convexity due to the introduction of likelihood ratio and absence of efficient change of measure prescriptions which hold uniformly well simultaneosuly over feasible $(u, \boldsymbol{\theta})$ have been the primary bottlenecks in using IS, in general, for optimization.

Since the loss structure is explicitly known in optimization settings, the growth rate ρ in Assumption 1b is typically readily known and the IS transformation $\boldsymbol{T}$ in (4) employed with $\boldsymbol{\kappa} = \boldsymbol{\kappa}_1$ is particularly well-suited to overcome the above difficulties. In particular, by (i) changing variable as in $[u/l] = s$ where $s \in [1, \infty)$, and (ii) using $\boldsymbol{\kappa} = \boldsymbol{\kappa}^{(1)}$ in (7), the resulting IS weighted objective (EC.61) remains convex in $(u, \boldsymbol{\theta})$ when $\ell(\boldsymbol{X}, \boldsymbol{\theta})$ is convex in $\boldsymbol{\theta}$. Except for the knowledge of ρ, the model agnostic nature of the change makes the change of measure induced by $\boldsymbol{T}$ to possess low variance at every feasible $(u,\boldsymbol{\theta})$. Parameterizing the stretch factor s as $s = h\log\log(1/\beta)$, the selection of hyperparameter h at any given feasible $(u,\boldsymbol{\theta})$ can be accomplished by minimizing the second moment as in Step 2 of Algorithm 3. Algorithm 4 below incorporates this selection in every stage of Retrospective Approximation of the CVaR objective. We refer the readers to a follow-up work Deo *et al.* (2022) for further implementation details.

Algorithm 4: IS based CVaR Optimisation

Input: Tail probability level β, samples $\boldsymbol{X}_1,\ldots,$ from $f_{\boldsymbol{X}}(\cdot)$, initializations $u_0,\boldsymbol{\theta}_0,h_0$.

For $k \geq 1$**, do** the following steps until stopping criterion is met

Step 1(IS-Weighted CVaR optimisation): With a sample size of m_k and error tolerance ε_k:

a) Transform the samples: For each sample $i = 1,\ldots,m_k$, compute the transformation,

$$\boldsymbol{Z}_i = \boldsymbol{T}(\boldsymbol{X}_i) := \boldsymbol{X}_i[s]^{\boldsymbol{\kappa}^{(1)}(\boldsymbol{X}_i)},$$

with $s = h_{k-1}\log\log(1/\beta)$.

b) Minimize the IS weighted CVaR objective

$$\inf_{u,\boldsymbol{\theta}}\left\{u + \frac{1}{m_k\beta}\sum_{i=1}^{m_k}\left[\ell(\boldsymbol{Z}_i,\boldsymbol{\theta}) - u\right]^+\mathcal{L}_{h,i}\right\}, \tag{EC.61}$$

with the initial solution iterate set to $(u_{k-1},\boldsymbol{\theta}_{k-1})$ and $h = h_{k-1}$. Let $(u_k,\boldsymbol{\theta}_k)$ denote the optimiser returned after reaching an error tolerance $\varepsilon_k = m_k^{-1/2}$.

2 Update the cross validation parameter: With the initial solution iterate set to h_{k-1}, minimize the sample second moment estimate as in

$$\inf_{h>0}\frac{1}{m_k}\sum_{i=1}^{m_k}\left\{\mathbf{I}\left(\ell(\boldsymbol{T}_h(\boldsymbol{X}_i),\boldsymbol{\theta}_k)\geq u_k\right)\mathcal{L}_{h,i}\right\}^2,$$

where $\boldsymbol{T}_h(\boldsymbol{X}_i) = \boldsymbol{X}_i[h\log\log(1/\beta)]^{\boldsymbol{\kappa}^{(1)}(\boldsymbol{X}_i)}$. Let h_k denote the solution obtained by solving until reaching error-tolerance $\varepsilon_k = m_k^{-1/2}$.

Numerical results: Consider the constrained minimum CVaR portfolio optimisation problem: Here $\ell(\boldsymbol{x},\boldsymbol{\theta}) = \boldsymbol{\theta}^\intercal\boldsymbol{x}$ and the set $\Theta = \{\boldsymbol{\theta} : \boldsymbol{\theta}^\intercal\mathbf{1} = 1, \boldsymbol{\theta}^\intercal\boldsymbol{\mu}\geq r\}$, where $\boldsymbol{\mu}$ denotes the expected returns. The marginals of the loss realizations $\boldsymbol{X}$ are taken to have the c.d.f.s $F_i(x) = P(X_i\leq x) = 1 - \mathrm{e}^{-x^{\alpha_i}}$ where $\alpha_i = 0.5\ \forall i$. Dependence is modelled through a Gaussian copula whose covariance matrix $\boldsymbol{R}$ is designed to capture a realistic degree of correlation among various asset returns. In order to compare the effort required to obtain a desired out of sample accuracy, we give (i) the number of samples required by each method to obtain 1% relative regret (relative error between the optimal CVaR and CVaR computed at the solution proposed by the respective algorithm) and (ii) the out-of-sample regret when the number of loss evaluations used by each algorithm is restricted to 2500. For the former case, with $\beta = 0.037$, for IS, this is ≈ 600, while SAA requires ≈ 5500 samples. This difference is even more pronounced when $\beta = 0.003$, where SAA requires roughly 28000 samples, while IS only requires 1175. For the latter, at $\beta = 0.037$, IS gives a regret of 2% while SAA gives a regret of 5%. We refer the reader to Deo *et al.* (2022) for more details on the numerical experiments and the explicit specifications for $\{m_k, \varepsilon_k : k \geq 1\}$.

Code availability: Python implementations for importance sampling using self-structuring transformations are available at https://github.com/ananddeo161093/BBIS_Source_Codes.