Supplementary Material to "Sample path large deviations for multivariate heavy-tailed Hawkes processes and related Lévy processes"

JOSE BLANCHET[1,a], ROGER J. A. LAEVEN[2,b], XINGYU WANG[2,c] and BERT ZWART[3,d]

[1] *Department of Management Science and Engineering, Stanford University,* [a] *Jose.Blanchet@stanford.edu*
[2] *Department of Quantitative Economics, University of Amsterdam,* [b] *R.J.A.Laeven@uva.nl,* [c] *X.Wang4@uva.nl*
[3] *Centrum Wiskunde & Informatica (CWI),* [d] *Bert.Zwart@cwi.nl*

Contents

A Technical Details for Section 2 . . . 1
A.1 Proof of Lemma 2.3 . . . 1
A.2 Details for Tail Asymptotics of Hawkes Process Clusters . . . 3
B Counterexamples of Topology and Tail Behavior . . . 4
C Proof for Large Deviations of Lévy Processes with $\mathcal{MHRV}$ Increments . . . 11
C.1 Proof of Theorem 3.2 . . . 11
C.2 Proofs of Propositions C.2 and C.3 . . . 15
D Proof for Large Deviations of Multivariate Heavy-Tailed Hawkes Processes . . . 28
D.1 Adapting Theorem 3.2 to $\mathbb{D}[0,\infty)$ . . . 28
D.2 Proof of Proposition 3.4 . . . 34
D.3 Proof of Proposition 3.5 . . . 37
D.4 Univariate Results for M_1 Topology of $\mathbb{D}[0,T]$ . . . 47
E Sample Path Large Deviations for Lévy Processes with $\mathcal{MHRV}$ Increments (General Case) . . . 52
F Theorem Tree . . . 56
References . . . 57

Appendix A: Technical Details for Section 2

Section A.1 provides the proof of Lemma 2.3. Section A.2 collects additional details for tail asymptotics of Hawkes process clusters in Theorem 2.6, which also address Monte Carlo simulation for the limiting measures in Theorem 3.3.

A.1. Proof of Lemma 2.3

Proof of Lemma 2.3. Take any Borel measurable $B \subseteq \mathbb{S}$ that is bounded away from $\mathbb{C}$. W.l.o.g., in this proof we only consider $\Delta > 0$ small enough that $\boldsymbol{d}(B,\mathbb{C}) > 2\Delta$, and hence $B^{2\Delta}$ is still bounded away from $\mathbb{C}$.

First, for any $n \geq 1$ and $\delta > 0$,

$$\begin{aligned}&\mathbf{P}(X_n \in B)\\ &= \mathbf{P}\big(X_n \in B;\ \boldsymbol{d}(X_n, Y_n^\delta) \leq \Delta\big) + \mathbf{P}\big(X_n \in B;\ \boldsymbol{d}(X_n, Y_n^\delta) > \Delta\big)\end{aligned}$$

$$\le \mathbf{P}\big(Y_n^\delta \in B^\Delta\big) + \mathbf{P}\big(X_n \in B \text{ or } Y_n^\delta \in B;\ \boldsymbol{d}(X_n, Y_n^\delta) > \Delta\big)$$
$$= \sum_{v\in\mathcal{V}} \mathbf{P}\big(Y_n^\delta \in B^\Delta,\ V_n^\delta = v\big) + \mathbf{P}\big(Y_n^\delta \in B^\Delta,\ V_n^\delta \notin \mathcal{V}\big) + \mathbf{P}\big(X_n \in B \text{ or } Y_n^\delta \in B;\ \boldsymbol{d}(X_n, Y_n^\delta) > \Delta\big).$$

Then, by Condition (ii), there exists some $\delta_0 > 0$ such that for all but countably many $\delta \in (0, \delta_0)$,

$$\limsup_{n\to\infty} \epsilon_n^{-1}\mathbf{P}(X_n \in B) \le \sum_{v\in\mathcal{V}} \mu_v(B^{2\Delta}) + \limsup_{n\to\infty} \epsilon_n^{-1}\mathbf{P}\Big(\boldsymbol{d}\big(X_n, Y_n^\delta\big)\mathbb{I}\big(X_n \in B \text{ or } Y_n^\delta \in B\big) > \Delta\Big).$$

This allows us to pick a sequence $\delta_k \downarrow 0$ such that Condition (ii) of Lemma 2.3 holds under each $\delta = \delta_k$, which implies

$$\begin{aligned}
&\limsup_{n\to\infty} \epsilon_n^{-1}\mathbf{P}(X_n \in B) \\
&\le \sum_{v\in\mathcal{V}} \mu_v(B^{2\Delta}) + \limsup_{k\to\infty} \limsup_{n\to\infty} \epsilon_n^{-1}\mathbf{P}\Big(\boldsymbol{d}\big(X_n, Y_n^{\delta_k}\big)\mathbb{I}\big(X_n \in B \text{ or } Y_n^{\delta_k} \in B\big) > \Delta\Big) \\
&= \sum_{v\in\mathcal{V}} \mu_v(B^{2\Delta}) \qquad \text{by Condition (i).}
\end{aligned} \tag{1}$$

Analogously, observe the lower bound (for each $n \ge 1$ and $\delta > 0$)

$$\begin{aligned}
\mathbf{P}(X_n \in B) &\ge \mathbf{P}\big(X_n \in B;\ \boldsymbol{d}(X_n, Y_n^\delta) \le \Delta\big) \\
&\ge \mathbf{P}\big(Y_n^\delta \in B_\Delta;\ \boldsymbol{d}(X_n, Y_n^\delta) \le \Delta\big) \\
&\ge \mathbf{P}\big(Y_n^\delta \in B_\Delta\big) - \mathbf{P}\big(Y_n^\delta \in B_\Delta;\ \boldsymbol{d}(X_n, Y_n^\delta) > \Delta\big) \\
&\ge \mathbf{P}\big(Y_n^\delta \in B_\Delta\big) - \mathbf{P}\big(Y_n^\delta \in B \text{ or } X_n \in B;\ \boldsymbol{d}(X_n, Y_n^\delta) > \Delta\big) \\
&\ge \sum_{v\in\mathcal{V}} \mathbf{P}\big(Y_n^\delta \in B_\Delta,\ V_n^\delta = v\big) - \mathbf{P}\big(Y_n^\delta \in B \text{ or } X_n \in B;\ \boldsymbol{d}(X_n, Y_n^\delta) > \Delta\big).
\end{aligned}$$

Again, by Condition (ii), one can pick a sequence $\delta_k \downarrow 0$ such that

$$\begin{aligned}
&\liminf_{n\to\infty} \epsilon_n^{-1}\mathbf{P}(X_n \in B) \\
&\ge \liminf_{k\to\infty} \liminf_{n\to\infty} \sum_{v\in\mathcal{V}} \epsilon_n^{-1}\mathbf{P}\big(Y_n^{\delta_k} \in B_\Delta,\ V_n^{\delta_k} = v\big) \\
&\qquad - \limsup_{k\to\infty} \limsup_{n\to\infty} \epsilon_n^{-1}\mathbf{P}\Big(\boldsymbol{d}\big(X_n, Y_n^{\delta_k}\big)\mathbb{I}\big(X_n \in B \text{ or } Y_n^{\delta_k} \in B\big) > \Delta\Big) \\
&\ge \sum_{v\in\mathcal{V}} \mu_v(B_{2\Delta}) \qquad \text{by Conditions (i) and (ii).}
\end{aligned} \tag{2}$$

Since $|\mathcal{V}| < \infty$, $\mu_v \in \mathbb{M}(\mathbb{S} \setminus \mathbb{C})$ for each $v \in \mathcal{V}$, and $B^{2\Delta}$ is bounded away from $\mathbb{C}$, we have $\sum_{v\in\mathcal{V}} \mu_v(B^{2\Delta}) < \infty$. By sending $\Delta \downarrow 0$ in (1) and (2), it then follows from the continuity of measures that

$$\sum_{v\in\mathcal{V}} \mu_v(B^\circ) \le \liminf_{n\to\infty} \epsilon_n^{-1}\mathbf{P}(X_n \in B) \le \limsup_{n\to\infty} \epsilon_n^{-1}\mathbf{P}(X_n \in B) \le \sum_{v\in\mathcal{V}} \mu_v(B^-).$$

By the arbitrariness of our choice of B, we conclude the proof using Theorem 2.2. □

A.2. Details for Tail Asymptotics of Hawkes Process Clusters

To precisely define the limiting measures $\mathbf{C}_i^{\boldsymbol{j}}$ in (18), we review a few definitions in [2].

Definition A.1 (Type). We say that $\boldsymbol{I} = (I_{k,j})_{k\geq 1,\ j\in[d]}$ is a **type** if

- $I_{k,j} \in \{0,1\}$ for each $k \geq 1$ and $j \in [d]$;
- there exists $\mathcal{K}^{\boldsymbol{I}} \in \mathbb{Z}_+$ such that $\sum_{j\in[d]} I_{k,j} = 0\ \forall k > \mathcal{K}^{\boldsymbol{I}}$ and $\sum_{j\in[d]} I_{k,j} \geq 1\ \forall 1 \leq k \leq \mathcal{K}^{\boldsymbol{I}}$;
- for each $j \in [d]$, we have $\sum_{k\geq 1} I_{k,j} \leq 1$;
- for $k = 1$, the set $\{j \in [d] :\ I_{1,j} = 1\}$ is either empty or contains exactly one element.

Let $\mathscr{I}$ be the collection of all types. For each $\boldsymbol{I} \in \mathscr{I}$, let

$$\boldsymbol{j}^{\boldsymbol{I}} \stackrel{\text{def}}{=} \big\{j \in [d] :\ I_{k,j} = 1 \text{ for some } k \geq 1\big\}, \qquad \boldsymbol{j}_k^{\boldsymbol{I}} \stackrel{\text{def}}{=} \big\{j \in [d] :\ I_{k,j} = 1\big\}\ \ \forall k \geq 1. \tag{3}$$

Intuitively speaking, $\mathcal{K}^{\boldsymbol{I}}$ is the *depth* of $\boldsymbol{I}$, $\boldsymbol{j}^{\boldsymbol{I}}$ is the set of *active indices* in $\boldsymbol{I}$, and $\boldsymbol{j}_k^{\boldsymbol{I}}$ is the set of active indices of $\boldsymbol{I}$ at depth k. For any $\boldsymbol{I} \in \mathscr{I}$, note that: (i) $\boldsymbol{j}^{\boldsymbol{I}} = \emptyset$ (and hence $\mathcal{K}^{\boldsymbol{I}} = 0$) if and only if $I_{k,j} \equiv 0$ for any k, j; and (ii) when $\mathcal{K}^{\boldsymbol{I}} \geq 1$, there uniquely exists some $j_1^{\boldsymbol{I}} \in [d]$ such that $\boldsymbol{j}_1^{\boldsymbol{I}} = \{j_1^{\boldsymbol{I}}\}$. Next, given $\beta > 0$, we define a Borel measure on $(0,\infty)$ by

$$\nu_\beta(dw) \stackrel{\text{def}}{=} \frac{\beta dw}{w^{\beta+1}} \mathbb{I}\{w > 0\}. \tag{4}$$

Given non-empty sets $\mathcal{I} \subseteq [d]$ and $\mathcal{J} \subseteq [d]$, we say that $\{\mathcal{J}(i) :\ i \in \mathcal{I}\}$ is an **assignment of $\mathcal{J}$ to $\mathcal{I}$** if

$$\mathcal{J}(i) \subseteq \mathcal{J} \quad \forall i \in \mathcal{I}; \qquad \bigcup_{i\in\mathcal{I}} \mathcal{J}(i) = \mathcal{J}; \qquad \mathcal{J}(i) \cap \mathcal{J}(i') = \emptyset \quad \forall i \neq i'. \tag{5}$$

We use $\mathbb{T}_{\mathcal{I}\leftarrow\mathcal{J}}$ to denote the set containing all assignments of $\mathcal{J}$ to $\mathcal{I}$. Given non-empty $\mathcal{I} \subseteq [d]$ and $\mathcal{J} \subseteq [d]$, define the mapping

$$g_{\mathcal{I}\leftarrow\mathcal{J}}(\boldsymbol{w}) \stackrel{\text{def}}{=} \sum_{\{\mathcal{J}(i):\ i\in\mathcal{I}\}\in\mathbb{T}_{\mathcal{I}\leftarrow\mathcal{J}}} \prod_{i\in\mathcal{I}} \prod_{j\in\mathcal{J}(i)} w_i \bar{s}_{l^*(j)\leftarrow i}, \quad \forall \boldsymbol{w} = (w_i)_{i\in\mathcal{I}} \in [0,\infty)^{|\mathcal{I}|}. \tag{6}$$

Here, recall that $\bar{\boldsymbol{s}}_j = \mathbf{E}\boldsymbol{S}_j$, $\bar{s}_{j\leftarrow i} = \mathbf{E}S_{j\leftarrow i}$ (see (14)), and $l^*(j)$ is defined in (15).

Given a type $\boldsymbol{I} \in \mathscr{I}$ whose active index set $\boldsymbol{j}^{\boldsymbol{I}}$ is non-empty, recall that when $\boldsymbol{j}^{\boldsymbol{I}} \neq \emptyset$, there uniquely exists some $j_1^{\boldsymbol{I}} \in [d]$ such that $\boldsymbol{j}_1^{\boldsymbol{I}} = \{j_1^{\boldsymbol{I}}\}$. Let

$$\nu^{\boldsymbol{I}}(d\boldsymbol{w}) \stackrel{\text{def}}{=} \bigtimes_{k=1}^{\mathcal{K}^{\boldsymbol{I}}} \Big(\bigtimes_{j\in\boldsymbol{j}_k^{\boldsymbol{I}}} \nu_{\alpha^*(j)}(dw_{k,j}) \Big), \tag{7}$$

$$\mathbf{C}_i^{\boldsymbol{I}}(\cdot) \stackrel{\text{def}}{=} \int \mathbb{I}\Big\{ \sum_{k=1}^{\mathcal{K}^{\boldsymbol{I}}} \sum_{j\in\boldsymbol{j}_k^{\boldsymbol{I}}} w_{k,j} \bar{\boldsymbol{s}}_j \in \cdot \Big\} \Big(\bar{s}_{l^*(j_1^{\boldsymbol{I}})\leftarrow i} \prod_{k=1}^{\mathcal{K}^{\boldsymbol{I}}-1} g_{\boldsymbol{j}_k^{\boldsymbol{I}}\leftarrow\boldsymbol{j}_{k+1}^{\boldsymbol{I}}}(\boldsymbol{w}_k) \Big) \nu^{\boldsymbol{I}}(d\boldsymbol{w}), \tag{8}$$

where we adopt the notations $\boldsymbol{w}_k = (w_{k,j})_{j\in\boldsymbol{j}_k^{\boldsymbol{I}}}$ and $\boldsymbol{w} = (\boldsymbol{w}_k)_{k\in[\mathcal{K}^{\boldsymbol{I}}]}$. Besides, note that $\mathbf{C}_i^{\boldsymbol{I}}(\cdot)$ is supported on the cone $\mathbb{R}^d(\boldsymbol{j}^{\boldsymbol{I}})$ (see (6)). The measures $\mathbf{C}_i^{\boldsymbol{j}}(\cdot)$ in (18) admit the form

$$\mathbf{C}_i^{\boldsymbol{j}} = \sum_{\boldsymbol{I}\in\mathscr{I}:\ \boldsymbol{j}^{\boldsymbol{I}}=\boldsymbol{j}} \mathbf{C}_i^{\boldsymbol{I}}. \tag{9}$$

Next, continuing the discussions in Remark 7, we note that the evaluation of the limiting measures $\check{\mathbf{C}}^{[0,\infty)}_{\mathcal{K};\boldsymbol{\mu}_N}(\cdot)$ in (49) can be addressed by rejection sampling for $\mathbf{C}^{\boldsymbol{I}}_i(\cdot)$ in (8). Indeed, as noted in Remark 7, the key task is to sample under the law $\bar{\mathbf{C}}_{\boldsymbol{j},\bar{\delta}}(\cdot)$ in (50). Furthermore, the evaluation of $\mathbf{C}^{\boldsymbol{I}}_i\big(\bar{\mathbb{R}}^d_>(\boldsymbol{j},\bar{\delta})\big)$ is detailed in Remark 4 of [2]. Then, by (45) and (9), it suffices to check how to sample from probability measures

$$\hat{\mathbf{C}}^{\boldsymbol{I}}_{i,\bar{\delta}}(\cdot) \stackrel{\text{def}}{=} \mathbf{C}^{\boldsymbol{I}}_i\Big(\ \cdot\ \cap\bar{\mathbb{R}}^d_>(\boldsymbol{j},\bar{\delta})\Big)\Big/\mathbf{C}^{\boldsymbol{I}}_i\Big(\bar{\mathbb{R}}^d_>(\boldsymbol{j},\bar{\delta})\Big),$$

where $\boldsymbol{I}\in\mathscr{I}$ is such that $\boldsymbol{j}^{\boldsymbol{I}}=\boldsymbol{j}$ (see Definition A.1). Furthermore, the proof of Lemma 4.10 (b) of [2] implies the existence of some constant $M'<\infty$ such that (we write $\boldsymbol{w}_k=(w_{k,j})_{j\in\boldsymbol{j}^{\boldsymbol{I}}_k}$)

$$\left[\prod_{k=1}^{\mathcal{K}^{\boldsymbol{I}}-1} g_{\boldsymbol{j}^{\boldsymbol{I}}_k\leftarrow\boldsymbol{j}^{\boldsymbol{I}}_{k+1}}(\boldsymbol{w}_k)\right]\cdot\prod_{k=1}^{\mathcal{K}^{\boldsymbol{I}}}\prod_{j\in\boldsymbol{j}^{\boldsymbol{I}}_k} w_{k,j}^{-\alpha^*(j)-1}\le M'\cdot\prod_{k=1}^{\mathcal{K}^{\boldsymbol{I}}}\prod_{j\in\boldsymbol{j}^{\boldsymbol{I}}_k} w_{k,j}^{-\alpha^*(j)}$$

whenever $\sum_{k=1}^{\mathcal{K}^{\boldsymbol{I}}}\sum_{j\in\boldsymbol{j}^{\boldsymbol{I}}_k} w_{k,j}\bar{\boldsymbol{s}}_j\in\bar{\mathbb{R}}^d_>(\boldsymbol{j},\bar{\delta})$. Here, the mapping $g_{\mathcal{I}\leftarrow\mathcal{J}}(\cdot)$ is defined in (6), $\alpha^*(\cdot)$ is defined in (15), and the $\mathcal{K}^{\boldsymbol{I}}$ and $\boldsymbol{j}^{\boldsymbol{I}}_k$'s are specified in Definition A.1. Then, by definitions in (8), to sample from $\hat{\mathbf{C}}^{\boldsymbol{I}}_{i,\bar{\delta}}(\cdot)$ it suffices to set

$$\hat{g}^{\boldsymbol{I}}(\boldsymbol{w})\stackrel{\text{def}}{=}\left[\prod_{k=1}^{\mathcal{K}^{\boldsymbol{I}}-1} g_{\boldsymbol{j}^{\boldsymbol{I}}_k\leftarrow\boldsymbol{j}^{\boldsymbol{I}}_{k+1}}(\boldsymbol{w}_k)\right]\cdot\prod_{k=1}^{\mathcal{K}^{\boldsymbol{I}}}\prod_{j\in\boldsymbol{j}^{\boldsymbol{I}}_k}\left(\frac{\delta}{w_{k,j}}\right)^{\alpha^*(j)+1},\quad \hat{f}^{\boldsymbol{I}}(\boldsymbol{w})\stackrel{\text{def}}{=}\prod_{k=1}^{\mathcal{K}^{\boldsymbol{I}}}\prod_{j\in\boldsymbol{j}^{\boldsymbol{I}}_k}\left(\frac{\delta}{w_{k,j}}\right)^{\alpha^*(j)},$$

pick $\delta>0$ small enough and $M>0$ large enough, and run rejection sampling as follows.

- Independently for each $k\in[\mathcal{K}^{\boldsymbol{I}}]$ and $j\in\boldsymbol{j}^{\boldsymbol{I}}_k$, generate a Pareto (i.e., exact power-law) random variable $W^{(\delta)}_{k,j}$ with lower bound δ and power-law index $\alpha^*(j)-1$; Write $\boldsymbol{W}^{(\delta)}=\big(W^{(\delta)}_{k,j}\big)_{k\in[\mathcal{K}^{\boldsymbol{I}}],j\in\boldsymbol{j}^{\boldsymbol{I}}_k}$.
- Generate $U\sim\mathrm{Unif}(0,1)$.
- If $\sum_{k\in[\mathcal{K}^{\boldsymbol{I}}]}\sum_{j\in\boldsymbol{j}} W^{(\delta)}_{k,j}\bar{\boldsymbol{s}}_j\in\bar{\mathbb{R}}^d_>(\boldsymbol{j},\bar{\delta})$ and $U<\hat{g}^{\boldsymbol{I}}(\boldsymbol{W}^{(\delta)})\big/\big[M\cdot\hat{f}^{\boldsymbol{I}}(\boldsymbol{W}^{(\delta)})\big]$, return

$$\sum_{k\in[\mathcal{K}^{\boldsymbol{I}}]}\sum_{j\in\boldsymbol{j}} W^{(\delta)}_{k,j}\bar{\boldsymbol{s}}_j.$$

  Otherwise, rerun this procedure.

Appendix B: Counterexamples of Topology and Tail Behavior

Example 1. In this two-dimensional example, we demonstrate that the LDP (49) stated in Theorem 3.3 would generally fail w.r.t. the product M_1 topology of $\mathbb{D}[0,T]$, the càdlàg space with compact domain. The intuition behind the counterexample is that, for any immigrant arriving near the end of the time interval $[0,T]$, its cluster size vector may be arbitrarily truncated at time T due to randomness in the birth times of its descendants. As a result, the LDP in (49) generally fails under the product M_1 topology on $\mathbb{D}[0,T]$ for rare event sets B that involve the value of the process at time T.

Without loss of generality, we fix $T=1$, and lighten the notations by writing

$$\mathbb{D}=\mathbb{D}\big([0,1],\mathbb{R}^d\big),\qquad \check{\mathbb{D}}^{\epsilon}_{\mathcal{K}}=\check{\mathbb{D}}^{\epsilon}_{\mathcal{K};\boldsymbol{\mu}_N}[0,1],\qquad \check{\mathbb{D}}^{\epsilon}_{\leqslant\boldsymbol{k}}=\check{\mathbb{D}}^{\epsilon}_{\leqslant\boldsymbol{k};\boldsymbol{\mu}_N}[0,1],$$
$$\check{\mathbf{C}}_{\mathcal{K}}=\check{\mathbf{C}}^{[0,1]}_{\mathcal{K};\boldsymbol{\mu}_N},\qquad \bar{N}_n=\bar{N}^{[0,1]}_n,\qquad \boldsymbol{d}_{\mathcal{P}}=\boldsymbol{d}^{[0,1]}_{\mathcal{P}}.$$

We focus on the case with $d = 2$ and adopt all assumptions imposed in Theorem 3.3. Besides, we impose the following conditions on the fertility functions (see (1)):

$$f_{1\leftarrow 1}(t) = 0, \qquad \forall t > 1, \tag{10}$$

$$f_{2\leftarrow 1}(t) = 0, \qquad \forall t \in [0, 2]. \tag{11}$$

Here, condition (10) implies that the time a type-1 parent waits to give birth to a type-1 child would, with probability 1, take values over $[0, 1]$; condition (11) implies that the time a type-1 parent waits to give birth to a type-2 child is always strictly larger than 2. We note that: (i) it is easy to identify concrete cases that are compatible with both (10)–(11) and the tail condition (48) in Theorem 3.3 under any $\boldsymbol{k}$; for instance, one can take $f_{1\leftarrow 1}(t) = \mathbb{I}_{(0,1)}(t)$, which induces a uniform distribution over $(0, 1)$, and $f_{2\leftarrow 1}(t) = \mathbb{I}_{(2,\infty)}(t) \cdot \exp(-(t-2))$, which induces an exponential distribution with +2 offset; (ii) the exact form of conditions (10)–(11) are not pivotal to the analysis below and are imposed only for simplicity; for instance, the arguments below would still hold after minor modifications if we set $f_{j\leftarrow i}(t) = A_{j\leftarrow i} \exp(-b_{j\leftarrow i} \cdot t)$ (i.e., with exact exponential tails in decay functions, which make the conditional intensity process of the Hawkes process Markovian).

Regarding the tail indices for the clusters, we assume that

$$\alpha^*(1) = \alpha_{1\leftarrow 1} > 1, \qquad \alpha_{2\leftarrow 1} > \alpha_{1\leftarrow 1}, \qquad \alpha^*(2) > \alpha^*(1) + 1, \tag{12}$$

where $\alpha_{i\leftarrow j}$ is the regular variation index for $B_{i\leftarrow j}$ (see Assumption 2 and the discussion right above) and $\alpha^*(\cdot)$ is defined in (15).

We are interested in studying the asymptotics (as $n \to \infty$)

$$\mathbf{P}\big(n^{-1}\boldsymbol{N}(n) \in A\big) = \mathbf{P}\big(\bar{\boldsymbol{N}}_n \in E\big),$$

where $\bar{\boldsymbol{N}}_n = \{\boldsymbol{N}(nt)/n : \ t \in [0, 1]\}$ is the scaled sample path of the Hawkes process $\boldsymbol{N}(t)$ embedded in $\mathbb{D}$, and

$$A \stackrel{\text{def}}{=} \big\{(x_1, x_2)^{\mathrm{T}} \in \mathbb{R}_+^2 : \ x_1 > 1, \ cx_1 > x_2\big\}, \qquad E \stackrel{\text{def}}{=} \big\{\xi \in \mathbb{D} : \ \xi(1) \in A\big\}. \tag{13}$$

To specify the choice of c, we note that by Assumption 3, in $\bar{\boldsymbol{s}}_i = (\bar{s}_{1\leftarrow i}, \bar{s}_{2\leftarrow i})^\top$ for each $i \in \{1, 2\}$ (see (14)) and $\boldsymbol{\mu}_{\boldsymbol{N}} = (\mu_{\boldsymbol{N},1}, \mu_{\boldsymbol{N},2})^\top$ (see (47)), all coordinates are strictly positive. Therefore, we can fix $c > 0$ small enough such that the set $\{\boldsymbol{x} + \boldsymbol{\mu}_{\boldsymbol{N}} : \ \boldsymbol{x} \in \bar{\mathbb{R}}^2(\{1, 2\}, \epsilon)\}$ is bounded away from A for some $\epsilon > 0$; see (7) for the definition of cones $\bar{\mathbb{R}}^d(\boldsymbol{j}, \epsilon)$. More generally, given $a \in (0, \infty)$, there exists $M = M(c, a) \in (0, \infty)$ such that

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \le \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \le \begin{pmatrix} a \\ a \end{pmatrix}, \ y > M \qquad \Longrightarrow \qquad \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} + \begin{pmatrix} y \\ 0 \end{pmatrix} \in A. \tag{14}$$

We start by making a few observations. Recall that $\mathcal{P}_m$ is the collection of non-empty subsets of $[m]$, and hence $\mathcal{P}_2 = \{\{1\}, \{2\}, \{1, 2\}\}$. Also, in the context of Theorem 3.3, recall the definitions of

$$\check{\mathbb{D}}^{\epsilon}_{\leqslant \boldsymbol{k}} = \bigcup_{\substack{\mathcal{K} \in \mathbb{Z}_+^{\mathcal{P}_2}: \\ \mathcal{K} \notin \mathbb{A}(\boldsymbol{k}), \ \check{c}(\mathcal{K}) \le c(\boldsymbol{k})}} \check{\mathbb{D}}^{\epsilon}_{\mathcal{K}}, \qquad \check{c}(\mathcal{K}) = \sum_{\boldsymbol{j} \in \mathcal{P}_2} \mathcal{K}_{\boldsymbol{j}} \cdot \big(\boldsymbol{\alpha}(\boldsymbol{j}) - 1\big), \qquad c(\boldsymbol{k}) = \sum_{i=1,2} k_i \cdot \big(\alpha(\{i\}) - 1\big), \tag{15}$$

where $\boldsymbol{\alpha}(\boldsymbol{j}) = 1 + \sum_{i\in\boldsymbol{j}} \big(\alpha^*(i) - 1\big)$ is defined in (16). In particular, note that $\alpha(\{1\}) = \alpha^*(1)$, $\alpha(\{2\}) = \alpha^*(2)$, and $\alpha(\{1,2\}) = \alpha^*(1) + \alpha^*(2) - 1$. Also, we introduce notations $\mathcal{K}^{\boldsymbol{i}} = (\mathcal{K}^{\boldsymbol{i}}_{\boldsymbol{j}})_{\boldsymbol{j}\in\mathcal{P}_2} \in \mathbb{Z}_+^{\mathcal{P}_2}$ by setting $\mathcal{K}^{\boldsymbol{i}}_{\boldsymbol{j}} = 1$ if $\boldsymbol{j} = \boldsymbol{i}$, and $\mathcal{K}^{\boldsymbol{i}}_{\boldsymbol{j}} = 0$ otherwise. For instance, in $\mathcal{K}^{\{1,2\}} = (\mathcal{K}^{\{1,2\}}_{\boldsymbol{j}})_{\boldsymbol{j}\in\mathcal{P}_2}$, we have $\mathcal{K}^{\{1,2\}}_{\{1,2\}} = 1$ and $\mathcal{K}^{\{1,2\}}_{\{1\}} = \mathcal{K}^{\{1,2\}}_{\{2\}} = 0$. By (12) and (15), we have

$$\breve{\mathbb{D}}^{\epsilon}_{\leqslant(0,1)} \subseteq \bigcup_{\mathcal{K}\in\mathbb{Z}_+^{\mathcal{P}_2}:\ \mathcal{K}_{\{2\}}=\mathcal{K}_{\{1,2\}}=0} \breve{\mathbb{D}}^{\epsilon}_{\mathcal{K}}.$$

Furthermore, for any $\mathcal{K} \in \mathbb{Z}_+^{\mathcal{P}_2}$ such that $\mathcal{K}_{\{2\}} = \mathcal{K}_{\{1,2\}} = 0$ and any $\xi \in \breve{\mathbb{D}}^{\epsilon}_{\mathcal{K}}$, we have

$$\xi(1) \in \bar{\mathbb{R}}^2(\{1\}, \epsilon) + \boldsymbol{\mu}_{\boldsymbol{N}}.$$

Meanwhile, recall the definitions of the sets A and E in (13). It has been noted above that $\bar{\mathbb{R}}^2(\{1,2\}, \epsilon) + \boldsymbol{\mu}_{\boldsymbol{N}}$ is bounded away from A, which implies that the set E is bounded away from $\breve{\mathbb{D}}^{\epsilon}_{\leqslant(0,1)}$ under $\boldsymbol{d}_{\mathcal{P}}$. Now, suppose that the asymptotics (49) stated in Theorem 3.3 hold w.r.t. the product M_1 topology of $\mathbb{D}$ (i.e., on the metric space $(\mathbb{D}, \boldsymbol{d}_{\mathcal{P}})$). Then, we are led to believe that

$$\mathbf{P}\big(\bar{\boldsymbol{N}}_n \in E\big) = O\Big(\breve{\lambda}_{(0,1)}(n)\Big), \qquad \text{as } n \to \infty. \tag{16}$$

However, our analysis below disproves (16) by verifying that

$$\liminf_{n\to\infty} \mathbf{P}\big(\bar{\boldsymbol{N}}_n \in E\big) \big/ \mathbf{P}(B_{1\leftarrow 1} > n) > 0. \tag{17}$$

Indeed, by Assumption 2, we have $\mathbf{P}(B_{1\leftarrow 1} > n) \in \mathcal{RV}_{-\alpha_{1\leftarrow 1}}(n)$; by (12) and (15), we have $\breve{\lambda}_{(0,1)}(n) \in \mathcal{RV}_{-c((0,1))}(n)$ with $c\big((0,1)\big) = \alpha^*(2) - 1 > \alpha_{1\leftarrow 1}$. Then, by

$$\liminf_{n\to\infty} \frac{\mathbf{P}\big(\bar{\boldsymbol{N}}_n \in E\big)}{\breve{\lambda}_{(0,1)}(n)} \geq \liminf_{n\to\infty} \frac{\mathbf{P}\big(\bar{\boldsymbol{N}}_n \in E\big)}{\mathbf{P}(B_{1\leftarrow 1} > n)} \cdot \liminf_{n\to\infty} \frac{\mathbf{P}(B_{1\leftarrow 1} > n)}{\breve{\lambda}_{(0,1)}(n)},$$

we end up with $\liminf_{n\to\infty} \mathbf{P}\big(\bar{\boldsymbol{N}}_n \in E\big) \big/ \breve{\lambda}_{(0,1)}(n) = \infty$, which is a clear contradiction to (16). We thus confirm that it is generally not possible to strengthen the characterizations in (49) stated in Theorem 3.3 w.r.t. the product M_1 topology of $\mathbb{D}[0,\infty)$ to those on $\mathbb{D}[0,T]$.

Below, we verify Claim (17). To proceed, we adopt the notations in (99)–(103) for the cluster representation of Hawkes processes. Specifically, for each $j \in \{1,2\}$, we use the sequence $\big(T^{C}_{j;k}\big)_{k\geq 1}$ to denote the arrival times of type-j immigrants, which are generated by a Poisson process $G_j(t)$ with rate c_j (see (1)). For $\boldsymbol{N}^{O}_{(T^{C}_{j;k},j)}$, the cluster process induced by the k^{th} type-j immigrant, we denote its size by

$$S^{(k)}_{i\leftarrow j} \stackrel{\text{def}}{=} \sum_{m\geq 0} \mathbb{I}\Big\{A^{O}_{j;k}(m) = i\Big\}, \qquad \boldsymbol{S}^{(k)}_{j} \stackrel{\text{def}}{=} \Big(S^{(k)}_{1\leftarrow j}, S^{(k)}_{2\leftarrow j}, \ldots, S^{(k)}_{d\leftarrow j}\Big)^{\mathrm{T}}.$$

Next, we define the event (with $\boldsymbol{\mu}_{\boldsymbol{N}}$ defined in (47))

$$(\mathrm{I}) = \left\{ n^{-1} \sum_{j\in\{1,2\}} \sum_{k\geq 1: T^{C}_{j;k} \leq n-2} \boldsymbol{S}^{(k)}_{j} \leq 2\boldsymbol{\mu}_{\boldsymbol{N}} \right\},$$

which represents the case where the accumulated size of all clusters arrived by time $n-2$ is upper bounded by the vector $2n\boldsymbol{\mu}_{\boldsymbol{N}}$. Also, recall that we use $G_j(t)$ to denote the Poisson process generating the arrival times $T^C_{j;k}$ for the immigrants. Let

$$\text{(II)} = \Big\{G_1(n-1) - G_1(n-2) = 1;\ G_1(n) - G_1(n-1) = 0;\ G_2(n) - G_2(n-2) = 0\Big\}.$$

That is, on event (II), there is only one type-1 immigrant who arrived during the time window $(n-2, n]$, with the arrival time lying in $(n-2, n-1]$, and no type-2 immigrant arrived during $(n-2, n]$. On this event, for the only type-1 immigrant that arrived during $(n-2, n]$, we use $B_{1\leftarrow 1}$ to specifically denote the number of its type-1 children (in one generation). Let

$$\text{(III)} = \big\{B_{1\leftarrow 1} > nM\big\}.$$

Here, by (14), we fix some M large enough such that

$$\begin{pmatrix}0\\0\end{pmatrix} \le \begin{pmatrix}w_1\\w_2\end{pmatrix} \le 2\boldsymbol{\mu}_{\boldsymbol{N}},\ y > M \qquad \Longrightarrow \qquad \begin{pmatrix}w_1\\w_2\end{pmatrix} + \begin{pmatrix}y\\0\end{pmatrix} \in A. \tag{18}$$

Now, we make some observations. First, on the event (II), by (10) we know that all the $B_{1\leftarrow 1}$ children of the type-1 ancestor must have arrived by time n, and they may further give birth to more type-1 offspring by time n. On the other hand, by (11), we know that any type-2 offspring in this cluster, induced by the type-1 immigrant who arrived during $(n-2, n]$, will have to arrive after time n. Then, by (18), on event (I) $\cap$ (II) $\cap$ (III) we must have $n^{-1}\boldsymbol{N}(n) \in A$. This leads to

$$\liminf_{n\to\infty} \frac{\mathbf{P}\big(\bar{\boldsymbol{N}}_n \in E\big)}{\mathbf{P}(B_{1\leftarrow 1} > n)} = \liminf_{n\to\infty} \frac{\mathbf{P}\big(n^{-1}\boldsymbol{N}(n) \in A\big)}{\mathbf{P}(B_{1\leftarrow 1} > n)} \ge \lim_{n\to\infty} \mathbf{P}\big((\text{I})\big) \cdot \mathbf{P}\big((\text{II})\big) \cdot \lim_{n\to\infty} \frac{\mathbf{P}(B_{1\leftarrow 1} > nM)}{\mathbf{P}(B_{1\leftarrow 1} > n)},$$

where the inequality follows from the independent and stationary increments in Poisson processes $G_j(t)$. By the law of large numbers, we get $\lim_{n\to\infty} \mathbf{P}\big((\text{I})\big) = 1$. By the law of Poisson processes, we have

$$\mathbf{P}\big((\text{II})\big) = \mathbf{P}\Big(G_1(1) = 1;\ G_1(2) - G_1(1) = 0;\ G_2(2) = 0\Big) > 0.$$

Lastly, by the regularly varying law of $B_{1\leftarrow 1}$, we get $\lim_{n\to\infty} \frac{\mathbf{P}(B_{1\leftarrow 1} > nM)}{\mathbf{P}(B_{1\leftarrow 1} > n)} = M^{-\alpha_{1\leftarrow 1}} > 0$. This concludes the proof of Claim (17).

Example 2. This example demonstrates that, even in the univariate setting, we need sufficiently tight power-law bounds on the tail behavior of the decay functions (like in (48)) for the large deviations asymptotics stated in Theorem 3.3 to hold. The intuition of this is that, without sufficiently tight tail bounds on the decay functions, some rare events may be driven by a single large cluster with a long lifetime—effectively a big jump spread broadly across the entire timeline—rather than by multiple large clusters.

Specifically, we consider a univariate Hawkes process $N(t)$ with initial value $N(0) = 0$ and conditional intensity $h(t) = c + \int_0^t \tilde{B}(s) f(t-s) dN(s)$, where the constant $c > 0$ is the immigration rate, the decay function $f(\cdot)$ takes the form

$$f(t) = \frac{\theta}{(1+t)^{\theta+1}}, \qquad \forall t \ge 0, \tag{19}$$

for some $\theta > 1$ (i.e., $\int_t^\infty f(s)ds = (1+t)^{-\theta}$, and $\|f\|_1 = \int_0^\infty f(t)dt = 1$), and the excitation rates $\tilde{B}(s)$ are independent copies of some non-negative random variable $\tilde{B}$ with law

$$\mathbf{P}(\tilde{B} > x) = \frac{1}{\max\{1, x/\tilde{c}\}^\alpha}, \qquad \forall x \in \mathbb{R}. \tag{20}$$

Here, we have $\alpha > 1$ for the power-law tail index; the constant $\tilde{c} > 0$ is small enough such that $\mathbf{E}[\tilde{B}] \cdot \|f\|_1 = \mathbf{E}[\tilde{B}] < 1$ (so the sub-critical condition in Assumption 1 also holds). By (47), in this univariate setting we have $\mu_N = \frac{c}{1-\mathbf{E}[\tilde{B}]\cdot\|f\|_1} = \frac{c}{1-\mathbf{E}[\tilde{B}]}$.

Recall that $\bar{N}_n(t) = N(nt)/n$, and that $\bar{\boldsymbol{N}}_n^{[0,\infty)} \stackrel{\text{def}}{=} \{\bar{N}_n(t) : \ t \geq 0\}$ is the scaled sample path of the Hawkes process $N(t)$. In this example, we study events of the form $\{\bar{\boldsymbol{N}}_n^{[0,\infty)} \in A(k)\}$ with

$$A(k) \stackrel{\text{def}}{=} \big\{\xi \in \mathbb{D}([0,\infty), \mathbb{R}) : \ \xi(t_l + u) - \xi(t_l) > \mu_N \ \forall l = 0, 1, \ldots, k-1\big\}, \tag{21}$$

where

- k is some positive integer;
- the t_l's are such that $0 = t_0 < t_1 < t_2 < \ldots < t_{k-1} < 1$;
- Furthermore, $u \in (0,1)$ is small enough such that, across $l = 0, 1, \ldots, k-1$, the intervals $I_l = [t_l, t_l + u]$ are mutually disjoint and are all contained within $[0,1]$.

In other words, for any path ξ that belongs to the set $A(k)$, the increment of ξ over each of the mutually disjoint intervals I_l is bounded from below by μ_N. In particular, note that $A(k)$ is bounded away from (by adapting (27) and (37) to the univariate setting)

$$\breve{\mathbb{D}}^\epsilon_{\leqslant k;\mu_N}[0,\infty) = \bigcup_{k'=0,1,\ldots,k-1} \breve{\mathbb{D}}^\epsilon_{k';\mu_N}[0,\infty) = \breve{\mathbb{D}}^\epsilon_{k-1;\mu_N}[0,\infty)$$

under $\boldsymbol{d}_{M_1}^{[0,\infty)}$ for any $\epsilon > 0$ small enough: indeed, a linear path of slope μ_N would only increase by $u \cdot \mu_N < \mu_N$ over each interval I_l, meaning that at least one jump needs to be added to the linear path over each $(I_l)_{l=0,1,\ldots,k-1}$ for it to enter the set $A(k)$. Also, when there is no ambiguity about the choice of k, we simply write $A = A(k)$.

We note that, regardless of the value of the tail indices θ in (19) for the decay function and α in (20) for the law of the excitation rates, the tail condition (48) would be violated eventually for all k large enough. Indeed, by (17) and (35), in this univariate setting we have

$$\breve{\lambda}_k(x) = \Big(x\mathbf{P}(\tilde{B} > x)\Big)^k = O(x^{-k(\alpha-1)}).$$

Therefore, condition (48) fails if $\theta < k(\alpha-1)+1$. Suppose that Theorem 3.3 holds even after dropping the tail condition (48) on the decay functions, then we are led to believe that

$$\mathbf{P}\big(\bar{\boldsymbol{N}}_n^{[0,\infty)} \in A\big) = O\big(\breve{\lambda}_k(n)\big) = O\big(n^{-k(\alpha-1)}\big), \quad \text{as } n \to \infty. \tag{22}$$

However, our analysis below derives the lower bound

$$\liminf_{n\to\infty} \mathbf{P}\big(\bar{\boldsymbol{N}}_n^{[0,\infty)} \in A\big) \cdot n^{\alpha(\theta+1)-1} > 0. \tag{23}$$

This clearly contradicts (22) for all k large enough such that $k(\alpha-1) > \alpha(\theta+1) - 1$. In summary, even in the univariate setting, it is generally not possible to remove tail bounds on the decay functions from Theorem 3.3.

The rest of this example is devoted to establishing the lower bound (23), and we proceed by defining a few events. Let $0 < T_1 < T_2 < \dots$ be a sequence generated by a Poisson process with rate c. This sequence denotes the arrival time sequence of the immigrants in $N(t)$. Also, let $(B_i)_{i\geq 1}$ be independent copies of $B \overset{\mathcal{D}}{=} \text{Poisson}(\tilde{B} \cdot \|f\|_1) \overset{\mathcal{D}}{=} \text{Poisson}(\tilde{B})$ (recall that we have $\|f\|_1 = 1$). Note that $\mathbf{P}(B > x) \sim \mathbf{P}(\tilde{B} > x)$ as $x \to \infty$. In the cluster representation of $N(t)$, each B_i denotes the count of first-generation offspring of the immigrant arriving at time T_i. We fix $C \in (0, \infty)$ large enough such that

$$\frac{C\theta u}{4} > \mu_N, \tag{24}$$

and define the event

$$(A : *) \overset{\text{def}}{=} \left\{ \exists i \text{ such that } T_i \leq \frac{nu}{2},\ B_i > Cn^{\theta+1} \right\}. \tag{25}$$

That is, on the event $(A : *)$, an immigrant arrives before time $\frac{nu}{2}$ and plans to give birth to more than $n^{\theta+1}$ descendants in one generation. Next, on this event, we use i^* to denote the smallest i such that the conditions in (25) hold. Let the sequence $(T_k^{(i^*)})_{k\geq 1}$ be independent copies with density $f(\cdot)/\|f\|_1 = f(\cdot)$; this sequence denotes the gap between the arrival times of the i^*-th immigrant and its first-generation offspring. Using these notations, we define the event

$$(A : 0) \overset{\text{def}}{=} \left\{ \#\left\{ k \leq \lceil Cn^{\theta+1} \rceil :\ T_k^{(i^*)} \leq nu/2 \right\} > \mu_N \cdot n \right\}. \tag{26}$$

On the event $(A : *) \cap (A : 0)$, note that in the Hawkes process $N(t)$ we must have more than $\mu_N \cdot n$ points arriving by time nu. Indeed, on this event, there is an immigrant who arrives by time $nu/2$ and gives birth to more than $n \cdot \mu_N$ of its first-generation descendants within $nu/2$ time after its arrival. Then, under the $O(n)$ scaling for both time and space in $\bar{N}_n^{[0,\infty)}$, we have $\bar{N}_n(u) - \bar{N}_n(0) > \mu_N$. Besides, for each $l \in [k-1]$ we define

$$(A : l) \overset{\text{def}}{=} \left\{ \#\left\{ k \leq \lceil Cn^{\theta+1} \rceil :\ T_k^{(i^*)} \in \left[nt_l,\ nt_l + \frac{nu}{2} \right] \right\} > \mu_N \cdot n \right\}. \tag{27}$$

Similarly, on the event $(A : *) \cap (A : l)$ we must have more than $\mu_N \cdot n$ points arriving during $[nt_l, n(t_l + u)]$. Then, under the $O(n)$ scaling for both time and space in $\bar{N}_n^{[0,\infty)}$, we have $\bar{N}_n(t_l + u) - \bar{N}_n(t_l) > \mu_N$. In summary, on the event $(A : *) \cap \left(\bigcup_{l=0,1,\dots,k} (A : l) \right)$, we have $\bar{N}_n^{[0,\infty)} \in A$. Furthermore, if we can verify that

$$\liminf_{n\to\infty} \mathbf{P}\big((A : *)\big) \cdot n^{\alpha(\theta+1)-1} > 0, \tag{28}$$

$$\lim_{n\to\infty} \mathbf{P}\big((A : l) \,\big|\, (A : *)\big) = 1, \qquad \forall l = 0, 1, \dots, k-1, \tag{29}$$

then, the lower bound (23) follows immediately.

Verification of Claim (28). Throughout, we use $\text{Poisson}(\lambda)$ to denote a generic copy of a Poisson random variable with rate λ. By the law of a compound Poisson process,

$$\mathbf{P}\big((A : *)\big) = \mathbf{P}\left(\text{Poisson}\left(\frac{un}{2} \cdot c \cdot \mathbf{P}(B > Cn^{\theta+1}) \right) \geq 1 \right) = 1 - \exp\left(-\frac{cu}{2} \cdot n\mathbf{P}(B > Cn^{\theta+1}) \right).$$

Then, by $\mathbf{P}(B > x) \sim \mathbf{P}(\tilde{B} > x)$ as $x \to \infty$ and (20),

$$\lim_{n\to\infty} \mathbf{P}\big((A:*)\big) \cdot n^{\alpha(\theta+1)-1} = \frac{cu}{2} \cdot \lim_{n\to\infty} n^{\alpha(\theta+1)-1} \cdot n\mathbf{P}(B > Cn^{\theta+1}) = \frac{cu}{2} \cdot \big(C/\tilde{c}\big)^{-\alpha} > 0.$$

Verification of Claim (29), ($l = 0$). We first verify the case where $l = 0$. Since the $T_k^{(i^*)}$'s are independent copies with density $f(\cdot)$, we have

$$\begin{aligned}
&\mathbf{P}\Big(\big((A:0)\big)^c \;\Big|\; (A:*)\Big) \\
&= \mathbf{P}\Big(\text{Binomial}\Big(\lceil Cn^{\theta+1}\rceil, \int_0^{nu/2} f(t)dt\Big) \le \mu_N n\Big) \\
&\le \mathbf{P}\Big(\text{Binomial}\Big(\lceil Cn^{\theta+1}\rceil, 1/2\Big) \le \mu_N n\Big) \quad \text{for any } n \text{ large enough due to } \|f\|_1 = 1 \\
&\le \mathbf{P}\Big(\text{Binomial}\Big(\lceil Cn^{\theta+1}\rceil, 1/2\Big) \le \lceil Cn^{\theta+1}\rceil/4\Big) \quad \text{for any } n \text{ large enough due to } \theta > 0.
\end{aligned}$$

Then, by the law of large numbers,

$$\lim_{n\to\infty} \mathbf{P}\Big(\big((A:0)\big)^c \;\Big|\; (A:*)\Big) \le \lim_{n\to\infty} \mathbf{P}\Big(\text{Binomial}\Big(\lceil Cn^{\theta+1}\rceil, 1/2\Big) < \lceil Cn^{\theta+1}\rceil/4\Big) = 0.$$

This verifies that $\lim_{n\to\infty} \mathbf{P}\big((A:0) \;\big|\; (A:*)\big) = 1$.

Verification of Claim (29), ($l = 1, 2, \ldots, k-1$). Next, we take any $l \in [k-1]$. Note that

$$\mathbf{P}\Big(\big((A:l)\big)^c \;\Big|\; (A:*)\Big) = \mathbf{P}\Big(\text{Binomial}\Big(\lceil Cn^{\theta+1}\rceil, \underbrace{\int_{nt_l}^{n(t_l+\frac{u}{2})} f(t)dt}_{\stackrel{\text{def}}{=} F_l}\Big) \le \mu_N n\Big).$$

By (19) and that $t_l + u < 1$,

$$F_l = (nt_l)^{-\theta} - \Big(nt_l + \frac{nu}{2}\Big)^{-\theta} \ge \frac{\theta}{(1+n)^{\theta+1}} \cdot \frac{nu}{2}.$$

As a result, for any n large enough we have $F_l \ge \frac{\theta u}{4} \cdot n^{-\theta}$, and hence

$$\mathbf{P}\Big(\big((A:l)\big)^c \;\Big|\; (A:*)\Big) \le \mathbf{P}\Big(\text{Binomial}\Big(\lceil Cn^{\theta+1}\rceil, \frac{\theta u}{4} \cdot n^{-\theta}\Big) \le \mu_N n\Big).$$

By Chernoff bounds, for any $w > 0$ we have

$$\mathbf{P}\Big(\big((A:l)\big)^c \;\Big|\; (A:*)\Big) \le \exp(w\mu_N n) \cdot \Big(1 - \frac{\theta u}{4n^\theta} + \frac{\theta u e^{-w}}{4n^\theta}\Big)^{Cn^{\theta+1}}.$$

Then, due to $(1+x_n)^n \sim \exp(nx_n)$ when $\lim_{n\to\infty} x_n = 0$, we get (for any $w > 0$)

$$\limsup_{n\to\infty} \mathbf{P}\Big(\big((A:l)\big)^c \;\Big|\; (A:*)\Big) \le \exp\bigg(-Cn^{\theta+1}\cdot(1-e^{-w})\cdot\frac{\theta u}{4n^\theta} + w\mu_N n\bigg)$$
$$= \exp\bigg(-n\cdot\Big(\frac{C\theta u(1-e^{-w})}{4} - w\mu_N\Big)\bigg).$$

By our choice of C in (24), it holds for any $w > 0$ small enough that $\frac{C\theta u(1-e^{-w})}{4} - w\mu_N > 0$, which implies $\limsup_{n\to\infty} \mathbf{P}\Big(\big((A:l)\big)^c \;\Big|\; (A:*)\Big) = 0$ and hence $\lim_{n\to\infty} \mathbf{P}\big((A:l) \;\big|\; (A:*)\big) = 1$.

Appendix C: Proof for Large Deviations of Lévy Processes with $\mathcal{MHRV}$ Increments

Without loss of generality, we prove Theorem 3.2 for $T = 1$. Besides, considering the arbitrariness of the drift constant $\boldsymbol{c_L}$ in (24), we can w.l.o.g. impose the following assumption and focus on centered Lévy processes for the proof of Theorem 3.2.

Assumption 1 (WLOG Assumption for Theorem 3.2). $\boldsymbol{\mu_L} = \mathbf{0}$ *in* (24)*, and* $T = 1$.

Henceforth in Section C, we also lighten notations in the proofs by writing

$$\check{\mathbb{D}}^{\epsilon}_{\mathcal{K}} \stackrel{\text{def}}{=} \check{\mathbb{D}}^{\epsilon}_{\mathcal{K};\mathbf{0}}[0,1], \quad \check{\mathbb{D}}^{\epsilon}_{\leqslant \boldsymbol{k}} \stackrel{\text{def}}{=} \check{\mathbb{D}}^{\epsilon}_{\leqslant \boldsymbol{k};\mathbf{0}}[0,1], \quad \check{\mathbf{C}}_{\mathcal{K}} \stackrel{\text{def}}{=} \check{\mathbf{C}}^{[0,1]}_{\mathcal{K};\mathbf{0}}, \quad \bar{\boldsymbol{L}}_n \stackrel{\text{def}}{=} \bar{\boldsymbol{L}}^{[0,1]}_n, \quad \boldsymbol{d}_{J_1} \stackrel{\text{def}}{=} \boldsymbol{d}^{[0,1]}_{J_1}. \tag{30}$$

C.1. Proof of Theorem 3.2

We start by identifying the *large jumps* in $\bar{\boldsymbol{L}}_n$. Given $n \in \mathbb{N}$ and $\delta > 0$, let

$$\tau_n^{>\delta}(k) \stackrel{\text{def}}{=} \inf\Big\{t > \tau_n^{>\delta}(k-1) : \big\|\Delta\bar{\boldsymbol{L}}_n(t)\big\| > \delta\Big\}, \;\; \forall k \ge 1, \qquad \tau_n^{>\delta}(0) \stackrel{\text{def}}{=} 0; \tag{31}$$

$$\boldsymbol{W}_n^{>\delta}(k) \stackrel{\text{def}}{=} \Delta\bar{\boldsymbol{L}}_n\Big(\tau_n^{>\delta}(k)\Big), \qquad \forall k \ge 1. \tag{32}$$

In (31), note that $\Delta\bar{\boldsymbol{L}}_n(t) = \Delta\boldsymbol{L}(nt)/n$. Intuitively speaking, the sequence $\big(\tau_n^{>\delta}(k)\big)_{k\ge1}$ marks the arrival times of jumps in $\bar{\boldsymbol{L}}_n(t)$ with size (in terms of L_1 norm) larger than δ, which correspond to jumps of size larger than $n\delta$ in $\boldsymbol{L}_n(t)$; the sequence $\big(\boldsymbol{W}_n^{>\delta}(k)\big)_{k\ge1}$ represents the sizes of the large jumps.

Building upon the definitions in (31)–(32), we introduce a *large-jump approximation* for $\bar{\boldsymbol{L}}_n(t)$. Specifically, given $\delta > 0$ and $n \ge 1$, let

$$\hat{\boldsymbol{L}}_n^{>\delta}(t) \stackrel{\text{def}}{=} \sum_{k\ge1} \boldsymbol{W}_n^{>\delta}(k)\mathbb{I}_{[\tau_n^{>\delta}(k),1]}(t), \qquad \forall t \in [0,1], \tag{33}$$

and $\hat{\boldsymbol{L}}_n^{>\delta} \stackrel{\text{def}}{=} \{\hat{\boldsymbol{L}}_n^{>\delta}(t) : \; t \in [0,1]\}$. Clearly, $\hat{\boldsymbol{L}}_n^{>\delta}(t)$ is a step function (i.e., piece-wise constant) in $\mathbb{D} = \mathbb{D}\big([0,1],\mathbb{R}^d\big)$ that vanishes at the origin and approximates $\bar{\boldsymbol{L}}_n(t)$ by only keeping the large jumps (under threshold δ).

Our proof of Theorem 3.2 hinges on Propositions C.2 and C.3 that outline the following key steps: (i) first, we establish the asymptotic equivalence between $\bar{\boldsymbol{L}}_n$ and $\hat{\boldsymbol{L}}_n^{>\delta}$ as in Lemma 2.3; (ii) next, we

provide a detailed asymptotic analysis for the law of large jumps. To carry out these two steps, we further classify the large jumps $\boldsymbol{W}_n^{>\delta}(k)$ based on their locations w.r.t. some cones $\bar{\mathbb{R}}^{>\delta}(\boldsymbol{j})$. Specifically, let

$$\bar{\mathbb{R}}^{>\delta}(\boldsymbol{j}) \stackrel{\text{def}}{=} \left\{\boldsymbol{x} \in \mathbb{R}_+^d : \|\boldsymbol{x}\| > \delta,\ \boldsymbol{x} \in \bar{\mathbb{R}}^d(\boldsymbol{j}, \delta) \setminus \bar{\mathbb{R}}^d_{\lneq}(\boldsymbol{j}, \delta)\right\}, \qquad \forall \delta > 0,\ \boldsymbol{j} \in \mathcal{P}_d. \tag{34}$$

We highlight several useful properties regarding $\bar{\mathbb{R}}^{>\delta}(\boldsymbol{j})$.

- First, note that $\bigcup_{\boldsymbol{j} \in \mathcal{P}_d} \bar{\mathbb{R}}^{>\delta}(\boldsymbol{j}) = \{\boldsymbol{x} \in \mathbb{R}_+^d :\ \|\boldsymbol{x}\| > \delta,\ \boldsymbol{x} \in \bar{\mathbb{R}}([d], \delta)\}$. Indeed, it is obvious that $\bigcup_{\boldsymbol{j} \in \mathcal{P}_d} \bar{\mathbb{R}}^{>\delta}(\boldsymbol{j}) \subseteq \{\boldsymbol{x} \in \mathbb{R}_+^d :\ \|\boldsymbol{x}\| > \delta,\ \boldsymbol{x} \in \bar{\mathbb{R}}([d], \delta)\}$. On the other hand, given any $\delta > 0$ and $\boldsymbol{x} \in \bar{\mathbb{R}}^d([d], \delta)$ with $\|\boldsymbol{x}\| > \delta$, the argument minimum

  $$\boldsymbol{j}(\boldsymbol{x}) \stackrel{\text{def}}{=} \operatorname*{arg\,min}_{\boldsymbol{j} \in \mathcal{P}_d :\ \boldsymbol{x} \in \bar{\mathbb{R}}^d(\boldsymbol{j}, \delta)} \alpha(\boldsymbol{j})$$

  uniquely exists under the condition $\alpha(\boldsymbol{j}) \neq \alpha(\boldsymbol{j}')\ \forall \boldsymbol{j}, \boldsymbol{j}' \in \mathcal{P}_d$ in Assumption 5. Then, we must have $\boldsymbol{x} \in \bar{\mathbb{R}}^{>\delta}(\boldsymbol{j}(\boldsymbol{x}))$ due to the definition of $\bar{\mathbb{R}}^d_{\lneq}(\boldsymbol{j}, \delta) = \bigcup_{\boldsymbol{j}' \subseteq [k]:\ \boldsymbol{j}' \neq \boldsymbol{j},\ \alpha(\boldsymbol{j}') \leq \alpha(\boldsymbol{j})} \bar{\mathbb{R}}^d(\boldsymbol{j}', \delta)$. This confirms that $\bigcup_{\boldsymbol{j} \in \mathcal{P}_d} \bar{\mathbb{R}}^{>\delta}(\boldsymbol{j}) \supseteq \{\boldsymbol{x} \in \mathbb{R}_+^d :\ \|\boldsymbol{x}\| > \delta,\ \boldsymbol{x} \in \bar{\mathbb{R}}([d], \delta)\}$.
- Next, we note that $\bar{\mathbb{R}}^{>\delta}(\boldsymbol{j}) \cap \bar{\mathbb{R}}^{>\delta}(\boldsymbol{j}') = \emptyset$ holds for any $\boldsymbol{j}, \boldsymbol{j}' \in \mathcal{P}_d$ with $\boldsymbol{j} \neq \boldsymbol{j}'$. To see why, we assume w.l.o.g. that $\alpha(\boldsymbol{j}') \leq \alpha(\boldsymbol{j})$. Then, due to $\boldsymbol{j}' \neq \boldsymbol{j}$ and $\alpha(\boldsymbol{j}') \leq \alpha(\boldsymbol{j})$, we have $\bar{\mathbb{R}}^{>\delta}(\boldsymbol{j}') \subseteq \bar{\mathbb{R}}^d_{\lneq}(\boldsymbol{j}, \delta)$. Therefore, for any $\boldsymbol{x} \in \bar{\mathbb{R}}^{>\delta}(\boldsymbol{j}')$, we must have $\boldsymbol{x} \notin \bar{\mathbb{R}}^d(\boldsymbol{j}, \delta) \setminus \bar{\mathbb{R}}^d_{\lneq}(\boldsymbol{j}, \delta)$. Repeating this argument for all $\boldsymbol{j}, \boldsymbol{j}' \in \mathcal{P}_d$, we confirm that $\bar{\mathbb{R}}^{>\delta}(\boldsymbol{j}) \cap \bar{\mathbb{R}}^{>\delta}(\boldsymbol{j}') = \emptyset$.
- In summary, the collection of sets $\bar{\mathbb{R}}^{>\delta}(\boldsymbol{j})$ provide a partition of $\{\boldsymbol{x} \in \mathbb{R}_+^d :\ \|\boldsymbol{x}\| > \delta,\ \boldsymbol{x} \in \bar{\mathbb{R}}^d([d], \delta)\}$. Meanwhile, note that

  $$\sum_{\boldsymbol{j} \in \mathcal{P}_d} \mathbf{C}_{\boldsymbol{j}}\left(\partial \bar{\mathbb{R}}^{>\delta}(\boldsymbol{j})\right) = 0, \qquad \text{for all but countably many } \delta > 0, \tag{35}$$

  where we use $\partial E \stackrel{\text{def}}{=} E^- \setminus E^\circ$ to denote the boundary of the set E. This claim is a consequence of the definition of $\mathcal{MHRV}$, and we state the proof below.

Lemma C.1. *Let Assumption 5 hold. Let $\bar{\mathbb{R}}^{>\delta}(\boldsymbol{j})$ be defined as in (34). For each $\boldsymbol{j} \in \mathcal{P}_d$,*

$$\mathbf{C}_{\boldsymbol{j}}\left(\partial \bar{\mathbb{R}}^{>\delta}(\boldsymbol{j})\right) = 0, \qquad \textit{for all but countably many } \delta > 0. \tag{36}$$

Proof. Take any $\delta_0 > 0$. For each $\delta > \delta_0$, note that

$$\begin{aligned}\partial \bar{\mathbb{R}}^{>\delta}(\boldsymbol{j}) \subseteq\ & \underbrace{\left\{\boldsymbol{x} \in \mathbb{R}_+^d \setminus \bar{\mathbb{R}}^d_{\lneq}(\boldsymbol{j}, \delta_0) :\ \|\boldsymbol{x}\| = \delta\right\}}_{\stackrel{\text{def}}{=} E_1(\delta)} \cup \underbrace{\left\{\boldsymbol{x} \in \mathbb{R}_+^d \setminus \bar{\mathbb{R}}^d_{\lneq}(\boldsymbol{j}, \delta_0) :\ \boldsymbol{x} \in \partial \bar{\mathbb{R}}^d(\boldsymbol{j}, \delta),\ \|\boldsymbol{x}\| > \delta_0\right\}}_{\stackrel{\text{def}}{=} E_2(\delta)} \\ & \cup \underbrace{\left\{\boldsymbol{x} \in \mathbb{R}_+^d \setminus \bar{\mathbb{R}}^d_{\lneq}(\boldsymbol{j}, \delta_0) :\ \boldsymbol{x} \in \partial \bar{\mathbb{R}}^d_{\lneq}(\boldsymbol{j}, \delta),\ \|\boldsymbol{x}\| > \delta_0\right\}}_{\stackrel{\text{def}}{=} E_3(\delta)}.\end{aligned}$$

We first analyze the set $E_1(\delta)$. Note that $E_1(\delta) \cap E_1(\delta') = \emptyset$ for any $\delta_0 < \delta < \delta'$, and $\bigcup_{\delta > \delta_0} E_1(\delta) \subseteq \mathbb{R}_+^d \setminus \bar{\mathbb{R}}^d_{\lneq}(\boldsymbol{j}, \delta_0)$. Meanwhile, under the $\mathcal{MHRV}$ condition in Assumption 5, we have

$$\mathbf{C}_{\boldsymbol{j}}\left(\left(\bar{\mathbb{R}}^d_{\lneq}(\boldsymbol{j}, \delta_0)\right)^c\right) < \infty;$$

see Definition 2.4. As a result, given any $\epsilon > 0$, there exists at most finitely many $\delta \in (\delta_0, \infty)$ such that $\mathbf{C}_{\boldsymbol{j}}\big(E_1(\delta)\big) > \epsilon$. Sending $\epsilon \downarrow 0$, we confirm that $\mathbf{C}_{\boldsymbol{j}}\big(E_1(\delta)\big) = 0$ for all but countably many $\delta \in (\delta_0, \infty)$.

Similarly, note that $\bigcup_{\delta > \delta_0} E_2(\delta) \subseteq \mathbb{R}^d_+ \setminus \bar{\mathbb{R}}^d_{\leqslant}(\boldsymbol{j}, \delta_0)$, and that $E_2(\delta) \cap E_2(\delta') = \emptyset$ for any $\delta' > \delta > \delta_0$. The same arguments above confirm that $\mathbf{C}_{\boldsymbol{j}}\big(E_2(\delta)\big) = 0$ for all but countably many $\delta \in (\delta_0, \infty)$. Next, by the definition of $\bar{\mathbb{R}}^d_{\leqslant}(\boldsymbol{j}, \delta)$, we have $E_3(\delta) \subseteq \bigcup_{\boldsymbol{j}' \in \mathcal{P}_d:\, \boldsymbol{j}' \neq \boldsymbol{j},\ \alpha(\boldsymbol{j}') \leq \alpha(\boldsymbol{j})} E_3^{\boldsymbol{j}'}(\delta)$, where

$$E_3^{\boldsymbol{j}'}(\delta) \stackrel{\text{def}}{=} \big\{\boldsymbol{x} \in \mathbb{R}^d_+ \setminus \bar{\mathbb{R}}^d_{\leqslant}(\boldsymbol{j}, \delta_0) :\ \boldsymbol{x} \in \partial \bar{\mathbb{R}}^d(\boldsymbol{j}', \delta),\ \|\boldsymbol{x}\| > \delta_0\big\}.$$

For each $\boldsymbol{j}' \in \mathcal{P}_d$ with $\boldsymbol{j}' \neq \boldsymbol{j},\ \alpha(\boldsymbol{j}') \leq \alpha(\boldsymbol{j})$, one can see that $\bigcup_{\delta \in (\delta_0, \infty)} E_3^{\boldsymbol{j}'}(\delta) \subseteq \mathbb{R}^d_+ \setminus \bar{\mathbb{R}}^d_{\leqslant}(\boldsymbol{j}, \delta_0)$, and $E_3^{\boldsymbol{j}'}(\delta) \cap E_3^{\boldsymbol{j}'}(\delta') = \emptyset$ for any $\delta' > \delta > \delta_0$. Again, we get

$$\mathbf{C}_{\boldsymbol{j}}\big(E_3(\delta)\big) \leq \sum_{\boldsymbol{j}' \in \mathcal{P}_d:\, \boldsymbol{j}' \neq \boldsymbol{j},\ \alpha(\boldsymbol{j}') \leq \alpha(\boldsymbol{j})} \mathbf{C}_{\boldsymbol{j}}\big(E_3^{\boldsymbol{j}'}(\delta)\big) = 0$$

for all but countably many $\delta \in (\delta_0, \infty)$.

In summary, we have verified that the claim $\mathbf{C}_{\boldsymbol{j}}\big(\partial \bar{\mathbb{R}}^{>\delta}(\boldsymbol{j})\big) = 0$ holds for all but countably many $\delta \in (\delta_0, \infty)$. We conclude the proof by sending $\delta_0 \downarrow 0$. □

To proceed, given $\delta > 0$, $\boldsymbol{j} \in \mathcal{P}_d$, and $n \in \mathbb{N}$, let (for each $k \geq 1$)

$$\tau_n^{>\delta}(k; \boldsymbol{j}) \stackrel{\text{def}}{=} \inf\Big\{t > \tau_n^{>\delta}(k-1; \boldsymbol{j}) :\ \Delta \bar{\boldsymbol{L}}_n(t) \in \bar{\mathbb{R}}^{>\delta}(\boldsymbol{j})\Big\}, \tag{37}$$

$$\boldsymbol{W}_n^{>\delta}(k; \boldsymbol{j}) \stackrel{\text{def}}{=} \Delta \bar{\boldsymbol{L}}_n\Big(\tau_n^{>\delta}(k; \boldsymbol{j})\Big), \tag{38}$$

and we adopt the convention $\tau_n^{>\delta}(0; \boldsymbol{j}) = 0$. Analogous to the definitions in (31)–(32), the $\tau_n^{>\delta}(k; \boldsymbol{j})$'s and $\boldsymbol{W}_n^{>\delta}(k; \boldsymbol{j})$'s are the arrival times and sizes of the k^{th} large jump that belongs to $\bar{\mathbb{R}}^{>\delta}(\boldsymbol{j})$. To provide the individual count for different types of large jumps, we define

$$K_n^{>\delta}(\boldsymbol{j}) \stackrel{\text{def}}{=} \max\{k \geq 0 :\ \tau_n^{>\delta}(k; \boldsymbol{j}) \leq 1\}, \qquad \forall \boldsymbol{j} \in \mathcal{P}_d, \tag{39}$$

and keep track of all counts by defining

$$\boldsymbol{K}_n^{>\delta} \stackrel{\text{def}}{=} \Big(K_n^{>\delta}(\boldsymbol{j})\Big)_{\boldsymbol{j} \in \mathcal{P}_d} \in \mathbb{Z}_+^{\mathcal{P}_d}. \tag{40}$$

Similarly, we define

$$\widetilde{\tau}_n^{>\delta}(k) \stackrel{\text{def}}{=} \inf\big\{t > \tau_n^{>\delta}(k-1) :\ \Delta \bar{\boldsymbol{L}}_n(t) \notin \bar{\mathbb{R}}^d([d], \delta)\big\}, \qquad \tilde{\tau}_n^{>\delta}(0) = 0, \tag{41}$$

$$\widetilde{K}_n^{>\delta} \stackrel{\text{def}}{=} \max\{k \geq 0 :\ \widetilde{\tau}_n^{>\delta}(k) \leq 1\}, \tag{42}$$

for big jumps that are not aligned with any direction of the cone $\bar{\mathbb{R}}^d([d], \delta)$. Given $\boldsymbol{\mathcal{K}} = (\mathcal{K}_{\boldsymbol{j}})_{\boldsymbol{j} \in \mathcal{P}_d} \in \mathbb{Z}_+^{\mathcal{P}_d}$, note that on the event $\{\boldsymbol{K}_n^{>\delta} = \boldsymbol{\mathcal{K}}\} \cap \{\widetilde{K}_n^{>\delta} = 0\}$, the process $\hat{\boldsymbol{L}}_n^{>\delta}(t)$ admits the form

$$\hat{\boldsymbol{L}}_n^{>\delta}(t) = \sum_{\boldsymbol{j} \in \mathcal{P}_d} \sum_{k \in [\mathcal{K}_{\boldsymbol{j}}]} \boldsymbol{W}_n^{>\delta}(k; \boldsymbol{j}) \mathbb{I}_{[\tau_n^{>\delta}(k; \boldsymbol{j}), 1]}(t), \qquad \forall t \in [0, 1] \tag{43}$$

with the $\tau_n^{>\delta}(k; \boldsymbol{j})$'s and $\boldsymbol{W}_n^{>\delta}(k; \boldsymbol{j})$'s defined in (37)–(38). In particular, on the event $\{\boldsymbol{K}_n^{>\delta} = \boldsymbol{\mathcal{K}}\} \cap \{\widetilde{K}_n^{>\delta} = 0\}$, we have $\hat{\boldsymbol{L}}_n^{>\delta} \in \mathbb{D}^{\delta}_{\boldsymbol{\mathcal{K}}} = \mathbb{D}^{\delta}_{\boldsymbol{\mathcal{K}};\boldsymbol{0}}[0, 1],$; see (27).

We are now ready to state Propositions C.2 and C.3. Specifically, let

$$\bar{\boldsymbol{K}}_n^{>\delta} \stackrel{\text{def}}{=} \big(\boldsymbol{K}_n^{>\delta}, \widetilde{K}_n^{\delta}\big) \tag{44}$$

be the concatenation of $\boldsymbol{K}_n^{>\delta}$ and $\widetilde{K}_n^{>\delta}$ defined in (39)–(42). Let

$$\bar{\mathbb{A}}(\boldsymbol{k}) \stackrel{\text{def}}{=} \big\{(\mathcal{K}, 0) : \ \mathcal{K} \in \mathbb{A}(\boldsymbol{k})\big\}, \qquad \forall \boldsymbol{k} \in \mathbb{Z}_+^d, \tag{45}$$

where $\mathbb{A}(\boldsymbol{k})$ is introduced in Definition 3.1. That is, we augment each allocation $\mathcal{K} \in \mathbb{A}(\boldsymbol{k})$ with the same constant 0.

Proposition C.2. *Let Assumptions 5 and 1 hold. For each $\boldsymbol{k} = (k_j)_{j\in[d]} \in \mathbb{Z}_+^d \setminus \{\boldsymbol{0}\}$ and $\Delta > 0$,*

$$\lim_{n\to\infty} \mathbf{P}\Big(\boldsymbol{d}_{J_1}\big(\hat{\boldsymbol{L}}_n^{>\delta}, \bar{\boldsymbol{L}}_n\big) > \Delta\Big) \Big/ \breve{\lambda}_{\boldsymbol{k}}(n) = 0, \qquad \forall \delta > 0 \text{ small enough,} \tag{46}$$

where $\breve{\lambda}_{\boldsymbol{k}}(n)$ is defined in (32).

Proposition C.3. *Let Assumptions 5 and 1 hold. Let $f : \mathbb{D} \to [0,\infty)$ be bounded (i.e., $\|f\| \stackrel{def}{=} \sup_{\xi\in\mathbb{D}} |f(\xi)| < \infty$) and continuous (w.r.t. the J_1 topology of $\mathbb{D}$). Let $\boldsymbol{k} = (k_j)_{j\in[d]} \in \mathbb{Z}_+^d \setminus \{\boldsymbol{0}\}$ and $\epsilon > 0$. Suppose that $B = supp(f)$ is bounded away from $\breve{\mathbb{D}}_{\leqslant\boldsymbol{k}}^{\epsilon}$ under $\boldsymbol{d}_{J_1}$. Then,*

(a) *for any $\delta \in (0, \epsilon)$,*

$$\lim_{n\to\infty} \frac{\mathbf{E}\Big[f(\hat{\boldsymbol{L}}_n^{>\delta})\mathbb{I}\big\{\bar{\boldsymbol{K}}_n^{>\delta} \notin \bar{\mathbb{A}}(\boldsymbol{k})\big\}\Big]}{\breve{\lambda}_{\boldsymbol{k}}(n)} = 0;$$

(b) *there exists $\delta_0 > 0$ such that for all but countably many $\delta \in (0, \delta_0)$,*

$$\lim_{n\to\infty} \frac{\mathbf{E}\Big[f(\hat{\boldsymbol{L}}_n^{>\delta})\mathbb{I}\big\{\bar{\boldsymbol{K}}_n^{>\delta} = (\mathcal{K}, 0)\big\}\Big]}{\breve{\lambda}_{\boldsymbol{k}}(n)} = \breve{\mathbf{C}}_{\mathcal{K}}(f) < \infty, \qquad \forall \mathcal{K} \in \mathbb{A}(\boldsymbol{k}),$$

where the measure $\breve{\mathbf{C}}_{\mathcal{K}}(\cdot) = \breve{\mathbf{C}}_{\mathcal{K};\boldsymbol{0}}^{[0,1]}(\cdot)$ is defined in (30).

We defer their proofs to Section C.2, and conclude this subsection by applying Propositions C.2 and C.3 and establishing Theorem 3.2.

Proof of Theorem 3.2. Without loss of generality, we prove Theorem 3.2 under Assumption 1. For any $\epsilon > 0$ and any Borel set $B \subseteq \mathbb{D}$ that is bounded away from $\breve{\mathbb{D}}_{\leqslant\boldsymbol{k}}^{\epsilon}$, note that B^{Δ} is also bounded away from $\breve{\mathbb{D}}_{\leqslant\boldsymbol{k}}^{\epsilon}$ under any $\Delta > 0$ sufficiently small. By Urysohn's Lemma, one can identify some bounded and continuous $f : \mathbb{D} \to [0,1]$ such that $\mathbb{I}_B \le f \le \mathbb{I}_{B^{\Delta}}$. Then, part (b) of Proposition C.3 confirms that $\sum_{\mathcal{K}\in\mathbb{A}(\boldsymbol{k})} \breve{\mathbf{C}}_{\mathcal{K}}(B) \le \sum_{\mathcal{K}\in\mathbb{A}(\boldsymbol{k})} \breve{\mathbf{C}}_{\mathcal{K}}(f) < \infty$. This establishes the finite upper bound in Claim (39) and verifies that $\sum_{\mathcal{K}\in\mathbb{A}(\boldsymbol{k})} \breve{\mathbf{C}}_{\mathcal{K}} \in \mathbb{M}\big(\mathbb{D} \setminus \breve{\mathbb{D}}_{\leqslant\boldsymbol{k}}^{\epsilon}\big)$, thus allowing us to apply Lemma 2.3. Next, given $\epsilon > 0$ and under the choice of

$$(\mathbb{S}, \boldsymbol{d}) = (\mathbb{D}, \boldsymbol{d}_{J_1}), \quad \mathbb{C} = \breve{\mathbb{D}}_{\leqslant\boldsymbol{k}}^{\epsilon}, \quad \epsilon_n = \breve{\lambda}_{\boldsymbol{k}}(n), \quad X_n = \bar{\boldsymbol{L}}_n, \quad Y_n^{\delta} = \hat{\boldsymbol{L}}_n^{>\delta}, \quad V_n^{\delta} = \bar{\boldsymbol{K}}_n^{>\delta}, \quad \mathcal{V} = \bar{\mathbb{A}}(\boldsymbol{k}), \tag{47}$$

Condition (i) of Lemma 2.3, is verified by Proposition C.2, and Condition (ii) is verified by Proposition C.3. In particular, in Condition (i) of Lemma 2.3, note that

$$\mathbf{P}\big(\boldsymbol{d}(X_n, Y_n^\delta)\mathbb{I}(X_n \in B \text{ or } Y_n^\delta \in B) > \Delta\big) \le \mathbf{P}\big(\boldsymbol{d}(X_n, Y_n^\delta) > \Delta\big),$$

and the claim $\mathbf{P}\big(\boldsymbol{d}(X_n, Y_n^\delta) > \Delta\big) = o(\epsilon_n)$ as $n \to \infty$, under the choices in (47) with $\delta > 0$ small enough, is exactly the content of Proposition C.2. As for Condition (ii), note that the claims are trivial when $B = \emptyset$. In case that $B \neq \emptyset$, since B is bounded away from $\check{\mathbb{D}}^\epsilon_{\leqslant \boldsymbol{k}}$, we pick $\Delta > 0$ small enough B^Δ is still bounded away from $\mathbb{C}$. By Urysohn's Lemma, one can find some bounded and continuous $f : \mathbb{D} \to [0,1]$ such that $\mathbb{I}_B \le f \le \mathbb{I}_{B^\Delta}$. Then, by applying part (b) of Proposition C.3 onto such f, one can identify some $\delta_0 = \delta_0(B, \Delta) > 0$ such that for all but countably many $\delta \in (0, \delta_0)$,

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{\mathbf{P}\big(\hat{\boldsymbol{L}}_n^{>\delta} \in B,\ \bar{\boldsymbol{K}}_n^{>\delta} = (\mathcal{K}, 0)\big)}{\check{\lambda}_{\boldsymbol{k}}(n)} \\
&\le \lim_{n\to\infty} \frac{\mathbf{E}\big[f(\hat{\boldsymbol{L}}_n^{>\delta})\mathbb{I}\{\bar{\boldsymbol{K}}_n^{>\delta} = (\mathcal{K}, 0)\}\big]}{\check{\lambda}_{\boldsymbol{k}}(n)} = \check{\mathbf{C}}_{\mathcal{K}}(f) \le \check{\mathbf{C}}_{\mathcal{K}}(B^\Delta), \qquad \forall \mathcal{K} \in \mathbb{A}(\boldsymbol{k}).
\end{aligned}$$

By a similar argument using Urysohn's Lemma, one can identify some bounded and continuous f with $\mathbb{I}_{B_\Delta} \le f \le \mathbb{I}_B$. Using part (b) of Proposition C.3 again (and by picking a smaller $\delta_0 > 0$ if necessary), it follows for all but countably many $\delta \in (0, \delta_0)$ that

$$\liminf_{n\to\infty} \frac{\mathbf{P}\big(\hat{\boldsymbol{L}}_n^{>\delta} \in B, \bar{\boldsymbol{K}}_n^{>\delta} = (\mathcal{K}, 0)\big)}{\check{\lambda}_{\boldsymbol{k}}(n)} \ge \check{\mathbf{C}}_{\mathcal{K}}(B_\Delta), \qquad \forall \mathcal{K} \in \mathbb{A}(\boldsymbol{k}).$$

Likewise, by part (a) of of Proposition C.3, for the continuous and bounded f with $\mathbb{I}_B \le f \le \mathbb{I}_{B^\Delta}$ identified above, it holds for any $\delta > 0$ small enough that

$$\limsup_{n\to\infty} \frac{\mathbf{P}\big(\hat{\boldsymbol{L}}_n^{>\delta} \in B,\ \bar{\boldsymbol{K}}_n^{>\delta} \notin \bar{\mathbb{A}}(\boldsymbol{k})\big)}{\check{\lambda}_{\boldsymbol{k}}(n)} \le \lim_{n\to\infty} \frac{\mathbf{E}\big[f(\hat{\boldsymbol{L}}_n^{>\delta})\mathbb{I}\{\bar{\boldsymbol{K}}_n^{>\delta} \notin \bar{\mathbb{A}}(\boldsymbol{k})\}\big]}{\check{\lambda}_{\boldsymbol{k}}(n)} = 0.$$

These calculations verify Condition (ii) of Lemma 2.3 under the choices in (47). Using Lemma 2.3, we conclude the proof of Theorem 3.2. □

C.2. Proofs of Propositions C.2 and C.3

This subsection collects the proofs of Propositions C.2 and C.3. To this end, we prepare a few technical lemmas. First, Lemma C.4 establishes a concentration inequality for $\bar{\boldsymbol{L}}_n(t)$ before the arrival of any large jump.

Lemma C.4. *Let Assumption 5 hold. Given any $\epsilon, \beta, T > 0$, there exists $\delta_0 = \delta_0(\epsilon, \beta)$ such that*

$$\lim_{n\to\infty} n^{-\beta} \cdot \mathbf{P}\left(\sup_{t\in[0,T]:\ t<\tau_n^{>\delta}(1)} \left\| \bar{\boldsymbol{L}}_n(t) - \boldsymbol{\mu}_{\boldsymbol{L}} t \right\| > \epsilon \right) = 0 \qquad \forall \delta \in (0, \delta_0).$$

Proof. Without loss of generality, we work with the assumption that $\mu_{\boldsymbol{L}} = \mathbf{0}$ and $T = 1$. Recall the Lévy–Itô decomposition stated in (21):

$$\boldsymbol{L}(t) \overset{\mathcal{D}}{=} \underbrace{\boldsymbol{\Sigma}_{\boldsymbol{L}}^{1/2}\boldsymbol{B}(t)}_{\overset{\text{def}}{=}\boldsymbol{L}_1(t)} + \underbrace{\int_{\|\boldsymbol{x}\|\le 1} \boldsymbol{x}\big[\mathrm{PRM}_\nu([0,t]\times d\boldsymbol{x}) - t\nu(d\boldsymbol{x})\big]}_{\overset{\text{def}}{=}\boldsymbol{L}_2(t)} + \underbrace{\boldsymbol{c}_{\boldsymbol{L}}t + \int_{\|\boldsymbol{x}\|>1} \boldsymbol{x}\mathrm{PRM}_\nu([0,t]\times d\boldsymbol{x})}_{\overset{\text{def}}{=}\boldsymbol{L}_3(t)},$$

where ν is the Lévy measure supported on $\mathbb{R}_+^d \setminus \{\mathbf{0}\}$ satisfying $\int_{\boldsymbol{x}\in\mathbb{R}_+^d\setminus\{\mathbf{0}\}}(\|\boldsymbol{x}\|^2 \wedge 1)\nu(d\boldsymbol{x}) < \infty$. By definitions in (31), $\tau_n^{>\delta}(1)$ is the arrival time of the first discontinuity in $\bar{\boldsymbol{L}}_n(t)$ with norm larger than δ. Therefore, after the $1/n$ time-scaling, $n\tau_n^{>\delta}(1)$ is the arrival time of the first discontinuity in $\boldsymbol{L}(t)$ larger than $n\delta$. Then, it suffices to show that (as $n \to \infty$)

$$\mathbf{P}\Big(\sup_{t\in[0,n]} \|\boldsymbol{L}_1(t)\| > n\epsilon/3\Big) = o(n^{-\beta}), \tag{48}$$

$$\mathbf{P}\Big(\sup_{t\in[0,n]} \|\boldsymbol{L}_2(t)\| > n\epsilon/3\Big) = o(n^{-\beta}), \tag{49}$$

$$\mathbf{P}\Big(\sup_{t\in[0,n]:\, t<n\tau_n^{>\delta}(1)} \|\boldsymbol{L}_3(t)\| > n\epsilon/3\Big) = o(n^{-\beta}), \qquad \forall \delta > 0 \text{ small enough.} \tag{50}$$

Proof of Claim (48). We use $\sigma_{i,j}^{1/2}$ to denote the element in the i^{th} row and j^{th} column of the positive semi-definite matrix $\boldsymbol{\Sigma}_{\boldsymbol{L}}^{1/2} \in \mathbb{R}^{d\times d}$, and write $\boldsymbol{B}(t) = \big(B_1(t), \ldots, B_d(t)\big)^{\mathrm{T}}$. Let $C \overset{\text{def}}{=} \max_{i,j\in[d]} |\sigma_{i,j}^{1/2}|$. Suppose that $\sup_{t\in[0,n]} |B_j(t)| \le \frac{n\epsilon}{3Cd^2}$ holds for each $j \in [d]$. Then, under the L_1 norm,

$$\sup_{t\in[0,n]} \|\boldsymbol{L}_1(t)\| \le \sum_{i\in[d]}\sum_{j\in[d]} |\sigma_{i,j}^{1/2}| \cdot \sup_{t\in[0,n]} |B_j(t)| \le \sum_{j\in[d]} \sup_{t\in[0,n]} |B_j(t)| \cdot Cd \le d \cdot \frac{n\epsilon}{3Cd^2} \cdot Cd \le \epsilon/3.$$

Therefore, it suffices to show that $\mathbf{P}\big(\sup_{t\in[0,n]} |B(t)| > \frac{n\epsilon}{3Cd^2}\big) = o(n^{-\beta})$ holds for a standard Brownian motion $B(t)$ in $\mathbb{R}^1$. By Doob's maximal inequality and the MGF of $B(t)$, we get

$$\mathbf{P}\bigg(\sup_{t\in[0,n]} |B(t)| > \frac{n\epsilon}{3Cd^2}\bigg) \le 2\cdot\exp\bigg[-\frac{1}{2n}\cdot\bigg(\frac{n\epsilon}{3Cd^2}\bigg)^2\bigg] = 2\cdot\exp\bigg(-n\cdot\frac{\epsilon^2}{18C^2d^4}\bigg) = o(n^{-\beta}),$$

and conclude the proof of Claim (48).

Proof of Claim (49). Note that each coordinate of $\boldsymbol{L}_2(t) = (L_{2,1}(t), \ldots, L_{2,d}(t))^\top$ is a Lévy process with bounded jumps and also a martingale. Therefore, each $L_{2,j}(t)$ has finite moments of any order; see Theorem 34 in Chapter I of [7]. Meanwhile, by Burkholder-Davis-Gundy inequalities (see Theorem 48 in Chapter IV of [7]), for each $p \in [1,\infty)$, there exists some constant $c_p \in (0,\infty)$ (whose value does not vary with the law of $\boldsymbol{L}$) such that

$$\mathbf{E}\bigg[\sup_{t\in[0,n]} |L_{2,j}(t)|^p\bigg] \le c_p\mathbf{E}\bigg[\Big([L_{2,j},L_{2,j}](n)\Big)^{p/2}\bigg], \qquad \forall j\in[d],\ n\ge 1,$$

where $V_j(t) \overset{\text{def}}{=} [L_{2,j}, L_{2,j}](t)$ is the quadratic variation process of $L_{2,j}(t)$. Due to the independent and stationary increments of Lévy processes $L_{2,j}(t)$, the quadratic variation $V_j(t) = [L_{2,j}, L_{2,j}](t)$ also

has stationary and independent increments. Then, by the Minkowski inequality, for each $p \in [1,\infty)$ we have

$$\left(\mathbf{E}\Big[\sup_{t\in[0,n]}|L_{2,j}(t)|^p\Big]\right)^{2/p} \le (c_p)^{2/p}\cdot\left(\mathbf{E}\Big[\Big([L_{2,j},L_{2,j}](n)\Big)^{p/2}\Big]\right)^{2/p}$$
$$\le (c_p)^{2/p}\cdot n\cdot\Big(\underbrace{\mathbf{E}\big[(V_j(1))^{p/2}\big]}_{\stackrel{\text{def}}{=}v_{p,j}<\infty}\Big)^{2/p}.$$

As a result, for each $j\in[d]$, $p\in[1,\infty)$, and $n\ge 1$,

$$\mathbf{P}\Big(\sup_{t\in[0,n]}|L_{2,j}(t)|>\frac{n\epsilon}{3d}\Big)\le\Big(\frac{3d}{\epsilon}\Big)^p\cdot\frac{\mathbf{E}\big[\sup_{t\in[0,n]}|L_{2,j}(t)|^p\big]}{n^p}$$
$$\le\Big(\frac{3d}{\epsilon}\Big)^p\cdot\frac{n^{p/2}\cdot c_p v_{p,j}}{n^p}=\Big(\frac{3d}{\epsilon}\Big)^p\cdot\frac{c_p v_{p,j}}{n^{p/2}}.$$

Picking $p>2\beta$, we conclude the proof of Claim (49).

Proof of Claim (50). Note that $\boldsymbol{L}_3(t)$ is a compound Poisson process with drift, i.e., $\boldsymbol{L}_3(t)=\boldsymbol{c}_{\boldsymbol{L}}t+\sum_{i=1}^{N(t)}\boldsymbol{Z}_n$, where $N(t)$ is a Poisson process with rate $\lambda=\nu\big(\{\|\boldsymbol{x}\|\in\mathbb{R}^d_+:\ \|\boldsymbol{x}\|>1\}\big)$, and the $\boldsymbol{Z}_n$'s are i.i.d. copies under the law

$$\mathbf{P}(\boldsymbol{Z}\in\cdot\,)=\frac{\nu\big(\,\cdot\,\cap\{\|\boldsymbol{x}\|\in\mathbb{R}^d_+:\ \|\boldsymbol{x}\|>1\}\big)}{\nu\{\|\boldsymbol{x}\|\in\mathbb{R}^d_+:\ \|\boldsymbol{x}\|>1\}}.$$

In particular, under our running assumption $\mu_{\boldsymbol{L}}=\boldsymbol{0}$, by (24) we have $\boldsymbol{c}_{\boldsymbol{L}}=-\lambda\mathbf{E}\boldsymbol{Z}$ for the linear drift, and hence

$$\boldsymbol{L}_3(t)=\sum_{i=1}^{N(t)}\boldsymbol{Z}_n-t\cdot\lambda\mathbf{E}\boldsymbol{Z}.$$

Also, by Assumption 5 and the $\mathcal{MHRV}$ condition in Definition 2.4, $\mathbf{P}(\|\boldsymbol{Z}\|>n)\in\mathcal{RV}_{-\alpha(\{i^*\})}(n)$ with $\alpha(\{i^*\})>1$, where $i^*\stackrel{\text{def}}{=}\arg\min_{i\in[d]}\alpha(\{i\})$. This confirms that $\|\mathbf{E}\boldsymbol{Z}\|<\infty$ and $\|\boldsymbol{Z}\|$ has regularly varying law with tail index larger than 1.

To proceed, let $j^*=j^*(n,\delta)$ be the index of the first $\boldsymbol{Z}_j$ with $\|\boldsymbol{Z}_j\|>n\delta$, and $\tau^*=\tau^*(n,\delta)$ be its arrival time in the compound Poisson process $\sum_{i=1}^{N(t)}\boldsymbol{Z}_n$. Given $\hat{\epsilon}>0$, on the event $\big\{\sup_{t\in[0,n]}|N(t)-\lambda t|\le n\hat{\epsilon}\big\}$, observe that

$$\sup_{t\in[0,n]:\ t<n\tau_n^{>\delta}(1)}\|\boldsymbol{L}_3(t)\|=\sup_{t\in[0,n]:\ t<\tau^*}\left\|\sum_{i=1}^{N(t)}\boldsymbol{Z}_n-t\cdot\lambda\mathbf{E}\boldsymbol{Z}\right\|$$
$$\le\sup_{t\in[0,n]:\ t<\tau^*}\left\|\sum_{i=1}^{N(t)}\boldsymbol{Z}_n-N(t)\cdot\mathbf{E}\boldsymbol{Z}\right\|+\sup_{t\in[0,n]:\ t<\tau^*}|N(t)-\lambda t|\cdot\|\mathbf{E}\boldsymbol{Z}\|$$
$$\le\max_{j\le(\lambda+\hat{\epsilon})\cdot n:\ j<j^*}\left\|\sum_{i=1}^{j}\boldsymbol{Z}_i-i\cdot\mathbf{E}\boldsymbol{Z}\right\|+n\hat{\epsilon}\cdot\|\mathbf{E}\boldsymbol{Z}\|$$

$$\leq \max_{j \leq (\lambda+\hat{\epsilon})\cdot n} \left\| \sum_{i=1}^{j} \mathbf{Z}_i \mathbb{I}\{\|\mathbf{Z}_i\| \leq n\delta\} - i \cdot \mathbf{E}\mathbf{Z} \right\| + n\hat{\epsilon} \cdot \|\mathbf{E}\mathbf{Z}\|.$$

In particular, by picking $\hat{\epsilon} > 0$ small enough, we get $\hat{\epsilon} \cdot \|\mathbf{E}\mathbf{Z}\| < \epsilon/6$. Therefore, to prove Claim (50), it suffices to show that

$$\mathbf{P}\left(\sup_{t \in [0,n]} |N(t) - \lambda t| > n\hat{\epsilon} \right) = o(n^{-\beta}), \tag{51}$$

$$\mathbf{P}\left(\max_{j \leq (\lambda+\hat{\epsilon})\cdot n} \left\| \sum_{i=1}^{j} \mathbf{Z}_i \mathbb{I}\{\|\mathbf{Z}_i\| \leq n\delta\} - i \cdot \mathbf{E}\mathbf{Z} \right\| > n\epsilon/6 \right) = o(n^{-\beta}), \quad \forall \delta > 0 \text{ small enough.} \tag{52}$$

However, the bound (51) follows from Doob's maximal inequality and Cramer's theorem, and the bound (52) follows from Lemma 3.1 in [9] (i.e., concentration inequalities for truncated regularly varying vectors). This concludes the proof of Claim (50). □

The next result justifies the use of $\hat{\boldsymbol{L}}_n^{>\delta}$ in (33) as an approximator to $\bar{\boldsymbol{L}}_n$.

Lemma C.5. *Let $k \in \mathbb{N}$ and $\epsilon > 0$. For any $n \geq 1$ and $\delta > 0$, it holds on event*

$$\left\{\tau_n^{>\delta}(k) \leq 1 < \tau_n^{>\delta}(k+1)\right\} \cap \left(\underbrace{\bigcap_{m=1}^{k+1} \left\{ \sup_{t \in [\tau_n^{>\delta}(m-1),1]:\, t < \tau_n^{>\delta}(m)} \left\| \bar{\boldsymbol{L}}_n(t) - \bar{\boldsymbol{L}}_n\left(\tau_n^{>\delta}(m-1)\right) \right\| \leq \frac{\epsilon}{k+1} \right\}}_{\stackrel{def}{=} A_n(m;\delta,k,\epsilon)} \right)$$

that

$$\sup_{t \in [0,1]} \left\| \hat{\boldsymbol{L}}_n^{>\delta}(t) - \bar{\boldsymbol{L}}_n(t) \right\| \leq \epsilon.$$

Proof. For any $\xi, \tilde{\xi} \in \mathbb{D}$ and any $0 \leq u < v \leq 1$, observe the elementary bound

$$\begin{aligned} &\sup_{t \in [0,v]} \left\| \xi(t) - \tilde{\xi}(t) \right\| \\ &\leq \sup_{t \in [0,u]} \left\| \xi(t) - \tilde{\xi}(t) \right\| + \sup_{t \in [u,v)} \left\| \left(\xi(t) - \xi(u)\right) - \left(\tilde{\xi}(t) - \tilde{\xi}(u)\right) \right\| + \left\| \Delta\xi(v) - \Delta\tilde{\xi}(v) \right\|. \end{aligned}$$

Furthermore, applying this bound inductively, it holds for any $0 < t_1 < t_2 < \ldots < t_k \leq 1$ that (under the convention that $t_0 = 0$ and $t_{k+1} = 1$)

$$\begin{aligned} &\sup_{t \in [0,1]} \left\| \xi(t) - \tilde{\xi}(t) \right\| \\ &\leq \sum_{m=1}^{k+1} \sup_{t \in [t_{m-1}, t_m)} \left\| \left(\xi(t) - \xi(t_{m-1})\right) - \left(\tilde{\xi}(t) - \tilde{\xi}(t_{m-1})\right) \right\| + \left\| \Delta\xi(t_m) - \Delta\tilde{\xi}(t_m) \right\|. \end{aligned}$$

Now, on event $\{\tau_n^{>\delta}(k) \le 1 < \tau_n^{>\delta}(k+1)\} \cap \big(\bigcap_{m=1}^{k+1} A_n(m;\delta,k,\epsilon)\big)$, we apply this bound with $\xi = \bar{\boldsymbol{L}}_n$, $\tilde{\xi} = \hat{\boldsymbol{L}}_n^{>\delta}$, and $t_m = \tau_n^{>\delta}(m)$. In particular, by (33), we have $\sup_{t\in[t_{m-1},t_m)} \big\|\tilde{\xi}(t) - \tilde{\xi}(t_{m-1})\big\| = 0$ and $\big\|\Delta\xi(t_m) - \Delta\tilde{\xi}(t_m)\big\| = 0$. The claims then follow directly from the condition in the event $A_n(m;\delta,k,\epsilon)$. □

In Lemma C.6, we develop useful asymptotics for $\tau_n^{>\delta}(k;\boldsymbol{j})$ and $\boldsymbol{W}_n^{>\delta}(k;\boldsymbol{j})$. Specifically, given $\delta > 0$, $n \ge 1$, we define the event

$$\widetilde{E}_n^{>\delta} \stackrel{\text{def}}{=} \big\{\widetilde{K}_n^{>\delta} = 0\big\}. \tag{53}$$

Furthermore, given $\mathcal{K} = (\mathcal{K}_{\boldsymbol{j}})_{\boldsymbol{j}\in\mathcal{P}_d} \in \mathbb{Z}_+^{\mathcal{P}_d}$, we define

$$E_n^{>\delta}(\mathcal{K}) \stackrel{\text{def}}{=} \big\{\boldsymbol{K}_n^{>\delta} = \mathcal{K}\big\} \cap \widetilde{E}_n^{>\delta} = \Bigg(\bigcap_{\boldsymbol{j}\in\mathcal{P}_d} \underbrace{\big\{K_n^{>\delta}(\boldsymbol{j}) = \mathcal{K}_{\boldsymbol{j}}\big\}}_{\stackrel{\text{def}}{=} E_{n;\boldsymbol{j}}^{>\delta}(\mathcal{K}_{\boldsymbol{j}})}\Bigg) \cap \big\{\widetilde{K}_n^{>\delta} = 0\big\}. \tag{54}$$

Given $\boldsymbol{j} \in \mathcal{P}_d$ and $c > 0$, let $\big(\boldsymbol{W}_*^{(c)}(k;\boldsymbol{j})\big)_{k\ge1}$ be sequences that are independent across $\boldsymbol{j} \in \mathcal{P}_d$, where each $\boldsymbol{W}_*^{(c)}(k;\boldsymbol{j})$ is an i.i.d. copy of $\boldsymbol{W}_*^{(c)}(\boldsymbol{j})$ with law

$$\mathbf{P}\Big(\boldsymbol{W}_*^{(c)}(\boldsymbol{j}) \in \cdot\Big) \stackrel{\text{def}}{=} \mathbf{C}_{\boldsymbol{j}}\Big(\cdot \cap \bar{\mathbb{R}}^{>c}(\boldsymbol{j})\Big)\Big/\mathbf{C}_{\boldsymbol{j}}\Big(\bar{\mathbb{R}}^{>c}(\boldsymbol{j})\Big), \tag{55}$$

where the $\mathbf{C}_{\boldsymbol{j}}$'s are the limiting measures in the $\mathcal{MHRV}$ condition of Assumption 5, and the $\bar{\mathbb{R}}^{>c}(\boldsymbol{j})$'s are defined in (34). Besides, let the $U_{\boldsymbol{j},k}$'s be i.i.d. copies of Unif(0, 1), and, for each $k \ge 1$ and $\boldsymbol{j} \in \mathcal{P}_d$, let $U_{\boldsymbol{j},(1:k)} \le U_{\boldsymbol{j},(2:k)} \le \ldots \le U_{\boldsymbol{j},(k:k)}$ be the order statistics of $(U_{\boldsymbol{j},q})_{q\in[k]}$.

Lemma C.6. *Let Assumptions 5 and 1 hold. Let $\mathcal{K} = (\mathcal{K}_{\boldsymbol{j}})_{\boldsymbol{j}\in\mathcal{P}_d} \in \mathbb{Z}_+^{\mathcal{P}_d}$, and let $\boldsymbol{k} \in \mathbb{Z}_+^d$ satisfy (33) (i.e., $\mathcal{K}$ is an allocation of $\boldsymbol{k}$). First,*

$$\lim_{n\to\infty} n^{\gamma} \cdot \mathbf{P}\Big(\big(\widetilde{E}_n^{>\delta}\big)^c\Big) = 0, \qquad \forall \gamma > 0,\ \delta > 0. \tag{56}$$

Next, for any $\delta > 0$ if Claim (35) holds, then

$$\lim_{n\to\infty} \frac{\mathbf{P}\big(E_n^{>\delta}(\mathcal{K})\big)}{\breve{\lambda}_{\boldsymbol{k}}(n)} = \prod_{\boldsymbol{j}\in\mathcal{P}_d} \frac{1}{\mathcal{K}_{\boldsymbol{j}}!} \cdot \Big(\mathbf{C}_{\boldsymbol{j}}\big(\bar{\mathbb{R}}^{>\delta}(\boldsymbol{j})\big)\Big)^{\mathcal{K}_{\boldsymbol{j}}}, \tag{57}$$

where $\breve{\lambda}_{\boldsymbol{j}}(\cdot)$ is defined in (32). Furthermore, if $\mathcal{K} \ne \mathbf{0}$, then

$$\mathscr{L}\Bigg(\Big(\tau_n^{>\delta}(k;\boldsymbol{j}), \boldsymbol{W}_n^{>\delta}(k;\boldsymbol{j})\Big)_{\boldsymbol{j}\in\mathcal{P}_d,\ k\in[\mathcal{K}_{\boldsymbol{j}}]} \Bigg| E_n^{>\delta}(\mathcal{K})\Bigg) \to \mathscr{L}\Bigg(\Big(U_{\boldsymbol{j},(k:\mathcal{K}_{\boldsymbol{j}})}, \boldsymbol{W}_*^{(\delta)}(k;\boldsymbol{j})\Big)_{\boldsymbol{j}\in\mathcal{P}_d,\ k\in[\mathcal{K}_{\boldsymbol{j}}]}\Bigg) \tag{58}$$

as $n \to \infty$ in terms of weak convergence.

Proof. Recall the definition of $\nu_n(A) = \nu\{n\boldsymbol{x} : \boldsymbol{x} \in A\}$. First, by the law of the (scaled) Lévy process $\bar{\boldsymbol{L}}_n(t) = \frac{1}{n}\boldsymbol{L}(nt)$ in (21) and the definitions in (41)–(42),

$$\begin{aligned}\mathbf{P}\Big(\big(\widetilde{E}_n^{>\delta}\big)^c\Big) &= \mathbf{P}\left(\text{Poisson}\Big(n\cdot\nu_n\big(\mathbb{R}_+^d \setminus \bar{\mathbb{R}}^d([d],\delta)\big)\Big) \geq 1\right)\\ &\leq \mathbf{E}\left[\text{Poisson}\Big(n\cdot\nu_n\big(\mathbb{R}_+^d \setminus \bar{\mathbb{R}}^d([d],\delta)\big)\Big)\right] = n\cdot\nu_n\left(\Big(\bar{\mathbb{R}}^d([d],\delta)\Big)^c\right) \text{ by Markov's inequality.}\end{aligned}$$

By the $\mathcal{MHRV}$ condition in Assumption 5 (in particular, Claim (10) in Definition 2.4), we verify Claim (56).

Next, by the independence when splitting the Poisson random measure PRM_ν in (21),

$$\mathbf{P}\Big(E_n^{>\delta}(\mathcal{K})\Big) = \left(\prod_{\boldsymbol{j}\in\mathcal{P}_d}\mathbf{P}\Big(E_{n;\boldsymbol{j}}^{>\delta}(\mathcal{K}_{\boldsymbol{j}})\Big)\right)\cdot\mathbf{P}\big(\widetilde{E}_n^{>\delta}\big).$$

On the one hand, Claim (56) implies that $\lim_{n\to\infty}\mathbf{P}\big(\widetilde{E}_n^{>\delta}\big) = 1$. On the other hand, in light of property (35), to prove (57) it suffices to show that for any $\boldsymbol{j}\in\mathcal{P}_d$ and $k\geq 0$,

$$\lim_{n\to\infty}\frac{\mathbf{P}\big(E_{n;\boldsymbol{j}}^{>\delta}(k)\big)}{\big(n\lambda_{\boldsymbol{j}}(n)\big)^k} = \frac{1}{k!}\cdot\Big(\mathbf{C}_{\boldsymbol{j}}\big(\bar{\mathbb{R}}^{>\delta}(\boldsymbol{j})\big)\Big)^k. \tag{59}$$

By the law of the (scaled) Lévy process $\bar{\boldsymbol{L}}_n(t)$ in (21) and the definitions in (37)–(38),

$$\mathbf{P}\Big(E_{n;\boldsymbol{j}}^{>\delta}(k)\Big) = \mathbf{P}\left(\text{Poisson}\Big(n\cdot\underbrace{\nu_n\big(\bar{\mathbb{R}}^{>\delta}(\boldsymbol{j})\big)}_{\stackrel{\text{def}}{=}\theta_n(\boldsymbol{j},\delta)}\Big) = k\right) = \exp\big(-n\theta_n(\boldsymbol{j},\delta)\big)\cdot\frac{\big(n\theta_n(\boldsymbol{j},\delta)\big)^k}{k!}. \tag{60}$$

By the $\mathcal{MHRV}$ condition in Assumption 5,

$$\lim_{n\to\infty}\frac{\theta_n(\boldsymbol{j},\delta)}{\lambda_{\boldsymbol{j}}(n)} = \mathbf{C}_{\boldsymbol{j}}\Big(\bar{\mathbb{R}}^{>\delta}(\boldsymbol{j})\Big); \qquad \text{see (9) and our choice of } \delta \text{ in (35).}$$

This also implies $\theta_n(\boldsymbol{j},\delta) = O\big(\lambda_{\boldsymbol{j}}(n)\big) = o(n)$ (due to $\lambda_{\boldsymbol{j}}(n)\in\mathcal{RV}_{-\alpha(\boldsymbol{j})}(n)$ with $\alpha(\boldsymbol{j})>1$), and hence $\lim_{n\to\infty}\exp\big(-n\theta_n(\boldsymbol{j},\delta)\big) = 1$. Plugging these limits back into (60), we conclude the proof of Claim (59).

Next, we prove the weak convergence stated in (58). Again, by the independence of the Poisson splitting, it suffices to show that for any $\boldsymbol{j}\in\mathcal{P}_d$ and $K\geq 1$,

$$\mathscr{L}\left(\Big(\tau_n^{>\delta}(k;\boldsymbol{j}),\boldsymbol{W}_n^{>\delta}(k;\boldsymbol{j})\Big)_{k\in[K]}\,\middle|\,E_{n;\boldsymbol{j}}^{>\delta}(K)\right) \to \mathscr{L}\left(\Big(U_{\boldsymbol{j},(k:K)},\boldsymbol{W}_*^{(\delta)}(k;\boldsymbol{j})\Big)_{k\in[K]}\right). \tag{61}$$

Next, by the law of a compound Poisson process, the arrival times $\tau_n^{>\delta}(k;\boldsymbol{j})$ are independent of the jump sizes $\boldsymbol{W}_n^{>\delta}(k;\boldsymbol{j})$, which implies that the law of the jump size sequence $\big(\boldsymbol{W}_n^{>\delta}(k;\boldsymbol{j})\big)_{k\geq 1}$ is independent of the event $E_{n;\boldsymbol{j}}^{>\delta}(K)$. Furthermore, the conditional law of the $\tau_n^{>\delta}(k;\boldsymbol{j})$'s—the sequence of arrival times of jumps in a compound Poisson process conditioning on the number of jumps—admits

the form

$$\mathscr{L}\Big(\tau_n^{>\delta}(1;\boldsymbol{j}),\tau_n^{>\delta}(2;\boldsymbol{j}),\ldots,\tau_n^{>\delta}(K;\boldsymbol{j})\;\Big|\;E_{n;\boldsymbol{j}}^{>\delta}(K)\Big)=\mathscr{L}\Big(U_{\boldsymbol{j},(1:K)},U_{\boldsymbol{j},(2:K)},\ldots,U_{\boldsymbol{j},(K:K)}\Big).$$

Therefore, it only remains to study the marginal law for $\boldsymbol{W}_n^{>\delta}(1;\boldsymbol{j})$ (since all $\boldsymbol{W}_n^{>\delta}(k;\boldsymbol{j})$'s admit the same law) and show that

$$\mathbf{P}\Big(\boldsymbol{W}_n^{>\delta}(1;\boldsymbol{j})\in\,\cdot\,\Big)\to\mathbf{P}\Big(\boldsymbol{W}_*^{(\delta)}(\boldsymbol{j})\in\,\cdot\,\Big)$$

in terms of weak convergence, where the law of $\boldsymbol{W}_*^{(\delta)}(\boldsymbol{j})$ is defined in (55). To this end, note that by the definitions in (37)–(38),

$$\mathbf{P}\Big(\boldsymbol{W}_n^{>\delta}(1;\boldsymbol{j})\in\,\cdot\,\Big)=\frac{\nu_n\big(\bar{\mathbb{R}}^{>\delta}(\boldsymbol{j})\cap\,\cdot\,\big)}{\nu_n\big(\bar{\mathbb{R}}^{>\delta}(\boldsymbol{j})\big)},\qquad\forall\boldsymbol{j}\in\mathcal{P}_d.\tag{62}$$

Then, given a Borel set $A\subseteq\mathbb{R}_+^d$,

$$\begin{aligned}
&\limsup_{n\to\infty}\mathbf{P}\Big(\boldsymbol{W}_n^{>\delta}(1;\boldsymbol{j})\in A\Big)\\
&=\limsup_{n\to\infty}\frac{\nu_n\Big(A\cap\bar{\mathbb{R}}^{>\delta}(\boldsymbol{j})\Big)\Big/\lambda_{\boldsymbol{j}}(n)}{\nu_n\Big(\bar{\mathbb{R}}^{>\delta}(\boldsymbol{j})\Big)\Big/\lambda_{\boldsymbol{j}}(n)}\qquad\text{by (62)}\\
&\le\mathbf{C}_{\boldsymbol{j}}\Big(\Big(A\cap\bar{\mathbb{R}}^{>\delta}(\boldsymbol{j})\Big)^-\Big)\Big/\mathbf{C}_{\boldsymbol{j}}\Big(\Big(\bar{\mathbb{R}}^{>\delta}(\boldsymbol{j})\Big)^\circ\Big)\qquad\text{by the }\mathcal{MHRV}\text{ condition and (9)}\\
&\le\mathbf{C}_{\boldsymbol{j}}\Big(A^-\cap\bar{\mathbb{R}}^{>\delta}(\boldsymbol{j})\Big)\Big/\mathbf{C}_{\boldsymbol{j}}\Big(\bar{\mathbb{R}}^{>\delta}(\boldsymbol{j})\Big)\qquad\text{because of }(A\cap B)^-\subseteq A^-\cap B^-\text{ and (35)}\\
&=\mathbf{P}\Big(\boldsymbol{W}_*^{(\delta)}(\boldsymbol{j})\in A^-\Big);\qquad\text{see (55).}
\end{aligned}$$

Using the Portmanteau Theorem, we conclude the proof. □

Given $\mathcal{K}=(\mathcal{K}_{\boldsymbol{j}})_{\boldsymbol{j}\in\mathcal{P}_d}\in\mathbb{Z}_+^{\mathcal{P}_d}$ and $\epsilon\ge0$, recall that any path $\xi\in\breve{\mathbb{D}}_{\mathcal{K}}^{\epsilon}$ admits the form

$$\xi(t)=\sum_{\boldsymbol{j}\in\mathcal{P}_d}\sum_{k=1}^{\mathcal{K}_{\boldsymbol{j}}}\boldsymbol{w}_{\boldsymbol{j},k}\mathbb{I}_{[t_{\boldsymbol{j},k},1]}(t),\qquad\forall t\in[0,1],\tag{63}$$

where $\boldsymbol{w}_{\boldsymbol{j},k}\in\bar{\mathbb{R}}^d(\boldsymbol{j},\epsilon)$ and $t_{\boldsymbol{j},k}\in(0,1]$ for each $\boldsymbol{j}\in\mathcal{P}_d$ and $k\in[\mathcal{K}_{\boldsymbol{j}}]$, and any two elements in the sequence $(t_{\boldsymbol{j},k})_{\boldsymbol{j}\in\mathcal{P}_d,\ k\in[\mathcal{K}_{\boldsymbol{j}}]}$ will not coincide; see (27) and (28). Besides, recall that in Definition 3.1, we use $\mathbb{A}(\boldsymbol{k})$ to denote the set of all allocations of $\boldsymbol{k}\in\mathbb{Z}_+^d$. We prepare the next two lemmas to collect useful properties of sets $\breve{\mathbb{D}}_{\mathcal{K}}^{\epsilon}$ and measures $\breve{\mathbf{C}}_{\mathcal{K}}(\cdot)$ (see (30)).

Lemma C.7. *Let Assumption 5 hold. Let $\boldsymbol{k}=(k_j)_{j\in[d]}\in\mathbb{Z}_+^d$ with $\boldsymbol{k}\neq(0,\ldots,0)$, $\epsilon>0$, and let B be a Borel set of $\mathbb{D}$ equipped with Skorokhod J_1 topology. Suppose that B is bounded away from $\breve{\mathbb{D}}_{\leqslant\boldsymbol{k}}^{\epsilon}$ under $\boldsymbol{d}_{J_1}$ for some $\epsilon>0$. Then, there exist $\bar{\epsilon}>0$ and $\bar{\delta}>0$ such that the following claims hold:*

(a) $\boldsymbol{d}_{J_1}\big(B^{\bar\epsilon}, \breve{\mathbb{D}}^{\bar\epsilon}_{\leqslant \boldsymbol{k}}\big) > \bar\epsilon$;

(b) *Given any* $\mathcal{K} = (\mathcal{K}_{\boldsymbol{j}})_{\boldsymbol{j}\in\mathcal{P}_d} \in \mathbb{A}(\boldsymbol{k})$ *and* $\xi \in B^{\bar\epsilon} \cap \breve{\mathbb{D}}^{\bar\epsilon}_{\mathcal{K}}$, *in the expression* (63) *for* ξ *we have*

$$\big\|\boldsymbol{w}_{\boldsymbol{j},k}\big\| > \bar\delta \text{ and } \boldsymbol{w}_{\boldsymbol{j},k} \notin \bar{\mathbb{R}}^d_{\leqslant}(\boldsymbol{j}, \bar\delta), \qquad \forall \boldsymbol{j} \in \mathcal{P}_d,\ k \in [\mathcal{K}_{\boldsymbol{j}}], \tag{64}$$

where the set $\bar{\mathbb{R}}^d_{\leqslant}(\boldsymbol{j}, \epsilon) = \bar{\mathbb{R}}^d_{\leqslant}(\boldsymbol{j}, \epsilon; \bar{\boldsymbol{S}}, \boldsymbol{\alpha})$ *is defined in* (8).

Proof. Part (a) follows directly from the fact that B is bounded away from $\breve{\mathbb{D}}^{\epsilon}_{\leqslant \boldsymbol{k}}$ under $\boldsymbol{d}_{J_1}$ for some $\epsilon > 0$, as well as the monotonicity of $\breve{\mathbb{D}}^{\epsilon'}_{\mathcal{K}} \subseteq \breve{\mathbb{D}}^{\epsilon}_{\mathcal{K}}$ for any $\epsilon \geq \epsilon' \geq 0$ and $\mathcal{K} \in \mathbb{Z}_+^{\mathcal{P}_d}$.

Moving on to part (b), since there are only finitely elements in $\mathbb{A}(\boldsymbol{k})$, if suffices to fix some $\mathcal{K} = (\mathcal{K}_{\boldsymbol{j}})_{\boldsymbol{j}\in\mathcal{P}_d} \in \mathbb{A}(\boldsymbol{k})$ and prove the claim (64). Specifically, we take any $\bar\delta \in (0, \bar\epsilon)$ and proceed with a proof by contradiction. Suppose that, for some $\xi \in B^{\bar\epsilon} \cap \breve{\mathbb{D}}^{\bar\epsilon}_{\mathcal{K}}$, there are $\boldsymbol{j}^* \in \mathcal{P}_d$ and $k^* \in [\mathcal{K}_{\boldsymbol{j}^*}]$ (which, of course, requires that $\mathcal{K}_{\boldsymbol{j}^*} \geq 1$) such that $\big\|\boldsymbol{w}_{\boldsymbol{j}^*,k^*}\big\| \leq \bar\delta$ in the expression for ξ in (63). Then, with

$$\xi^*(t) \stackrel{\text{def}}{=} \sum_{\boldsymbol{j}\in\mathcal{P}_d\setminus\{\boldsymbol{j}^*\}} \sum_{k=1}^{\mathcal{K}_{\boldsymbol{j}}} \boldsymbol{w}_{\boldsymbol{j},k} \mathbb{I}_{[t_{\boldsymbol{j},k},1]}(t) + \sum_{k\in[\mathcal{K}_{\boldsymbol{j}^*}]:\ k\neq k^*} \boldsymbol{w}_{\boldsymbol{j}^*,k} \mathbb{I}_{[t_{\boldsymbol{j}^*,k},1]}(t), \qquad \forall t \in [0,1],$$

we construct a path ξ^* by removing the jump $\boldsymbol{w}_{\boldsymbol{j}^*,k^*}$ from ξ. By defining $\mathcal{K}' = (\mathcal{K}'_{\boldsymbol{j}})_{\boldsymbol{j}\in\mathcal{P}_d}$ as

$$\mathcal{K}'_{\boldsymbol{j}} \stackrel{\text{def}}{=} \begin{cases} \mathcal{K}_{\boldsymbol{j}} & \text{if } \boldsymbol{j} \neq \boldsymbol{j}^* \\ \mathcal{K}_{\boldsymbol{j}^*} - 1 & \text{if } \boldsymbol{j} = \boldsymbol{j}^* \end{cases},$$

we have $\xi^* \in \breve{\mathbb{D}}^{\bar\epsilon}_{\mathcal{K}'}$ and $\breve{c}(\mathcal{K}') < \breve{c}(\mathcal{K})$ (see (29) and (34)), thus implying $\xi^* \in \breve{\mathbb{D}}^{\bar\epsilon}_{\leqslant \boldsymbol{k}}$. However, due to

$$\boldsymbol{d}_{J_1}(\xi, \xi^*) \leq \sup_{t\in[0,1]} \|\xi(t) - \xi^*(t)\| = \big\|\boldsymbol{w}_{\boldsymbol{j}^*,k^*}\big\| \leq \bar\delta < \bar\epsilon$$

and $\xi \in B^{\bar\epsilon}$, we arrive at the contradiction that $\boldsymbol{d}_{J_1}(B^{\bar\epsilon}, \breve{\mathbb{D}}^{\bar\epsilon}_{\leqslant \boldsymbol{k}}) \leq \bar\epsilon$. In summary, we have shown that $\big\|\boldsymbol{w}_{\boldsymbol{j},k}\big\| > \bar\delta$ for any $\boldsymbol{j} \in \mathcal{P}_d,\ k \in [\mathcal{K}_{\boldsymbol{j}}]$ in the expression (63) for ξ.

Next, suppose that, for some $\xi \in B^{\bar\epsilon} \cap \breve{\mathbb{D}}^{\bar\epsilon}_{\mathcal{K}}$, there exist $\boldsymbol{j}^* \in \mathcal{P}_d$ and $k^* \in [\mathcal{K}_{\boldsymbol{j}^*}]$ (which, again, requires that $\mathcal{K}_{\boldsymbol{j}^*} \geq 1$) such that $\boldsymbol{w}_{\boldsymbol{j}^*,k^*} \in \bar{\mathbb{R}}^d_{\leqslant}(\boldsymbol{j}^*, \bar\delta)$. Then, by the definition of $\bar{\mathbb{R}}^d_{\leqslant}(\boldsymbol{j}^*, \bar\delta)$ in (8) and our choice of $\bar\delta \in (0, \bar\epsilon)$, there exists some $\boldsymbol{j}' \in \mathcal{P}_d$ with $\boldsymbol{j}' \neq \boldsymbol{j}^*$, $\alpha(\boldsymbol{j}') \leq \alpha(\boldsymbol{j}^*)$ such that $\boldsymbol{w}_{\boldsymbol{j}^*,k^*} \in \bar{\mathbb{R}}^d(\boldsymbol{j}', \bar\delta) \subseteq \bar{\mathbb{R}}^d(\boldsymbol{j}', \bar\epsilon)$. Now, define $\mathcal{K}' = (\mathcal{K}'_{\boldsymbol{j}})_{\boldsymbol{j}\in\mathcal{P}_d}$ by

$$\mathcal{K}'_{\boldsymbol{j}} \stackrel{\text{def}}{=} \begin{cases} \mathcal{K}_{\boldsymbol{j}^*} - 1 & \text{if } \boldsymbol{j} = \boldsymbol{j}^* \\ \mathcal{K}_{\boldsymbol{j}'} + 1 & \text{if } \boldsymbol{j} = \boldsymbol{j}' \\ \mathcal{K}_{\boldsymbol{j}} & \text{otherwise} \end{cases},$$

and note that $\xi \in \breve{\mathbb{D}}^{\bar\epsilon}_{\mathcal{K}'}$ due to $\boldsymbol{w}_{\boldsymbol{j}^*,k^*} \in \bar{\mathbb{R}}^d(\boldsymbol{j}', \bar\epsilon)$. Due to $\alpha(\boldsymbol{j}') \leq \alpha(\boldsymbol{j}^*)$, we have $\breve{c}(\mathcal{K}') \leq \breve{c}(\mathcal{K}) = c(\boldsymbol{k})$ (see (29)), and hence $\xi \in \breve{\mathbb{D}}^{\bar\epsilon}_{\leqslant \boldsymbol{k}}$ (see (37)). In light of the running assumption $\xi \in B^{\bar\epsilon}$, we arrive at the contradiction that $B^{\bar\epsilon} \cap \breve{\mathbb{D}}^{\bar\epsilon}_{\leqslant \boldsymbol{k}} \neq \emptyset$. This concludes the proof of part (b). □

Lemma C.8. *Let Assumption 5 hold. Let $\boldsymbol{k} = (k_j)_{j\in[d]} \in \mathbb{Z}_+^d$ such that $\boldsymbol{k} \neq \boldsymbol{0}$. Let $T, \epsilon > 0$ and $\boldsymbol{x} \in \mathbb{R}^d$. For any Borel set B of $(\mathbb{D}[0,T], \boldsymbol{d}_{J_1}^{[0,T]})$ that is bounded away from $\breve{\mathbb{D}}^{\epsilon}_{\leqslant \boldsymbol{k};\boldsymbol{x}}[0,T]$ under $\boldsymbol{d}_{J_1}^{[0,T]}$,*

$$\sum_{\mathcal{K}\in\mathbb{A}(\boldsymbol{k})} \breve{\mathbf{C}}^{[0,T]}_{\mathcal{K};\boldsymbol{x}}(B) < \infty.$$

Proof. Without loss of generality, we impose Assumption 1 (i.e., we focus on the case of $T = 1$ and $\boldsymbol{x} = \boldsymbol{0}$), and adopt the notations in (30). Also, since there are only finitely many elements in $\mathbb{A}(\boldsymbol{k})$, it suffices to fix some $\mathcal{K} = (\mathcal{K}_{\boldsymbol{j}})_{\boldsymbol{j}\in\mathcal{P}_d} \in \mathbb{A}(\boldsymbol{k})$ and some Borel set B of $(\mathbb{D}, \boldsymbol{d}_{J_1})$ that is bounded away from $\breve{\mathbb{D}}^{\epsilon}_{\leqslant \boldsymbol{k}}$, and show that $\breve{\mathbf{C}}_{\mathcal{K}}(B) < \infty$. Since the measure $\breve{\mathbf{C}}_{\mathcal{K}}(\cdot)$ is supported on $\breve{\mathbb{D}}^0_{\mathcal{K}}$ (see (30) and the remarks right after), it suffices to show that

$$\breve{\mathbf{C}}_{\mathcal{K}}\big(B \cap \breve{\mathbb{D}}^0_{\mathcal{K}}\big) < \infty.$$

Since B is bounded away from $\breve{\mathbb{D}}^{\epsilon}_{\leqslant \boldsymbol{k}}$, Lemma C.7 allows us to fix some $\bar{\delta} > 0$ such that the following claim holds: given any $\xi \in B \cap \breve{\mathbb{D}}^0_{\mathcal{K}}$, in the expression (63) for ξ we have

$$\big\|\boldsymbol{w}_{\boldsymbol{j},k}\big\| > \bar{\delta} \text{ and } \boldsymbol{w}_{\boldsymbol{j},k} \notin \bar{\mathbb{R}}^d_{\leqslant}(\boldsymbol{j}, \bar{\delta}), \qquad \forall \boldsymbol{j} \in \mathcal{P}_d,\ k \in [\mathcal{K}_{\boldsymbol{j}}].$$

Then, by the definition of $\breve{\mathbf{C}}_{\mathcal{K}}$ in (30),

$$\breve{\mathbf{C}}_{\mathcal{K}}\big(B \cap \breve{\mathbb{D}}^0_{\mathcal{K}}\big) \leq \frac{1}{\prod_{\boldsymbol{j}\in\mathcal{P}_d} \mathcal{K}_{\boldsymbol{j}}!} \cdot \prod_{\boldsymbol{j}\in\mathcal{P}_d} \Bigg(\underbrace{\mathbf{C}_{\boldsymbol{j}}\Big(\big\{\boldsymbol{w} \in \mathbb{R}^d_+ :\ \|\boldsymbol{w}\| > \bar{\delta},\ \boldsymbol{w} \notin \bar{\mathbb{R}}^d_{\leqslant}(\boldsymbol{j}, \bar{\delta})\big\}\Big)}_{\stackrel{\text{def}}{=} c_{\boldsymbol{j}}(\bar{\delta})} \Bigg)^{\mathcal{K}_{\boldsymbol{j}}}. \tag{65}$$

In particular, for any $\delta \in (0, \bar{\delta})$, the set $\big\{\boldsymbol{w} \in \mathbb{R}^d_+ :\ \|\boldsymbol{w}\| > \bar{\delta},\ \boldsymbol{w} \notin \bar{\mathbb{R}}^d_{\leqslant}(\boldsymbol{j}, \bar{\delta})\big\}$ is bounded away from $\bar{\mathbb{R}}^d_{\leqslant}(\boldsymbol{j}, \delta)$. Then, by the $\mathcal{MHRV}$ condition in Assumption 5 (in particular, Claim (9) in Definition 2.4) we get $c_{\boldsymbol{j}}(\bar{\delta}) < \infty$ for each $\boldsymbol{j} \in \mathcal{P}_d$. Plugging these bounds back into (65), we conclude the proof. □

Now, we are ready to state the proof of Propositions C.2 and C.3.

Proof of Proposition C.2. Recall the definition of $c : \mathbb{Z}^d_+ \to [0,\infty)$ in (31) and $\breve{c} : \mathbb{Z}^{\mathcal{P}_d}_+ \to [0,\infty)$ in (29). Define events

$$B_0 \stackrel{\text{def}}{=} \Big\{\boldsymbol{d}_{J_1}\big(\hat{\boldsymbol{L}}_n^{>\delta}, \bar{\boldsymbol{L}}_n\big) > \Delta\Big\}, \quad B_1 \stackrel{\text{def}}{=} \Big\{\widetilde{K}_n^{>\delta} = 0\Big\}, \quad B_2 \stackrel{\text{def}}{=} \Big\{\breve{c}(\boldsymbol{K}_n^{>\delta}) \leq c(\boldsymbol{k})\Big\}. \tag{66}$$

To establish Claim (46), it suffices to show that for any $\delta > 0$ small enough,

$$\lim_{n\to\infty} \mathbf{P}\big(B_0 \setminus B_1\big) \big/ \breve{\lambda}_{\boldsymbol{k}}(n) = 0, \tag{67}$$

$$\lim_{n\to\infty} \mathbf{P}\big((B_0 \cap B_1) \setminus B_2\big) \big/ \breve{\lambda}_{\boldsymbol{k}}(n) = 0, \tag{68}$$

$$\lim_{n\to\infty} \mathbf{P}\big(B_0 \cap B_1 \cap B_2\big) \big/ \breve{\lambda}_{\boldsymbol{k}}(n) = 0. \tag{69}$$

Proof of Claim (67). By Claim (56) of Lemma C.6, we get $\mathbf{P}\big((B_1)^c\big) = o\big(\breve{\lambda}_{\boldsymbol{k}}(n)\big)$ as $n \to \infty$ under any $\delta > 0$.

Proof of Claim (68). We prove this claim for any $\delta > 0$. Let

$$\mathbb{K}_{>} \stackrel{\text{def}}{=} \Big\{ \mathcal{K} \in \mathbb{Z}_{+}^{\mathcal{P}_d} : \ \check{c}(\mathcal{K}) > c(\boldsymbol{k}) \Big\}.$$

Note that $(B_0 \cap B_1) \setminus B_2 \subseteq (B_2)^c = \{\check{c}(\boldsymbol{K}_n^{>\delta}) > c(\boldsymbol{k})\} = \{\boldsymbol{K}_n^{>\delta} \in \mathbb{K}_{>}\}$. Meanwhile, by the definition of the linear function $\check{c}(\cdot)$ in (29) with coefficients $\alpha(\boldsymbol{j}) - 1 > 0$ for any $\boldsymbol{j} \in \mathcal{P}_d$ (see Assumption 5), there exists a subset $\mathbb{K}_{>}^* \subseteq \mathbb{K}_{>}$ such that $|\mathbb{K}_{>}^*| < \infty$ (i.e., containing only finitely many elements) and

$$\check{c}(\mathcal{K}') > c(\boldsymbol{k}) \quad \Longrightarrow \quad \mathcal{K}' \geq \mathcal{K} \text{ for some } \mathcal{K} \in \mathbb{K}_{>}^*.$$

Here, the ordering $\boldsymbol{x} \geq \boldsymbol{y}$ between two real vectors is equivalent to $x_i \geq y_i$ for all i. Then,

$$\big\{\check{c}(\boldsymbol{K}_n^{>\delta}) > c(\boldsymbol{k})\big\} \subseteq \bigcup_{\mathcal{K} \in \mathbb{K}_{>}^*} \big\{\boldsymbol{K}_n^{>\delta} \geq \mathcal{K}\big\}.$$

Therefore, to prove Claim (68), it suffices to fix some $\mathcal{K} \in \mathbb{K}_{>}^*$ (or, more generally, some $\mathcal{K} = (\mathcal{K}_{\boldsymbol{j}})_{\boldsymbol{j} \in \mathcal{P}_d} \in \mathbb{Z}_{+}^{\mathcal{P}_d}$ such that $\check{c}(\mathcal{K}) > c(\boldsymbol{k})$) and show that

$$\lim_{n \to \infty} \mathbf{P}\big(\boldsymbol{K}_n^{>\delta} \geq \mathcal{K}\big) \big/ \check{\lambda}_{\boldsymbol{k}}(n) = 0.$$

By the definition of $\boldsymbol{K}_n^{>\delta}$ in (40) and the law of $\boldsymbol{L}(t)$ in (21),

$$\begin{aligned}
&\mathbf{P}\big(\boldsymbol{K}_n^{>\delta} \geq \mathcal{K}\big) \qquad (70)\\
&= \prod_{\boldsymbol{j} \in \mathcal{P}_d} \mathbf{P}\Big(K_n^{>\delta}(\boldsymbol{j}) \geq \mathcal{K}_{\boldsymbol{j}}\Big) \quad \text{due to the independence of Poisson splitting}\\
&= \prod_{\boldsymbol{j} \in \mathcal{P}_d} \mathbf{P}\Big(\text{Poisson}\Big(n \underbrace{\nu_n\big(\bar{\mathbb{R}}^{>\delta}(\boldsymbol{j})\big)}_{\stackrel{\text{def}}{=} \theta_n(\boldsymbol{j}, \delta)}\Big) \geq \mathcal{K}_{\boldsymbol{j}}\Big) \qquad \text{by (37)–(39), where } \nu_n(A) = \nu\{n\boldsymbol{x} : \ \boldsymbol{x} \in A\}\\
&\leq \prod_{\boldsymbol{j} \in \mathcal{P}_d} \Big(n\theta_n(\boldsymbol{j}, \delta)\Big)^{\mathcal{K}_{\boldsymbol{j}}} \qquad \text{due to } \mathbf{P}\big(\text{Poisson}(\lambda) \geq k\big) \leq \lambda^k \ \forall \lambda \geq 0.
\end{aligned}$$

Under the $\mathcal{MHRV}$ condition in Assumption 5 (in particular, by Claim (9) in Definition 2.4), we have $\theta_n(\boldsymbol{j}, \delta) = O\big(\lambda_{\boldsymbol{j}}(n)\big)$ for each $\boldsymbol{j} \in \mathcal{P}_d$. Plugging these bounds into display (70), we get

$$\mathbf{P}\big(\boldsymbol{K}_n^{>\delta} \geq \mathcal{K}\big) = O\bigg(\prod_{\boldsymbol{j} \in \mathcal{P}_d} \big(n\lambda_{\boldsymbol{j}}(n)\big)^{\mathcal{K}_{\boldsymbol{j}}}\bigg), \qquad \text{as } n \to \infty.$$

By the definition of $\check{c}(\cdot)$ in (29) and $\lambda_{\boldsymbol{j}}(n) \in \mathcal{RV}_{-\alpha(\boldsymbol{j})}(n)$, we have $\prod_{\boldsymbol{j} \in \mathcal{P}_d} \big(n\lambda_{\boldsymbol{j}}(n)\big)^{\mathcal{K}_{\boldsymbol{j}}} \in \mathcal{RV}_{-\check{c}(\mathcal{K})}(n)$. Lastly, due to $\check{\lambda}_{\boldsymbol{k}}(n) \in \mathcal{RV}_{-c(\boldsymbol{k})}(n)$ (see (32)) and $\check{c}(\mathcal{K}) > c(\boldsymbol{k})$, we get $\mathbf{P}\big(\boldsymbol{K}_n^{>\delta} \geq \mathcal{K}\big) = o\big(\check{\lambda}_{\boldsymbol{k}}(n)\big)$. This concludes the proof of Claim (68).

Proof of Claim (69). We first note that the set

$$\mathbb{K}_{\leqslant} \stackrel{\text{def}}{=} \big\{ \mathcal{K} \in \mathbb{Z}_{+}^{\mathcal{P}_d} : \ \mathcal{K} \notin \mathbb{A}(\boldsymbol{k}), \ \check{c}(\mathcal{K}) \leq c(\boldsymbol{k}) \big\}$$

contains only finitely many elements. Therefore, it suffices to fix some $\mathcal{K} \in \mathbb{K}_{\leqslant}$ and show that

$$\lim_{n\to\infty} \mathbf{P}\Big(\big\{\boldsymbol{d}_{J_1}(\bar{\boldsymbol{L}}_n,\ \hat{\boldsymbol{L}}_n^{>\delta}) > \Delta\big\} \cap \big\{\boldsymbol{K}_n^{>\delta} = \mathcal{K},\ \widetilde{K}_n^{>\delta} = 0\big\}\Big) \Big/ \breve{\lambda}_{\boldsymbol{k}}(n) = 0, \qquad \forall \delta > 0 \text{ small enough.}$$

Let events $A_n(m;\delta,k,\Delta)$ be defined as in Lemma C.5, and let $|\mathcal{K}| = \sum_{\boldsymbol{j}\in\mathcal{P}_d} \mathcal{K}_{\boldsymbol{j}}$. By Lemma C.5,

$$\boldsymbol{d}_{J_1}(\bar{\boldsymbol{L}}_n,\ \hat{\boldsymbol{L}}_n^{>\delta}) \leq \Delta \text{ holds on the event } \big\{\boldsymbol{K}_n^{>\delta} = \mathcal{K},\ \widetilde{K}_n^{>\delta} = 0\big\} \cap \big(\cap_{m=1}^{|\mathcal{K}|+1} A_n(m;\delta,|\mathcal{K}|,\Delta)\big). \tag{71}$$

Then, it only remains to show that $\lim_{n\to\infty} \mathbf{P}\Big(\big(\cap_{m=1}^{|\mathcal{K}|+1} A_n(m;\delta,|\mathcal{K}|,\Delta)\big)^c\Big) \Big/ \breve{\lambda}_{\boldsymbol{k}}(n) = 0$. By the strong Markov property of the Lévy process $\bar{\boldsymbol{L}}_n$ at stopping times $\tau_n^{>\delta}(m)$ defined in (31), we have

$$\begin{aligned}
\mathbf{P}\left(\Big(\bigcap_{m=1}^{|\mathcal{K}|+1} A_n(m;\delta,|\mathcal{K}|,\Delta)\Big)^c\right) &\leq (|\mathcal{K}|+1)\cdot \mathbf{P}\left(\sup_{t\in[0,1]:\ t<\tau_n^{>\delta}(1)} \big\|\bar{\boldsymbol{L}}_n(t)\big\| > \frac{\Delta}{|\mathcal{K}|+1}\right) \\
&= o\Big(\breve{\lambda}_{\boldsymbol{k}}(n)\Big) \qquad \text{as } n\to\infty \text{ for any } \delta>0 \text{ small enough.}
\end{aligned}$$

The last line follows from Lemma C.4. This concludes the proof of Claim (69). □

Proof of Proposition C.3. (a) Note that

$$\begin{aligned}
& f(\hat{\boldsymbol{L}}_n^{>\delta})\mathbb{I}\big\{\bar{\boldsymbol{K}}_n^{>\delta} \notin \bar{\mathbb{A}}(\boldsymbol{k})\big\} \\
&\leq \|f\|\cdot \mathbb{I}\big\{\hat{\boldsymbol{L}}_n^{>\delta} \in B\big\}\mathbb{I}\big\{\bar{\boldsymbol{K}}_n^{>\delta} \notin \bar{\mathbb{A}}(\boldsymbol{k})\big\} \qquad \text{due to } B = \operatorname{supp}(f) \\
&\leq \|f\|\cdot \underbrace{\mathbb{I}\big\{\hat{\boldsymbol{L}}_n^{>\delta} \in B\big\}\mathbb{I}\big\{\widetilde{K}_n^{>\delta} \geq 1\big\}}_{\stackrel{\text{def}}{=} I_1(n,\delta)} + \|f\|\cdot \underbrace{\mathbb{I}\big\{\hat{\boldsymbol{L}}_n^{>\delta} \in B\big\}\mathbb{I}\big\{\breve{c}(\boldsymbol{K}_n^{>\delta}) > c(\boldsymbol{k})\big\}}_{\stackrel{\text{def}}{=} I_2(n,\delta)} \\
&\quad + \|f\|\cdot \underbrace{\mathbb{I}\big\{\hat{\boldsymbol{L}}_n^{>\delta} \in B\big\}\mathbb{I}\big\{\boldsymbol{K}_n^{>\delta} \notin \mathbb{A}(\boldsymbol{k}),\ \breve{c}(\boldsymbol{K}_n^{>\delta}) \leq c(\boldsymbol{k}),\ \widetilde{K}_n^{>\delta} = 0\big\}}_{\stackrel{\text{def}}{=} I_3(n,\delta)}.
\end{aligned}$$

Repeating the proof of Claims (67) and (68) for Proposition C.2, we get $\mathbf{E}\big[I_1(n,\delta)\big] = o\big(\breve{\lambda}_{\boldsymbol{k}}(n)\big)$ and $\mathbf{E}\big[I_2(n,\delta)\big] = o\big(\breve{\lambda}_{\boldsymbol{k}}(n)\big)$ (as $n\to\infty$) for any $\delta>0$. Next, let $\mathbb{K}_< \stackrel{\text{def}}{=} \big\{\mathcal{K}\in\mathbb{Z}_+^{\mathcal{P}_d}:\ \mathcal{K}\notin\mathbb{A}(\boldsymbol{k}),\ \breve{c}(\mathcal{K}) \leq c(\boldsymbol{k})\big\}$. By the definition of $\breve{\mathbb{D}}^{\epsilon}_{\leqslant\boldsymbol{k}} = \breve{\mathbb{D}}^{\epsilon}_{\leqslant\boldsymbol{k};\boldsymbol{0}}[0,1]$ in (39), for any $\mathcal{K}\in\mathbb{K}_<$ we have $\breve{\mathbb{D}}^{\delta}_{\mathcal{K}} \subset \breve{\mathbb{D}}^{\delta}_{\leqslant\boldsymbol{k}}$. Therefore, given $\mathcal{K}\in\mathbb{K}_<$, it holds on the event $\{\boldsymbol{K}_n^{>\delta} = \mathcal{K},\ \widetilde{K}_n^{>\delta} = 0\}$ that

$$\hat{\boldsymbol{L}}_n^{>\delta} \in \breve{\mathbb{D}}^{\delta}_{\mathcal{K}} \subseteq \breve{\mathbb{D}}^{\epsilon}_{\leqslant\boldsymbol{k}}, \qquad \text{due to property (43) and our choice of } \delta\in(0,\epsilon).$$

Since B is bounded away from $\breve{\mathbb{D}}^{\epsilon}_{\leqslant\boldsymbol{k}}$ under $\boldsymbol{d}_{J_1}$ we must have $\big\{\hat{\boldsymbol{L}}_n^{>\delta}\in B\big\} \cap \big\{\boldsymbol{K}_n^{>\delta}\notin\mathbb{A}(\boldsymbol{k}),\ \breve{c}(\boldsymbol{K}_n^{>\delta}) \leq c(\boldsymbol{k})\big\} = \emptyset$. In summary, we have shown that $I_3(n,\delta) = 0$ for each $\delta\in(0,\epsilon)$ and $n\geq 1$. This concludes the proof of part (a).

(b) Recall that there are only finitely many assignments for $\boldsymbol{k}$ (i.e., $|\mathbb{A}(\boldsymbol{k})|<\infty$). Therefore, it suffices to prove part (b) for some fixed $\mathcal{K} = (\mathcal{K}_{\boldsymbol{j}})_{\boldsymbol{j}\in\mathcal{P}_d} \in \mathbb{A}(\boldsymbol{k})$. By the definition of $B = \operatorname{supp}(f)$, we have $\breve{\mathbf{C}}_{\mathcal{K}}(f) \leq \|f\|\cdot\breve{\mathbf{C}}_{\mathcal{K}}(B)$. Using Lemma C.8, we confirm that $\breve{\mathbf{C}}_{\mathcal{K}}(f) < \infty$. Next, applying Lemma C.7

onto $B = \mathrm{supp}(f)$, one can fix some $\bar{\epsilon},\ \bar{\delta} > 0$ such that $\boldsymbol{d}_{J_1}(B^{\bar{\epsilon}}, \breve{\mathbb{D}}^{\bar{\epsilon}}_{\leqslant \boldsymbol{k}}) > \bar{\epsilon}$ and, given $\xi \in B^{\bar{\epsilon}} \cap \breve{\mathbb{D}}^{\bar{\epsilon}}_{\mathcal{K}}$, in the expression (63) for ξ we have

$$\big\|\boldsymbol{w}_{\boldsymbol{j},k}\big\| > \bar{\delta} \text{ and } \boldsymbol{w}_{\boldsymbol{j},k} \notin \bar{\mathbb{R}}^d_{\leqslant}(\boldsymbol{j}, \bar{\delta}), \qquad \forall \boldsymbol{j} \in \mathcal{P}_d,\ k \in [\mathcal{K}_{\boldsymbol{j}}]. \tag{72}$$

To proceed, let

$$I_{\mathcal{K}}(n,\delta) \stackrel{\text{def}}{=} f(\hat{\boldsymbol{L}}_n^{>\delta}) \mathbb{I}\big\{\boldsymbol{K}_n^{>\delta} = \mathcal{K},\ \widetilde{K}_n^{>\delta} = 0\big\}.$$

Note that $\{\boldsymbol{K}_n^{>\delta} = \mathcal{K},\ \widetilde{K}_n^{>\delta} = 0\} = E_n^{>\delta}(\mathcal{K})$; see (54). Also, note that Claim (35) holds for all but countably many $\delta > 0$, which is verified by Lemma C.1. Henceforth in this proof, we only consider $\delta \in (0, \bar{\delta})$ such that Claim (35) holds. Next, let $|\mathcal{K}| \stackrel{\text{def}}{=} \sum_{\boldsymbol{j} \in \mathcal{P}_d} \mathcal{K}_{\boldsymbol{j}}$, and define the mapping $h(\cdot)$ as follows: given $\mathbf{W} = (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_{|\mathcal{K}|}) \in (\mathbb{R}^d_+)^{|\mathcal{K}|}$ and $\boldsymbol{t} = (t_1, \ldots, t_{|\mathcal{K}|}) \in (0,1]^{|\mathcal{K}|}$, let

$$h(\mathbf{W}, \boldsymbol{t}) \stackrel{\text{def}}{=} \sum_{k=1,2,\ldots,|\mathcal{K}|} \boldsymbol{w}_k \mathbb{I}_{[t_k,1]}. \tag{73}$$

One can see that $h : (\mathbb{R}^d_+)^{|\mathcal{K}|} \times (0,1]^{|\mathcal{K}|} \to \mathbb{D}$ is continuous (w.r.t. the J_1 topology of $\mathbb{D}$) when restricted on the domain $(\mathbb{R}^d_+)^{|\mathcal{K}|} \times (0,1)^{|\mathcal{K}|,*}$, with $A^{k,*} \stackrel{\text{def}}{=} \big\{(t_1, \ldots, t_k) \in A^k :\ t_i \neq t_j\ \forall i, j \in [d] \text{ with } i \neq j\big\}$. This allows us to apply the continuous mapping theorem and study the asymptotic law of $f(\hat{\boldsymbol{L}}_n^{>\delta})$ conditioned on the event $E_n^{>\delta}(\mathcal{K})$. Specifically, on $\{\boldsymbol{K}_n^{>\delta} = \mathcal{K},\ \widetilde{K}_n^{>\delta} = 0\}$, we write

$$\mathbf{W}_n^{>\delta} = \Big(\boldsymbol{W}_n^{>\delta}(k;\boldsymbol{j})\Big)_{\boldsymbol{j} \in \mathcal{P}_d,\ k \in [\mathcal{K}_{\boldsymbol{j}}]}, \qquad \mathbf{T}_n^{>\delta} = \Big(\tau_n^{>\delta}(k;\boldsymbol{j})\Big)_{\boldsymbol{j} \in \mathcal{P}_d,\ k \in [\mathcal{K}_{\boldsymbol{j}}]}.$$

By (43), on the event $E_n^{>\delta}(\mathcal{K})$ we have $\hat{\boldsymbol{L}}_n^{>\delta} = h\big(\mathbf{W}_n^{>\delta}, \mathbf{T}_n^{>\delta}\big)$. As a result,

$$\frac{\mathbf{E}\big[I_{\mathcal{K}}(n,\delta)\big]}{\breve{\lambda}_{\boldsymbol{k}}(n)} = \mathbf{E}\Big[f\Big(h\big(\mathbf{W}_n^{>\delta}, \mathbf{T}_n^{>\delta}\big)\Big) \,\Big|\, E_n^{>\delta}(\mathcal{K})\Big] \cdot \frac{\mathbf{P}\big(E_n^{>\delta}(\mathcal{K})\big)}{\breve{\lambda}_{\boldsymbol{k}}(n)}.$$

Then, by Lemma C.6 and the continuity of $f\big(h(\cdot)\big)$ on $(\mathbb{R}^d_+)^{|\mathcal{K}|} \times (0,1)^{|\mathcal{K}|,*}$, we obtain the following results (note that $f(\xi) = 0\ \forall \xi \notin B$), where we use the $U_{\boldsymbol{j},k}$'s to denote i.i.d. copies of Unif$(0,1)$, $U_{\boldsymbol{j},(1:k)} \leq U_{\boldsymbol{j},(2:k)} \leq \ldots \leq U_{\boldsymbol{j},(k:k)}$ for the order statistics of $(U_{\boldsymbol{j},q})_{q \in [k]}$, $\mathcal{L}_I$ for the Lebesgue measure restricted on interval I, and $\mathbf{C}_{\boldsymbol{j}}(\cdot)$ for the location measure in the $\mathcal{MHRV}$ condition of Assumption 5 supported on $\mathbb{R}^d(\boldsymbol{j})$ (see Definition 2.4):

$$\begin{aligned}
&\lim_{n\to\infty} \mathbf{E}\Big[f\Big(h\big(\mathbf{W}_n^{>\delta}, \mathbf{T}_n^{>\delta}\big)\Big) \,\Big|\, E_n^{>\delta}(\mathcal{K})\Big] \\
&= \int f\Bigg(\sum_{\boldsymbol{j} \in \mathcal{P}_d} \sum_{k \in [\mathcal{K}_{\boldsymbol{j}}]} \boldsymbol{w}_{\boldsymbol{j},k} \mathbb{I}_{[t_{\boldsymbol{j},k},1]}\Bigg) \\
&\quad \cdot \mathbb{I}\Bigg\{\sum_{\boldsymbol{j} \in \mathcal{P}_d} \sum_{k \in [\mathcal{K}_{\boldsymbol{j}}]} \boldsymbol{w}_{\boldsymbol{j},k} \mathbb{I}_{[t_{\boldsymbol{j},k},1]} \in B\Bigg\} \cdot \mathbb{I}\Big\{\boldsymbol{w}_{\boldsymbol{j},k} \in \bar{\mathbb{R}}^{>\delta}(\boldsymbol{j})\ \forall \boldsymbol{j} \in \mathcal{P}_d,\ k \in [\mathcal{K}_{\boldsymbol{j}}]\Big\} \\
&\quad \cdot \Bigg(\bigtimes_{\boldsymbol{j} \in \mathcal{P}_d} \bigtimes_{k \in [\mathcal{K}_{\boldsymbol{j}}]} \frac{\mathbf{C}_{\boldsymbol{j}}(d\boldsymbol{w}_{\boldsymbol{j},k})}{\mathbf{C}_{\boldsymbol{j}}\big(\bar{\mathbb{R}}^{>\delta}(\boldsymbol{j})\big)}\Bigg)\Bigg(\bigtimes_{\boldsymbol{j} \in \mathcal{P}_d} \mathbf{P}\Big(\Big(U_{\boldsymbol{j},(1:\mathcal{K}_{\boldsymbol{j}})}, \ldots, U_{\boldsymbol{j},(\mathcal{K}_{\boldsymbol{j}}:\mathcal{K}_{\boldsymbol{j}})}\Big) \in d(t_{\boldsymbol{j},1}, \ldots, t_{\boldsymbol{j},\mathcal{K}_{\boldsymbol{j}}})\Big)\Bigg)
\end{aligned}$$

$$
\begin{aligned}
&\overset{(*)}{=} \int f\left(\sum_{\boldsymbol{j}\in\mathcal{P}_d}\sum_{k\in[\mathcal{K}_{\boldsymbol{j}}]} \boldsymbol{w}_{\boldsymbol{j},k}\mathbb{I}_{[t_{\boldsymbol{j},k},1]}\right) \\
&\quad\cdot \mathbb{I}\left\{\sum_{\boldsymbol{j}\in\mathcal{P}_d}\sum_{k\in[\mathcal{K}_{\boldsymbol{j}}]} \boldsymbol{w}_{\boldsymbol{j},k}\mathbb{I}_{[t_{\boldsymbol{j},k},1]} \in B\right\}\cdot \mathbb{I}\left\{\boldsymbol{w}_{\boldsymbol{j},k}\notin \bar{\mathbb{R}}^d_{\leqslant}(\boldsymbol{j},\delta)\ \forall \boldsymbol{j}\in\mathcal{P}_d,\ k\in[\mathcal{K}_{\boldsymbol{j}}]\right\} \\
&\quad\cdot\left(\bigtimes_{\boldsymbol{j}\in\mathcal{P}_d}\bigtimes_{k\in[\mathcal{K}_{\boldsymbol{j}}]}\frac{\mathbf{C}_{\boldsymbol{j}}(d\boldsymbol{w}_{\boldsymbol{j},k})}{\mathbf{C}_{\boldsymbol{j}}(\bar{\mathbb{R}}^{>\delta}(\boldsymbol{j}))}\right)\left(\bigtimes_{\boldsymbol{j}\in\mathcal{P}_d}\mathbf{P}\left(\left(U_{\boldsymbol{j},(1:\mathcal{K}_{\boldsymbol{j}})},\ldots,U_{\boldsymbol{j},(\mathcal{K}_{\boldsymbol{j}}:\mathcal{K}_{\boldsymbol{j}})}\right)\in d(t_{\boldsymbol{j},1},\ldots,t_{\boldsymbol{j},\mathcal{K}_{\boldsymbol{j}}})\right)\right) \\
&\overset{(\triangle)}{=} \int f\left(\sum_{\boldsymbol{j}\in\mathcal{P}_d}\sum_{k\in[\mathcal{K}_{\boldsymbol{j}}]} \boldsymbol{w}_{\boldsymbol{j},k}\mathbb{I}_{[t_{\boldsymbol{j},k},1]}\right) \\
&\quad\cdot\left(\bigtimes_{\boldsymbol{j}\in\mathcal{P}_d}\bigtimes_{k\in[\mathcal{K}_{\boldsymbol{j}}]}\frac{\mathbf{C}_{\boldsymbol{j}}(d\boldsymbol{w}_{\boldsymbol{j},k})}{\mathbf{C}_{\boldsymbol{j}}(\bar{\mathbb{R}}^{>\delta}(\boldsymbol{j}))}\right)\left(\bigtimes_{\boldsymbol{j}\in\mathcal{P}_d}\mathbf{P}\left(\left(U_{\boldsymbol{j},(1:\mathcal{K}_{\boldsymbol{j}})},\ldots,U_{\boldsymbol{j},(\mathcal{K}_{\boldsymbol{j}}:\mathcal{K}_{\boldsymbol{j}})}\right)\in d(t_{\boldsymbol{j},1},\ldots,t_{\boldsymbol{j},\mathcal{K}_{\boldsymbol{j}}})\right)\right).
\end{aligned}
$$

In the display above, the step $(*)$ holds since $\bar{\mathbb{R}}^{>\delta}(\boldsymbol{j}) = \bar{\mathbb{R}}^d(\boldsymbol{j},\delta)\setminus\bar{\mathbb{R}}^d_{\leqslant}(\boldsymbol{j},\delta)$, and $\mathbf{C}_{\boldsymbol{j}}(\cdot)$ is supported on $\mathbb{R}^d(\boldsymbol{j})\subseteq\bar{\mathbb{R}}^d(\boldsymbol{j},\delta)$; the step $(\triangle)$ follows from the definition of $B=\mathrm{supp}(f)$, our choice of $\bar{\delta}$ in (72) and $\delta\in(0,\bar{\delta})$. To proceed, we make a few observations. First, using $U_{(1:k)}<U_{(2:k)}<\ldots<U_{(k:k)}$ to denote the order statistics of k copies of $\mathrm{Unif}(0,1)$, we have

$$
\mathbf{P}\Big((U_{(1:k)},U_{(2:k)},\ldots,U_{(k:k)})\in d(t_1,\ldots,t_k)\Big) = k!\mathbb{I}\{0<t_1<t_2<\ldots<t_k<1\}dt_1dt_2\ldots dt_k.
$$

Second, the mapping h defined in (73) is invariant under permutation of its arguments. That is, given $\mathbf{W}=(\boldsymbol{w}_1,\ldots,\boldsymbol{w}_{|\mathcal{K}|})$ and $\boldsymbol{t}=(t_1,\ldots,t_{|\mathcal{K}|})$, we have

$$
h(\mathbf{W},\boldsymbol{t}) = h\Big((\boldsymbol{w}_{\sigma(i)})_{i=1,\ldots,|\mathcal{K}|},\ (t_{\sigma(i)})_{i=1,\ldots,|\mathcal{K}|}\Big)
$$

for any permutation $(\sigma(i))_{i=1,\ldots,|\mathcal{K}|}$ of $\{1,2,\ldots,|\mathcal{K}|\}$. These two properties allow us to re-evaluate step $(\triangle)$ as an integral over the domain $\{(t_{\boldsymbol{j},k})_{\boldsymbol{j},k}:\ t_{\boldsymbol{j},k}\in(0,1)\forall\boldsymbol{j},k\}$ instead of imposing the ordering constraint $t_{\boldsymbol{j},1}<t_{\boldsymbol{j},2}<\ldots<t_{\boldsymbol{j},\mathcal{K}_{\boldsymbol{j}}}$ for each $\boldsymbol{j}$. Specifically,

$$
\begin{aligned}
&\lim_{n\to\infty}\mathbf{E}\left[f\Big(h\big(\mathbf{W}_n^{>\delta},\mathbf{T}_n^{>\delta}\big)\Big)\,\middle|\,E_n^{>\delta}(\mathcal{K})\right] \qquad (74)\\
&= \left[\prod_{\boldsymbol{j}\in\mathcal{P}_d}\Big(\mathbf{C}_{\boldsymbol{j}}(\bar{\mathbb{R}}^{>\delta}(\boldsymbol{j}))\Big)^{-\mathcal{K}_{\boldsymbol{j}}}\right] \\
&\qquad\cdot\int f\left(\sum_{\boldsymbol{j}\in\mathcal{P}_d}\sum_{k\in[\mathcal{K}_{\boldsymbol{j}}]}\boldsymbol{w}_{\boldsymbol{j},k}\mathbb{I}_{[t_{\boldsymbol{j},k},1]}\right)\bigtimes_{\boldsymbol{j}\in\mathcal{P}_d}\bigtimes_{k\in[\mathcal{K}_{\boldsymbol{j}}]}\Big(\mathbf{C}_{\boldsymbol{j}}\times\mathcal{L}_{(0,1)}\Big(d(\boldsymbol{w}_{\boldsymbol{j},k},t_{\boldsymbol{j},k})\Big)\Big) \\
&= \left[\prod_{\boldsymbol{j}\in\mathcal{P}_d}\Big(\mathbf{C}_{\boldsymbol{j}}(\bar{\mathbb{R}}^{>\delta}(\boldsymbol{j}))\Big)^{-\mathcal{K}_{\boldsymbol{j}}}\right]\cdot\left(\prod_{\boldsymbol{j}\in\mathcal{P}_d}\mathcal{K}_{\boldsymbol{j}}!\right)\cdot\check{\mathbf{C}}_{\mathcal{K}}(f) \qquad \text{by definitions in (30) and (30).}
\end{aligned}
$$

On the other hand, applying the claim (57) in Lemma C.6, we get

$$\lim_{n\to\infty} \frac{\mathbf{P}\big(E_n^{>\delta}(\mathcal{K})\big)}{\breve{\lambda}_{\boldsymbol{k}}(n)} = \prod_{\boldsymbol{j}\in\mathcal{P}_d} \frac{1}{\mathcal{K}_{\boldsymbol{j}}!} \cdot \Big(\mathbf{C}_{\boldsymbol{j}}\big(\bar{\mathbb{R}}^{>\delta}(\boldsymbol{j})\big)\Big)^{\mathcal{K}_{\boldsymbol{j}}}. \tag{75}$$

Combining (74) and (75), we conclude the proof. □

Appendix D: Proof for Large Deviations of Multivariate Heavy-Tailed Hawkes Processes

D.1. Adapting Theorem 3.2 to $\mathbb{D}[0,\infty)$

We first establish sample path large deviations for Lévy processes with $\mathcal{MHRV}$ increments (i.e., a suitable modification of Theorem 3.2) w.r.t. the J_1 topology on $\mathbb{D}[0,\infty)$. Specifically, recall that the projection mapping $\phi_t : \mathbb{D}[0,\infty) \to \mathbb{D}[0,t]$ is defined by $\phi_t(\xi)(s) = \xi(s)$ for any $s \in [0,t]$, and let

$$\boldsymbol{d}_{J_1}^{[0,\infty)}(\xi^{(1)},\xi^{(2)}) \stackrel{\text{def}}{=} \int_0^\infty e^{-t} \cdot \Big[\boldsymbol{d}_{J_1}^{[0,t]}\Big(\phi_t\big(\xi^{(1)}\big),\ \phi_t\big(\xi^{(2)}\big)\Big) \wedge 1\Big] dt, \qquad \forall \xi^{(1)},\xi^{(2)} \in \mathbb{D}[0,\infty). \tag{76}$$

We need the following two lemmas. First, Lemma D.1 adapts Lemmas C.7 and C.8 to $\mathbb{D}[0,\infty)$.

Lemma D.1. *Let Assumption 5 hold. Let $\boldsymbol{k} = (k_j)_{j\in[d]} \in \mathbb{Z}_+^d \setminus \{\mathbf{0}\}$, and $\epsilon > 0$. Given any Borel set B of $(\mathbb{D}[0,\infty), \boldsymbol{d}_{J_1}^{[0,\infty)})$ that is bounded away from $\breve{\mathbb{D}}^{\epsilon}_{\leqslant \boldsymbol{k};\boldsymbol{\mu_L}}[0,\infty)$ under $\boldsymbol{d}_{J_1}^{[0,\infty)}$,*

(i) *there exist $T \in (0,\infty)$ and $\bar{\delta} \in (0,\infty)$ such that the following claim holds: for any $\mathcal{K} = (\mathcal{K}_{\boldsymbol{j}})_{\boldsymbol{j}\in\mathcal{P}_d} \in \mathbb{A}(\boldsymbol{k})$ and $\xi \in B \cap \breve{\mathbb{D}}^{\epsilon}_{\mathcal{K};\boldsymbol{\mu_L}}[0,\infty)$, in the expression (28) for ξ (i.e., under $I = [0,\infty)$ and $\boldsymbol{x} = \boldsymbol{\mu_L}$), we have*

$$t_{\boldsymbol{j},k} \le T, \quad \big\|\boldsymbol{w}_{\boldsymbol{j},k}\big\| > \bar{\delta}, \text{ and } \boldsymbol{w}_{\boldsymbol{j},k} \notin \bar{\mathbb{R}}^d_{\leqslant}(\boldsymbol{j},\bar{\delta}), \qquad \forall \boldsymbol{j} \in \mathcal{P}_d,\ k \in [\mathcal{K}_{\boldsymbol{j}}]; \tag{77}$$

(ii) $\sum_{\mathcal{K}\in\mathbb{A}(\boldsymbol{k})} \breve{\mathbf{C}}^{[0,\infty)}_{\mathcal{K};\boldsymbol{\mu_L}}(B) < \infty$.

Proof. Fix some $\Delta > 0$ such that $\boldsymbol{d}_{J_1}^{[0,\infty)}\big(B, \breve{\mathbb{D}}^{\epsilon}_{\leqslant\boldsymbol{k};\boldsymbol{\mu_L}}[0,\infty)\big) > \Delta$.

(i) Since $|\mathbb{A}(\boldsymbol{k})| < \infty$, it suffices to prove part (i) for some fixed $\mathcal{K} = (\mathcal{K}_{\boldsymbol{j}})_{\boldsymbol{j}\in\mathcal{P}_d} \in \mathbb{A}(\boldsymbol{k})$. We first show that Claim (77) holds for any $T > 0$ large enough such that

$$e^{-T} < \Delta. \tag{78}$$

We consider a proof by contradiction. Suppose that for some $\xi \in B \cap \breve{\mathbb{D}}^{\epsilon}_{\mathcal{K};\boldsymbol{\mu_L}}[0,\infty)$, in expression (28) there exist some $\boldsymbol{j}^* \in \mathcal{P}_d$ and $k^* \in [\mathcal{K}_{\boldsymbol{j}^*}]$ such that $t_{\boldsymbol{j}^*,k^*} > T$. By defining

$$\xi'(t) \stackrel{\text{def}}{=} \boldsymbol{\mu_L} t + \sum_{\boldsymbol{j}\in\mathcal{P}_d} \sum_{k=1}^{\mathcal{K}_{\boldsymbol{j}}} \boldsymbol{w}_{\boldsymbol{j},k} \mathbb{I}\{t_{\boldsymbol{j},k} \le T\} \cdot \mathbb{I}_{[t_{\boldsymbol{j},k},\infty)}(t), \qquad \forall t \ge 0,$$

we construct a path ξ' by removing any jump in ξ that arrives after time T. On the one hand, the existence of $t_{\boldsymbol{j}^*,k^*} > T$ implies that we are removing at least one jump when defining ξ', and hence

$\xi' \in \breve{\mathbb{D}}^{\epsilon}_{\leqslant \boldsymbol{k};\boldsymbol{\mu_L}}[0,\infty)$ by definitions in (37). On the other hand, since $\xi(t) = \xi'(t)$ for each $t \in [0,T]$, we have $\boldsymbol{d}^{[0,\infty)}_{J_1}(\xi,\xi') \leq \int_{t>T} e^{-t}dt = e^{-T} < \Delta$. Due to $\xi \in B$, we arrive at the contradiction that $\boldsymbol{d}^{[0,\infty)}_{J_1}\left(B, \breve{\mathbb{D}}^{\epsilon}_{\leqslant \boldsymbol{k};\boldsymbol{\mu_L}}[0,\infty)\right) < \Delta$, which verifies claim (77) under our choice of T in (78).

Next, through a proof by contradiction, we show that Claim (77) holds under

$$\bar{\delta} \in (0, \epsilon \wedge \Delta). \tag{79}$$

First, suppose that for some $\xi \in B \cap \breve{\mathbb{D}}^{\epsilon}_{\mathcal{K};\boldsymbol{\mu_L}}[0,\infty)$, in expression (28) there exist some $\boldsymbol{j}^* \in \mathcal{P}_d$ and $k^* \in [\mathcal{K}_{\boldsymbol{j}^*}]$ such that $\left\|\boldsymbol{w}_{\boldsymbol{j}^*,k^*}\right\| \leq \bar{\delta}$. By defining

$$\xi'(t) \stackrel{\text{def}}{=} \boldsymbol{\mu_L} t + \sum_{\boldsymbol{j}\in\mathcal{P}_d}\sum_{k=1}^{\mathcal{K}_{\boldsymbol{j}}} \boldsymbol{w}_{\boldsymbol{j},k}\mathbb{I}\Big\{(\boldsymbol{j},k) \neq (\boldsymbol{j}^*,k^*)\Big\} \cdot \mathbb{I}_{[t_{\boldsymbol{j},k},\infty)}(t), \qquad \forall t \geq 0,$$

we construct a path ξ' by removing the jump $\boldsymbol{w}_{\boldsymbol{j}^*,k^*}$—arriving at time $t_{\boldsymbol{j}^*,k^*}$—from ξ. Again, we have $\xi' \in \breve{\mathbb{D}}^{\epsilon}_{\leqslant \boldsymbol{k};\boldsymbol{\mu_L}}[0,\infty)$. On the other hand, due to $\left\|\boldsymbol{w}_{\boldsymbol{j}^*,k^*}\right\| \leq \bar{\delta}$, we have $\sup_{t\geq 0}\|\xi(t) - \xi'(t)\| \leq \bar{\delta}$, and hence $\boldsymbol{d}^{[0,\infty)}_{J_1}(\xi,\xi') \leq \int_0^\infty e^{-t}\cdot\bar{\delta}dt = \bar{\delta} < \Delta$; see (79). We then arrive at the contradiction that $\boldsymbol{d}^{[0,\infty)}_{J_1}\left(B, \breve{\mathbb{D}}^{\epsilon}_{\leqslant \boldsymbol{k};\boldsymbol{\mu_L}}[0,\infty)\right) < \Delta$. Second, suppose that for some $\xi \in B \cap \breve{\mathbb{D}}^{\epsilon}_{\mathcal{K};\boldsymbol{\mu_L}}[0,\infty)$, in expression (28) there exist some $\boldsymbol{j}^* \in \mathcal{P}_d$ and $k^* \in [\mathcal{K}_{\boldsymbol{j}^*}]$ such that $\boldsymbol{w}_{\boldsymbol{j}^*,k^*} \in \bar{\mathbb{R}}^d_{\leqslant}(\boldsymbol{j}^*,\bar{\delta}) \subseteq \bar{\mathbb{R}}^d_{\leqslant}(\boldsymbol{j}^*,\epsilon)$ (due to our choice of $\bar{\delta} < \epsilon$ in (79)). Then, by definitions in (8), there exists some $\boldsymbol{j}' \in \mathcal{P}_d$ with $\boldsymbol{j}' \neq \boldsymbol{j}^*$, $\alpha(\boldsymbol{j}') \leq \alpha(\boldsymbol{j}^*)$ such that $\boldsymbol{w}_{\boldsymbol{j}^*,k^*} \in \bar{\mathbb{R}}^d(\boldsymbol{j}',\epsilon)$. Now, define $\mathcal{K}' = (\mathcal{K}'_{\boldsymbol{j}})_{\boldsymbol{j}\in\mathcal{P}_d}$ by

$$\mathcal{K}'_{\boldsymbol{j}} \stackrel{\text{def}}{=} \begin{cases} \mathcal{K}_{\boldsymbol{j}^*} - 1 & \text{if } \boldsymbol{j} = \boldsymbol{j}^* \\ \mathcal{K}_{\boldsymbol{j}'} + 1 & \text{if } \boldsymbol{j} = \boldsymbol{j}' \\ \mathcal{K}_{\boldsymbol{j}} & \text{otherwise} \end{cases}.$$

First, $\boldsymbol{w}_{\boldsymbol{j}^*,k^*} \in \bar{\mathbb{R}}^d(\boldsymbol{j}',\epsilon)$ implies that $\xi \in \breve{\mathbb{D}}^{\epsilon}_{\mathcal{K}';\boldsymbol{\mu_L}}[0,\infty)$. Moreover, due to $\alpha(\boldsymbol{j}') \leq \alpha(\boldsymbol{j}^*)$, we have $\check{c}(\mathcal{K}') \leq \check{c}(\mathcal{K}) = c(\boldsymbol{k})$, and hence $\breve{\mathbb{D}}^{\epsilon}_{\mathcal{K}';\boldsymbol{\mu_L}}[0,\infty) \subseteq \breve{\mathbb{D}}^{\epsilon}_{\leqslant \boldsymbol{k};\boldsymbol{\mu_L}}[0,\infty)$. By $\xi \in B$, we now arrive at $B \cap \breve{\mathbb{D}}^{\epsilon}_{\leqslant \boldsymbol{k};\boldsymbol{\mu_L}}[0,\infty) \neq \emptyset$, which clearly contradicts $\boldsymbol{d}^{[0,\infty)}_{J_1}\left(B, \breve{\mathbb{D}}^{\epsilon}_{\leqslant \boldsymbol{k};\boldsymbol{\mu_L}}[0,\infty)\right) > \Delta$. This concludes the proof of claim (77) under our choice of $\bar{\delta}$ in (79).

(*ii*) Again, it suffices to fix some $\mathcal{K} = (\mathcal{K}_{\boldsymbol{j}})_{\boldsymbol{j}\in\mathcal{P}_d} \in \mathbb{A}(\boldsymbol{k})$ and show that $\breve{\mathbf{C}}^{[0,\infty)}_{\mathcal{K};\boldsymbol{\mu_L}}(B) < \infty$. Let T and $\bar{\delta}$ be characterized as in part (*i*). By the definition of $\breve{\mathbf{C}}^{[0,\infty)}_{\mathcal{K};\boldsymbol{\mu_L}}$ in (30),

$$\breve{\mathbf{C}}^{[0,\infty)}_{\mathcal{K};\boldsymbol{\mu_L}}(B) \leq \frac{1}{\prod_{\boldsymbol{j}\in\mathcal{P}_d}\mathcal{K}_{\boldsymbol{j}}!} \cdot \int \mathbb{I}\left\{t_{\boldsymbol{j},k} \leq T,\ \left\|\boldsymbol{w}_{\boldsymbol{j},k}\right\| > \bar{\delta},\ \text{and } \boldsymbol{w}_{\boldsymbol{j},k} \notin \bar{\mathbb{R}}^d_{\leqslant}(\boldsymbol{j},\bar{\delta}), \quad \forall \boldsymbol{j} \in \mathcal{P}_d,\ k \in [\mathcal{K}_{\boldsymbol{j}}]\right\}$$
$$\bigtimes_{\boldsymbol{j}\in\mathcal{P}_d}\bigtimes_{k\in[\mathcal{K}_{\boldsymbol{j}}]}\Big((\mathbf{C}_{\boldsymbol{j}} \times \mathcal{L}_{(0,\infty)})\Big(d(\boldsymbol{w}_{\boldsymbol{j},k}, t_{\boldsymbol{j},k})\Big)\Big)$$
$$\leq \frac{1}{\prod_{\boldsymbol{j}\in\mathcal{P}_d}\mathcal{K}_{\boldsymbol{j}}!} \cdot \prod_{\boldsymbol{j}\in\mathcal{P}_d}\Bigg(T \cdot \underbrace{\mathbf{C}_{\boldsymbol{j}}\Big(\big\{\boldsymbol{w}\in\mathbb{R}^d_+ :\ \|\boldsymbol{w}\| > \bar{\delta},\ \boldsymbol{w} \notin \bar{\mathbb{R}}^d_{\leqslant}(\boldsymbol{j},\bar{\delta})\big\}\Big)}_{\stackrel{\text{def}}{=} c_{\boldsymbol{j}}(\bar{\delta})}\Bigg)^{\mathcal{K}_{\boldsymbol{j}}}.$$

Furthermore, it has been shown in the proof of Lemma C.8 that $c_{\boldsymbol{j}}(\bar{\delta}) < \infty$. This concludes the proof of part (ii). □

Recall the definition of the projection mapping $\phi_t : \mathbb{D}[0,\infty) \to \mathbb{D}[0,t]$ in (43), i.e., $\phi_t(\xi)(s) = \xi(s)$ for any $s \in [0,t]$. Lemma D.2 connects the bounded-away-from-$\check{\mathbb{D}}^{\epsilon}_{\leqslant \boldsymbol{k};\boldsymbol{x}}[0,\infty)$ condition (under $\boldsymbol{d}^{[0,\infty)}_{J_1}$) to that of $\check{\mathbb{D}}^{\epsilon}_{\leqslant \boldsymbol{k};\boldsymbol{x}}[0,T]$ and $\boldsymbol{d}^{[0,T]}_{J_1}$.

Lemma D.2. *Let $\boldsymbol{k} \in \mathbb{Z}^d_+ \setminus \{\boldsymbol{0}\}$, $\epsilon > 0$, and $\boldsymbol{x} \in \mathbb{R}^d$. Given a Borel set B of $(\mathbb{D}[0,\infty), \boldsymbol{d}^{[0,\infty)}_{J_1})$ that is bounded away from $\check{\mathbb{D}}^{\epsilon}_{\leqslant \boldsymbol{k};\boldsymbol{x}}[0,\infty)$ under $\boldsymbol{d}^{[0,\infty)}_{J_1}$, it holds for any $T > 0$ large enough that $\phi_T(B)$ is bounded away from $\check{\mathbb{D}}^{\epsilon}_{\leqslant \boldsymbol{k};\boldsymbol{x}}[0,T]$ under $\boldsymbol{d}^{[0,T]}_{J_1}$.*

Proof. Fix some $\Delta > 0$ such that $\boldsymbol{d}^{[0,\infty)}_{J_1}\left(B, \check{\mathbb{D}}^{\epsilon}_{\leqslant \boldsymbol{k};\boldsymbol{x}}[0,\infty)\right) > \Delta$. Using a proof by contradiction, we show that the claim holds for any $T > 0$ large enough such that

$$e^{-T} < \Delta/2. \tag{80}$$

Specifically, suppose there are sequences $\xi_n \in \phi_T(B)$, $\xi'_n \in \check{\mathbb{D}}^{\epsilon}_{\leqslant \boldsymbol{k};\boldsymbol{x}}[0,T]$ such that

$$\lim_{n\to\infty} \boldsymbol{d}^{[0,T]}_{J_1}(\xi_n, \xi'_n) = 0. \tag{81}$$

Then, considering the definition of $\check{\mathbb{D}}^{\epsilon}_{\leqslant \boldsymbol{k};\boldsymbol{x}}[0,T]$ in (37) and by picking a sub-sequence if needed, we can w.l.o.g. assume the existence of $\mathcal{K} = (\mathcal{K}_{\boldsymbol{j}})_{\boldsymbol{j}\in\mathcal{P}_d} \in \mathbb{Z}_+^{\mathcal{P}_d}$ such that $\mathcal{K} \in \mathbb{A}(\boldsymbol{k})$, $\check{c}(\mathcal{K}) \le c(\boldsymbol{k})$, and $\xi'_n \in \check{\mathbb{D}}^{\epsilon}_{\mathcal{K};\boldsymbol{x}}[0,T]$ for each $n \ge 1$. In particular, due to $\boldsymbol{k} \neq \boldsymbol{0}$, we must have $\mathcal{K} \neq \boldsymbol{0}$. Note that $|\mathcal{K}| \stackrel{\text{def}}{=} \sum_{\boldsymbol{j}\in\mathcal{P}_d} \mathcal{K}_{\boldsymbol{j}}$ is the number of jumps for any path in $\check{\mathbb{D}}^{\epsilon}_{\mathcal{K};\boldsymbol{x}}[0,T]$. Since $\xi'_n \in \check{\mathbb{D}}^{\epsilon}_{\mathcal{K};\boldsymbol{x}}[0,T]$ for each n, by (28) we know that ξ'_n admits the form

$$\xi'_n(t) = t\boldsymbol{x} + \sum_{\boldsymbol{j}\in\mathcal{P}_d} \sum_{k=1}^{\mathcal{K}_{\boldsymbol{j}}} \boldsymbol{w}^{(n)}_{\boldsymbol{j},k} \mathbb{I}_{[t^{(n)}_{\boldsymbol{j},k},T]}(t), \qquad \forall t \in [0,T],$$

where each $\boldsymbol{w}^{(n)}_{\boldsymbol{j},k}$ belongs to the cone $\bar{\mathbb{R}}^d(\boldsymbol{j},\epsilon)$. Next, we fix some

$$\delta \in \left(0, \frac{\Delta}{2 + 4\|\boldsymbol{x}\| + 4|\mathcal{K}|}\right). \tag{82}$$

By (81), it holds for any n large enough that we have $\boldsymbol{d}^{[0,T]}_{J_1}(\xi_n, \xi'_n) < \delta$. In other words, for each n large enough there exists λ_n—a homeomorphism on $[0,T]$—such that

$$\sup_{t\in[0,T]} |\lambda_n(t) - t| \vee \left\|\xi'_n(\lambda_n(t)) - \xi_n(t)\right\| < \delta. \tag{83}$$

Under such large n, we consider $t \in (0,T)$ such that

$$t \notin [t^{(n)}_{\boldsymbol{j},k} - \delta, t^{(n)}_{\boldsymbol{j},k} + \delta], \qquad \forall \boldsymbol{j} \in \mathcal{P}_d,\ k \in [\mathcal{K}_{\boldsymbol{j}}]. \tag{84}$$

By (84), $\xi_n'(\cdot)$ does note make any jump over the interval $[t-\delta, t+\delta]\cap[0,T]$, meaning that, over this interval, the path $\xi_n'(\cdot)$ is a linear function with slope $\boldsymbol{x}$. Then, by (83), we get

$$\sup_{t_1,t_2\in[t-\delta,t+\delta]\cap[0,T]} \left\|\xi_n(t_1)-\xi_n'(t_2)\right\| \le \delta + \|\boldsymbol{x}\|\cdot 2\delta.$$

Therefore, by modifying $\lambda_n(\cdot)$ only over $[t-\delta,t]$, one can obtain $\hat{\lambda}_n^{(t)}$—a homeomorphism over $[0,t]$—such that

$$\sup_{u\in[0,t]} |\hat{\lambda}_n^{(t)}(u)-u| \vee \left\|\xi_n'\big(\hat{\lambda}_n^{(t)}(u)\big)-\xi_n(u)\right\| < \delta\cdot(1+2\|\boldsymbol{x}\|).$$

Applying this bound on $\boldsymbol{d}_{J_1}^{[0,t]}(\phi_t(\xi_n),\phi_t(\xi_n'))\wedge 1$ for any $t\in(0,T)$ satisfying (84), and applying the trivial upper bound 1 for any t where (84) doest not hold, we get (for all n large enough)

$$\begin{aligned}
&\int_0^T e^{-t}\cdot\left[\boldsymbol{d}_{J_1}^{[0,t]}\Big(\phi_t(\xi_n),\ \phi_t(\xi_n')\Big)\wedge 1\right]dt \qquad (85)\\
&\le \int_{t\in[0,T]:\ t\notin[t_{\boldsymbol{j},k}^{(n)}-\delta,t_{\boldsymbol{j},k}^{(n)}+\delta]\ \forall \boldsymbol{j}\in\mathcal{P}_d,\ k\in[\mathcal{K}_{\boldsymbol{j}}]} e^{-t}\cdot\delta\cdot(1+2\|\boldsymbol{x}\|)dt + |\mathcal{K}|\cdot 2\delta\\
&\le \delta\cdot\big(1+2\|\boldsymbol{x}\|+2|\mathcal{K}|\big) < \Delta/2 \qquad \text{by (82)}.
\end{aligned}$$

Furthermore, since $\xi_n\in\phi_t(B)$, for each $n\ge 1$ there is $\tilde{\xi}_n\in B$ such that $\tilde{\xi}_n(t)=\xi_n(t)\ \forall t\in[0,T]$. Also, by extending the path ξ_n' linearly with slope $\boldsymbol{x}$ over (T,∞), we obtain $\tilde{\xi}_n'\in\breve{\mathbb{D}}_{\mathcal{K};\boldsymbol{x}}^{\epsilon}[0,\infty)$ such that $\tilde{\xi}_n'(t)=\xi_n'(t)\ \forall t\in[0,T]$. Following from the inequality in display (85),

$$\begin{aligned}
\boldsymbol{d}_{J_1}^{[0,\infty)}(\tilde{\xi}_n,\tilde{\xi}_n') &< \frac{\Delta}{2}+\int_T^\infty e^{-t}\cdot\left[\boldsymbol{d}_{J_1}^{[0,t]}\Big(\phi_t(\tilde{\xi}_n),\ \phi_t(\tilde{\xi}_n')\Big)\wedge 1\right]dt\\
&< \Delta \qquad \text{by (80)},
\end{aligned}$$

which clearly contradicts $\boldsymbol{d}_{J_1}^{[0,\infty)}\big(B,\breve{\mathbb{D}}_{\leqslant\boldsymbol{k};\boldsymbol{x}}^{\epsilon}[0,\infty)\big)>\Delta$. This concludes the proof. □

Now, we are ready to prove Theorem D.3, which adapts Theorem 3.2 to $\big(\mathbb{D}[0,\infty),\boldsymbol{d}_{J_1}^{[0,\infty)}\big)$.

Theorem D.3. *Let Assumption 5 hold. Let $\boldsymbol{k}=(k_j)_{j\in[d]}\in\mathbb{Z}_+^d\setminus\{\boldsymbol{0}\}$, and let B be a Borel set of $\mathbb{D}[0,\infty)$ equipped with J_1 topology (i.e., induced by $\boldsymbol{d}_{J_1}^{[0,\infty)}$). Suppose that B is bounded away from $\breve{\mathbb{D}}_{\leqslant\boldsymbol{k};\boldsymbol{\mu_L}}^{\epsilon}[0,\infty)$ under $\boldsymbol{d}_{J_1}^{[0,\infty)}$ for some $\epsilon>0$. Then,*

$$\sum_{\mathcal{K}\in\mathbb{A}(\boldsymbol{k})}\breve{\mathbf{C}}_{\mathcal{K};\boldsymbol{\mu_L}}^{[0,\infty)}(B^\circ)\le\liminf_{n\to\infty}\frac{\mathbf{P}\big(\bar{\boldsymbol{L}}_n^{[0,\infty)}\in B\big)}{\breve{\lambda}_{\boldsymbol{k}}(n)}\le\limsup_{n\to\infty}\frac{\mathbf{P}\big(\bar{\boldsymbol{L}}_n^{[0,\infty)}\in B\big)}{\breve{\lambda}_{\boldsymbol{k}}(n)}\le\sum_{\mathcal{K}\in\mathbb{A}(\boldsymbol{k})}\breve{\mathbf{C}}_{\mathcal{K};\boldsymbol{\mu_L}}^{[0,\infty)}(B^-)<\infty, \qquad (86)$$

where $\breve{\lambda}_{\boldsymbol{k}}(n)$, $\boldsymbol{\mu_L}$, $\breve{\mathbf{C}}_{\mathcal{K};\boldsymbol{\mu_L}}^{[0,\infty)}$, and $\breve{\mathbb{D}}_{\leqslant\boldsymbol{k};\boldsymbol{\mu_L}}^{\epsilon}[0,\infty)$ are defined in (32), (24), (30), and (37), respectively.

Proof. Part (*ii*) of Lemma D.1 confirms that $\sum_{\mathcal{K}\in\mathbb{A}(\boldsymbol{k})}\breve{\mathbf{C}}_{\mathcal{K};\boldsymbol{\mu_L}}^{[0,\infty)}\in\mathbb{M}\big(\mathbb{D}[0,\infty)\setminus\breve{\mathbb{D}}_{\leqslant\boldsymbol{k};\boldsymbol{\mu_L}}^{\epsilon}[0,\infty)\big)$ (under $\boldsymbol{d}_{J_1}^{[0,\infty)}$) for any $\epsilon>0$, which verifies the finite upper bound in (86). Then, by Theorem 2.2 (the Portmanteau Theorem for $\mathbb{M}$-convergence), it suffices to prove the following claim: for any $f:\mathbb{D}[0,\infty)\to$

$[0,\infty)$ *uniformly continuous* (w.r.t. $(\mathbb{D}[0,\infty), \boldsymbol{d}_{J_1}^{[0,\infty)})$) and bounded (i.e., $\|f\| = \sup_{\xi\in\mathbb{D}[0,\infty)} f(\xi) < \infty$) such that $B \stackrel{\text{def}}{=} \text{supp}(f)$ is bounded away from $\breve{\mathbb{D}}^{\epsilon}_{\preccurlyeq \boldsymbol{k};\boldsymbol{\mu_L}}[0,\infty)$ under $\boldsymbol{d}_{J_1}^{[0,\infty)}$ for some $\epsilon > 0$, we have

$$\lim_{n\to\infty} \frac{\mathbf{E} f(\bar{\boldsymbol{L}}_n^{[0,\infty)})}{\breve{\lambda}_{\boldsymbol{k}}(n)} = \sum_{\mathcal{K}\in\mathbb{A}(\boldsymbol{k})} \breve{\mathbf{C}}^{[0,\infty)}_{\mathcal{K};\boldsymbol{\mu_L}}(f). \tag{87}$$

To proceed, for each $t \in [0,\infty)$ we define the mapping $f_t : \mathbb{D}[0,t] \to \mathbb{R}$ by (note that in the display (88) below, the path ξ is an element of $\mathbb{D}[0,t]$)

$$f_t(\xi) \stackrel{\text{def}}{=} f\big(\phi_t^{\text{inv}}(\xi)\big), \quad \text{where } \phi_t^{\text{inv}}(\xi)(s) \stackrel{\text{def}}{=} \xi(s\wedge t) + [0\vee(s-t)]\cdot \boldsymbol{\mu_L}\ \forall s \geq 0. \tag{88}$$

That is, ϕ_t^{inv} extends the path $\xi \in \mathbb{D}[0,t]$ to $\mathbb{D}[0,\infty)$ linearly with slope $\boldsymbol{\mu_L}$ over (t,∞), which, in a way, "inverts" the projection mapping ϕ_t in (43). Now, we make a few observations.

- Define $\hat{\phi}_t : \mathbb{D}[0,\infty) \to \mathbb{D}[0,\infty)$ by

  $$\hat{\phi}_t(\xi)(s) \stackrel{\text{def}}{=} \xi(t\wedge s) + [0\vee(s-t)]\cdot\boldsymbol{\mu_L}, \qquad \forall \xi\in\mathbb{D}[0,\infty),\ s\geq 0.$$

  By the definition of $\boldsymbol{d}_{J_1}^{[0,\infty)}$ in (76), for any $t > 0$ and $\xi \in \mathbb{D}[0,\infty)$, we have $\boldsymbol{d}_{J_1}^{[0,\infty)}\big(\hat{\phi}_t(\xi),\xi\big) \leq \int_{s>t} e^{-s}ds = e^{-t}$. Then, by the uniform continuity of the function f fixed in (87), given $\epsilon > 0$, there exists $T = T(\epsilon) > 0$ such that

  $$\big|f(\xi) - f\big(\hat{\phi}_t(\xi)\big)\big| < \epsilon, \qquad \forall t \geq T,\ \xi\in\mathbb{D}[0,\infty). \tag{89}$$

- Recall that $B = \text{supp}(f)$, and note that

  $$f(\xi) = 0 \text{ and } f\big(\hat{\phi}_t(\xi)\big) = 0, \qquad \forall t > 0,\ \xi\in\mathbb{D}[0,\infty) \text{ such that } \phi_t(\xi)\notin\phi_t(B). \tag{90}$$

- Combining (89) and (90), we know that given $\epsilon > 0$, there is $T = T(\epsilon) \in (0,\infty)$ such that

  $$\big|f(\xi) - f\big(\hat{\phi}_t(\xi)\big)\big| \leq \epsilon\cdot\mathbb{I}\big\{\phi_t(\xi)\in\phi_t(B)\big\}, \qquad \forall t\geq T,\ \xi\in\mathbb{D}[0,\infty). \tag{91}$$

- Recall that in (88), we use ξ to denote an element of $\mathbb{D}[0,t]$, which differs from our convention henceforth that $\xi \in \mathbb{D}[0,\infty)$. By definitions in (88), it holds for any $t > 0$ and $\xi\in\mathbb{D}[0,\infty)$ that $f_t\big(\phi_t(\xi)\big) = f\big(\hat{\phi}_t(\xi)\big)$. Then, by (91), given $\epsilon > 0$ there exists $T = T(\epsilon) > 0$ such that for any $t \geq T$,

  $$\limsup_{n\to\infty} \frac{\mathbf{E} f(\bar{\boldsymbol{L}}_n^{[0,\infty)})}{\breve{\lambda}_{\boldsymbol{k}}(n)} \leq \lim_{n\to\infty} \frac{\mathbf{E} f_t(\bar{\boldsymbol{L}}_n^{[0,t]})}{\breve{\lambda}_{\boldsymbol{k}}(n)} + \epsilon\cdot\limsup_{n\to\infty} \frac{\mathbf{P}(\bar{\boldsymbol{L}}_n^{[0,t]}\in\phi_t(B))}{\breve{\lambda}_{\boldsymbol{k}}(n)}, \tag{92}$$

  $$\liminf_{n\to\infty} \frac{\mathbf{E} f(\bar{\boldsymbol{L}}_n^{[0,\infty)})}{\breve{\lambda}_{\boldsymbol{k}}(n)} \geq \lim_{n\to\infty} \frac{\mathbf{E} f_t(\bar{\boldsymbol{L}}_n^{[0,t]})}{\breve{\lambda}_{\boldsymbol{k}}(n)} - \epsilon\cdot\limsup_{n\to\infty} \frac{\mathbf{P}(\bar{\boldsymbol{L}}_n^{[0,t]}\in\phi_t(B))}{\breve{\lambda}_{\boldsymbol{k}}(n)}. \tag{93}$$

Suppose we can show that

$$\lim_{t\to\infty}\lim_{n\to\infty} \frac{\mathbf{E} f_t(\bar{\boldsymbol{L}}_n^{[0,t]})}{\breve{\lambda}_{\boldsymbol{k}}(n)} = \sum_{\mathcal{K}\in\mathbb{A}(\boldsymbol{k})} \breve{\mathbf{C}}^{[0,\infty)}_{\mathcal{K};\boldsymbol{\mu_L}}(f), \tag{94}$$

$$\limsup_{t\to\infty}\limsup_{n\to\infty} \frac{\mathbf{P}(\bar{\boldsymbol{L}}_n^{[0,t]}\in\phi_t(B))}{\breve{\lambda}_{\boldsymbol{k}}(n)} < \infty. \tag{95}$$

Then, by plugging these results into (92)–(93) and sending $t \to \infty$, $\epsilon \downarrow 0$, we conclude the proof of claim (87). Now, it only remains to prove claims (94) and (95).

Proof of Claim (94). For any $t > 0$, the function f_t is continuous and bounded. Besides, note that $f_t(\xi) = 0$ whenever $\xi \notin \phi_t(B)$, where $B = \mathrm{supp}(f)$. On the other hand, Lemma D.2 shows that $\phi_t(B)$ is bounded away from $\breve{\mathbb{D}}^{\epsilon}_{\leqslant \boldsymbol{k};\boldsymbol{\mu_L}}[0,t]$ under $\boldsymbol{d}^{[0,t]}_{J_1}$ for any $t > 0$ large enough. This allows us to apply Theorem 3.2 and Theorem 2.2, the Portmanteau Theorem for $\mathbb{M}$-convergence, to show that

$$\lim_{n\to\infty} \frac{\mathbf{E} f_t(\bar{\boldsymbol{L}}^{[0,t]}_n)}{\breve{\lambda}_{\boldsymbol{k}}(n)} = \sum_{\mathcal{K}\in\mathbb{A}(\boldsymbol{k})} \breve{\mathbf{C}}^{[0,t]}_{\mathcal{K};\boldsymbol{\mu_L}}(f_t), \qquad \forall t > 0 \text{ sufficiently large.}$$

Also, note that $|\mathbb{A}(\boldsymbol{k})| < \infty$ (i.e., the set contains only finitely many elements). Therefore, to prove Claim (94), it suffices to fix some $\mathcal{K} = (\mathcal{K}_{\boldsymbol{j}})_{\boldsymbol{j}\in\mathcal{P}_d} \in \mathbb{A}(\boldsymbol{k})$ and show that $\breve{\mathbf{C}}^{[0,t]}_{\mathcal{K};\boldsymbol{\mu_L}}(f_t) = \breve{\mathbf{C}}^{[0,\infty)}_{\mathcal{K};\boldsymbol{\mu_L}}(f)$ for any $t > 0$ sufficiently large. Now, by the definition of f_t in (88) and $\breve{\mathbf{C}}^{[0,t]}_{\mathcal{K};\boldsymbol{\mu_L}}$ in (30),

$$\begin{aligned}
\breve{\mathbf{C}}^{[0,t]}_{\mathcal{K};\boldsymbol{\mu_L}}(f_t) &= \frac{1}{\prod_{\boldsymbol{j}\in\mathcal{P}_d}\mathcal{K}_{\boldsymbol{j}}!} \cdot \int f\left(\phi^{\mathrm{inv}}_t\left(\boldsymbol{\mu_L}\mathbf{1} + \sum_{\boldsymbol{j}\in\mathcal{P}_d}\sum_{k\in[\mathcal{K}_{\boldsymbol{j}}]} \boldsymbol{w}_{\boldsymbol{j},k}\mathbb{I}_{[t_{\boldsymbol{j},k},t]}\right)\right) \\
&\qquad \bigtimes_{\boldsymbol{j}\in\mathcal{P}_d}\bigtimes_{k\in[\mathcal{K}_{\boldsymbol{j}}]} \left((\mathbf{C}_{\boldsymbol{j}} \times \mathcal{L}_{(0,t)})\left(d(\boldsymbol{w}_{\boldsymbol{j},k}, t_{\boldsymbol{j},k})\right)\right) \\
&= \frac{1}{\prod_{\boldsymbol{j}\in\mathcal{P}_d}\mathcal{K}_{\boldsymbol{j}}!} \cdot \int f\left(\phi^{\mathrm{inv}}_t\left(\boldsymbol{\mu_L}\mathbf{1} + \sum_{\boldsymbol{j}\in\mathcal{P}_d}\sum_{k\in[\mathcal{K}_{\boldsymbol{j}}]} \boldsymbol{w}_{\boldsymbol{j},k}\mathbb{I}_{[t_{\boldsymbol{j},k},t]}\right)\right) \\
&\qquad \cdot \mathbb{I}\left\{t_{\boldsymbol{j},k} < t \ \forall \boldsymbol{j}\in\mathcal{P}_d,\ k\in[\mathcal{K}_{\boldsymbol{j}}]\right\} \bigtimes_{\boldsymbol{j}\in\mathcal{P}_d}\bigtimes_{k\in[\mathcal{K}_{\boldsymbol{j}}]} \left((\mathbf{C}_{\boldsymbol{j}} \times \mathcal{L}_{(0,\infty)})\left(d(\boldsymbol{w}_{\boldsymbol{j},k}, t_{\boldsymbol{j},k})\right)\right),
\end{aligned} \tag{96}$$

where $\mathcal{L}_I$ is the Lebesgue measure restricted on interval I. Furthermore, for $B = \mathrm{supp}(f)$, by part (i) of Lemma D.1, there exists some $T \in (0,\infty)$ such that the following claim holds: for any piece-wise linear function $\xi \in \mathbb{D}[0,\infty)$ of the form $\xi(t) = \boldsymbol{\mu_L} t + \sum_{\boldsymbol{j}\in\mathcal{P}_d}\sum_{k=1}^{\mathcal{K}_{\boldsymbol{j}}} \boldsymbol{w}_{\boldsymbol{j},k}\mathbb{I}_{[t_{\boldsymbol{j},k},\infty)}(t)$ with $\boldsymbol{w}_{\boldsymbol{j},k} \in \mathbb{R}^d(\boldsymbol{j})$ for each $\boldsymbol{j}\in\mathcal{P}_d,\ k\in[\mathcal{K}_{\boldsymbol{j}}]$, we have

$$\xi \in B \quad\Longrightarrow\quad t_{\boldsymbol{j},k} < T \ \forall \boldsymbol{j}\in\mathcal{P}_d,\ k\in[\mathcal{K}_{\boldsymbol{j}}]. \tag{97}$$

Therefore, provided that $\boldsymbol{w}_{\boldsymbol{j},k} \in \mathbb{R}^d(\boldsymbol{j})$ for each $\boldsymbol{j}\in\mathcal{P}_d$ and $k\in[\mathcal{K}_{\boldsymbol{j}}]$, given any $t > T$ we have

$$\begin{aligned}
&f\left(\phi^{\mathrm{inv}}_t\left(\boldsymbol{\mu_L}\mathbf{1} + \sum_{\boldsymbol{j}\in\mathcal{P}_d}\sum_{k\in[\mathcal{K}_{\boldsymbol{j}}]} \boldsymbol{w}_{\boldsymbol{j},k}\mathbb{I}_{[t_{\boldsymbol{j},k},t]}\right)\right) \cdot \mathbb{I}\left\{t_{\boldsymbol{j},k} < t \ \forall \boldsymbol{j}\in\mathcal{P}_d,\ k\in[\mathcal{K}_{\boldsymbol{j}}]\right\} \\
&= f\left(\boldsymbol{\mu_L}\mathbf{1} + \sum_{\boldsymbol{j}\in\mathcal{P}_d}\sum_{k\in[\mathcal{K}_{\boldsymbol{j}}]} \boldsymbol{w}_{\boldsymbol{j},k}\mathbb{I}_{[t_{\boldsymbol{j},k},\infty)}\right) \cdot \mathbb{I}\left\{t_{\boldsymbol{j},k} < t \ \forall \boldsymbol{j}\in\mathcal{P}_d,\ k\in[\mathcal{K}_{\boldsymbol{j}}]\right\} \quad \text{by (88)}, \\
&= f\left(\boldsymbol{\mu_L}\mathbf{1} + \sum_{\boldsymbol{j}\in\mathcal{P}_d}\sum_{k\in[\mathcal{K}_{\boldsymbol{j}}]} \boldsymbol{w}_{\boldsymbol{j},k}\mathbb{I}_{[t_{\boldsymbol{j},k},\infty)}\right) \qquad \text{due to (97) and } B = \mathrm{supp}(f).
\end{aligned}$$

Also, by the $\mathcal{MHRV}$ condition of Assumption 5, for each $\boldsymbol{j} \in \mathcal{P}_d$ the measure $\mathbf{C}_{\boldsymbol{j}}(\cdot)$ is supported on $\mathbb{R}^d(\boldsymbol{j})$ (see also Section 2.2). Then, in (96), it holds for any $t > T$ that

$$\begin{aligned}
&\check{\mathbf{C}}^{[0,t]}_{\mathcal{K};\boldsymbol{\mu_L}}(f_t) \\
&= \frac{1}{\prod_{\boldsymbol{j}\in\mathcal{P}_d}\mathcal{K}_{\boldsymbol{j}}!}\cdot\int f\left(\boldsymbol{\mu_L}\mathbf{1} + \sum_{\boldsymbol{j}\in\mathcal{P}_d}\sum_{k\in[\mathcal{K}_{\boldsymbol{j}}]}\boldsymbol{w}_{\boldsymbol{j},k}\mathbb{I}_{[t_{\boldsymbol{j},k},\infty)}\right)\bigtimes_{\boldsymbol{j}\in\mathcal{P}_d}\bigtimes_{k\in[\mathcal{K}_{\boldsymbol{j}}]}\Big((\mathbf{C}_{\boldsymbol{j}}\times\mathcal{L}_{(0,\infty)})\Big(d(\boldsymbol{w}_{\boldsymbol{j},k},t_{\boldsymbol{j},k})\Big)\Big) \\
&= \check{\mathbf{C}}^{[0,\infty)}_{\mathcal{K};\boldsymbol{\mu_L}}(f).
\end{aligned}$$

This concludes the proof of Claim (94).

Proof of Claim (95). By the definition of the projection mapping ϕ_t in (43), for any $t' > t > 0$ we have $\{\bar{\boldsymbol{L}}_n^{[0,t]} \in \phi_t(B)\} \supseteq \{\bar{\boldsymbol{L}}_n^{[0,t']} \in \phi_{t'}(B)\}$. This monotonicity w.r.t. t implies that, to prove Claim (95), it suffices to find $T \in (0,\infty)$ such that

$$\limsup_{n\to\infty}\frac{\mathbf{P}(\bar{\boldsymbol{L}}_n^{[0,T]} \in \phi_T(B))}{\check{\lambda}_{\boldsymbol{k}}(n)} < \infty. \tag{98}$$

Since $B = \mathrm{supp}(f)$ is bounded away from $\check{\mathbb{D}}^{\epsilon}_{\leqslant\boldsymbol{k};\boldsymbol{\mu_L}}[0,\infty)$ under $\boldsymbol{d}^{[0,\infty)}_{J_1}$ for some $\epsilon > 0$, Lemma D.2 shows that $\phi_T(B)$ is bounded away from $\check{\mathbb{D}}^{\epsilon}_{\leqslant\boldsymbol{k};\boldsymbol{\mu_L}}[0,T]$ under $\boldsymbol{d}^{[0,T]}_{J_1}$ for any $T > 0$ large enough. Claim (98) then follows from the finite upper bound in Claim (39) of Theorem 3.2. □

D.2. Proof of Proposition 3.4

To rigorously explain the coupling between the Hawkes process $\boldsymbol{N}(t)$ and the compound Poisson process $\boldsymbol{L}(t)$ in (52), we first review the cluster representation of Hawkes processes, which reveals the underlying branching structure in $\boldsymbol{N}(t)$. Intuitively speaking, the cluster representation shows that the Hawkes process $\boldsymbol{N}(t)$ can be constructed by first generating a sequence of "immigrants" (i.e., points of the 0^{th} generation) arriving at Poisson rates c_i and then, iteratively, letting each generation of points to give birth to the next generation of descendants. In particular, any type-j point gives birth to type-i points according to an inhomogeneous Poisson processes with rates determined by $\tilde{B}_{i\leftarrow j}f_{i\leftarrow j}(\cdot)$. This approach was first introduced in [4] in the univariate setting. See also Chapter 4 of [6] and Examples 6.3(c) and 6.4(c) of [3] for a treatment in the multivariate setting.

Appealing to the more general framework of Poisson cluster processes, we start by constructing a point process that identifies the *centers* of each cluster, and then augment each center with a separate point process representing the arrival times and types of the offspring in this cluster. First, we define a point process in the space $\boldsymbol{N}^C(\cdot)$ in $[0,\infty)\times[d]$, where the superscript C denotes "center". Recall the positive immigration rates c_i in (1). Independently for each $j \in [d]$, generate $0 < T^C_{i;1} < T^C_{i;2} < \ldots$ under a Poisson process with rate c_i. Let

$$\boldsymbol{N}^C(\cdot) \stackrel{\text{def}}{=} \sum_{j\in[d]}\sum_{k\geq 0}\boldsymbol{\delta}_{(T^C_{j;k},j)}(\cdot), \tag{99}$$

where $\boldsymbol{\delta}_x$ denotes the Dirac measure at x. This is equivalent to the superposition of several independent Poisson processes, where each arrival $T^C_{j;k}$ is augmented with a marker j to denote its type.

Next, we associate each point $(T^C_{j;k}, j)$ in the center process with an *offspring* process, which is a point process in the space $[0,\infty)\times[d]$ (we use the superscript O to denote "offspring"):

$$\boldsymbol{N}^O_{(T^C_{j;k},j)}(\cdot) = \sum_{m=0}^{K^O(T^C_{j;k},j)} \boldsymbol{\delta}_{(T^O_{j;k}(m),A^O_{j;k}(m))}(\cdot). \tag{100}$$

In particular, each $\boldsymbol{N}^O_{(T^C_{j;k},j)}(\cdot)$ is an independent copy of a point process C^O_j, whose law is described by the following iterative procedure.

Definition D.4 (Offspring Cluster Process C^O_j). For each $j\in[d]$, the point process $C^O_j(\cdot)$ in the space $[0,\infty)\times[d]$ is defined as follows.

(i) (**Ancestor Immigrant**) In this procedure, we use $R^{(n);j}_i$ to denote the number of type-i individuals in the n^{th} generation of the cluster C^O_j, and $T^{(n);j}_i(k)$ for the time between the arrival of the ancestor of the cluster and the birth of the k^{th} type-i individual in the n^{th} generation. Specifically, let $R^{(0);j}_j = 1$ and $R^{(0);j}_i = 0$ for each $i\in[d]$, $i\neq j$, which represents that the type-j ancestor is the only member in the 0^{th} generation of this cluster. Besides, we set $T^{(0);j}_j(1) = 0$.

(ii) (**Offspring in the n^{th} generation**) Iteratively, do the following for $n = 1,2,\ldots$, until we arrive at some n such that $R^{(n-1);j}_i = 0$ for any $i\in[d]$. For any $l\in[d]$ with $R^{(n-1);j}_l \geq 1$, and any $m\in[R^{(n-1);j}_l]$, independently for each $i\in[d]$, generate the sequence $\big(T^{(n,m);j}_{i\leftarrow l}(k)\big)_k$ according to an inhomogeneous Poisson process with rate

$$\tilde{B}_{i\leftarrow l} f_{i\leftarrow l}\big(\,\cdot - T^{(n-1);j}_l(m)\big). \tag{101}$$

Here, by sampling under the rate in (101), we mean the following.

- First, generate an independent copy of $\tilde{B}_{i\leftarrow l}$. Conditioning on $\tilde{B}_{i\leftarrow l} = b$ for some $b\geq 0$, generate $B^{(n,m);j}_{i\leftarrow l}$—the count of type-$i$ children (in the n^{th} generation) born by the m^{th} type-l individual in the $(n-1)^{\text{th}}$generation—by $B^{(n,m);j}_{i\leftarrow l}\sim \text{Poisson}\big(b\cdot\mu_{i\leftarrow l}\big)$, with $\mu_{i\leftarrow l}$ defined in (12).
- Next, let t_k be i.i.d. copies under the law with density $f_{i\leftarrow l}(\cdot)/\mu_{i\leftarrow l}$. Let $T^{(n,m);j}_{i\leftarrow l}(k) = t_k + T^{(n-1);j}_l(m)$ for each $k\in[B^{(n,m);j}_{i\leftarrow j}]$.

Then, given $i\in[d]$, by ordering the arrival times $T^{(n,m);j}_{i\leftarrow l}(k)$ across all $l\in[d]$, $m\in[R^{(n-1);j}_l]$, and $k\in[B^{(n,m);j}_{i\leftarrow l}]$, we get $0<T^{(n);j}_i(1)<T^{(n);j}_i(2)<\ldots<T^{(n);j}_i\big(R^{(n);j}_i\big)$, where we use $R^{(n);j}_i$ to denote the count of type-i individuals in the n^{th} generation. Note that this strictly increasing sequence is almost surely well-defined (i.e., with no ties) since the law of $T^{(n,m);j}_{i\leftarrow l}(k)$ is absolutely continuous w.r.t. the Lebesgue measure.

(iii) (**Definition of the cluster process**) Let $K^O_j \stackrel{\text{def}}{=} \max\{n\geq 0:\ R^{(n);j}_i\geq 1 \text{ for some } i\in[d]\}$ be the count of generations in this cluster. W.l.o.g. we set $R^{(n);j}_i = 0$ for any $n\geq K^O_j+1$ and $i\in[d]$. The cluster process is defined by

$$C^O_j(\cdot) \stackrel{\text{def}}{=} \sum_{n=0}^{K^O_j}\sum_{i\in[d]}\sum_{m=1}^{R^{(n);j}_i} \boldsymbol{\delta}_{(T^{(n);j}_i(m),i)}(\cdot). \tag{102}$$

In this paper, we work with conditions (see Section 2.3) which ensure that there are almost surely finitely many points in C_j^O, and hence the count of points $K^O(T_{j;k}^C, j)$ is almost surely finite in (100).

Define the time-shift operator $\boldsymbol{\theta}_t$ by $\boldsymbol{\theta}_t\mu(\cdot) \stackrel{\text{def}}{=} \sum_{k\geq 1}\boldsymbol{\delta}_{(t_k+t,a_k)}(\cdot)$ for any $t \geq 0$ and point process $\mu(\cdot) = \sum_{k\geq 1}\boldsymbol{\delta}_{(t_k,a_k)}$. Augmenting the center process $\boldsymbol{N}^C$ with each offspring cluster $\boldsymbol{N}^O_{(T_{j;k}^C,j)}$, we obtain a point process

$$\boldsymbol{N}(\cdot) \stackrel{\text{def}}{=} \sum_{j\in[d]}\sum_{k\geq 0}\boldsymbol{\theta}_{T_{j;k}^C}\boldsymbol{N}^O_{(T_{j;k}^C,j)}(\cdot) = \sum_{j\in[d]}\sum_{k\geq 0}\sum_{m=0}^{K^O(T_{j;k}^C,j)}\boldsymbol{\delta}_{(T_{j;k}^C+T_{j;k}^O(m),A_{j;k}^O(m))}(\cdot). \tag{103}$$

For each $t \geq 0$ and $i \in [d]$, we define the counting process

$$N_i(t) \stackrel{\text{def}}{=} \boldsymbol{N}\big([0,t]\times\{i\}\big) = \sum_{j\in[d]}\sum_{k\geq 0}\sum_{m=0}^{K^O(T_{j;k}^C,j)}\mathbb{I}\Big\{T_{j;k}^C + T_{j;k}^O(m) \leq t,\ A_{j;k}^O(m) = i\Big\}. \tag{104}$$

Under the sub-criticality condition regarding the offspring distributions (i.e., Assumption 1), it has been shown in [3,4] that the definitions in (99)–(104) using the cluster representation agree with Definition 1.1.

We formally define the size of the offspring cluster processes C_j^O. Using notations in Definition D.4, we define (for each $j \in [d]$)

$$S_{i\leftarrow j} \stackrel{\text{def}}{=} \sum_{n\geq 0}R_i^{(n);j}, \qquad \boldsymbol{S}_j \stackrel{\text{def}}{=} (S_{1\leftarrow j}, S_{2\leftarrow j}, \ldots, S_{d\leftarrow j})^\top. \tag{105}$$

That is, $\boldsymbol{S}_j$ is the size vector of the offspring cluster process $C_j^O(\cdot)$ induced by a type-j ancestor, with each element $S_{i\leftarrow j}$ representing the count of type-i points in the cluster. Under the law of $C_j^O(\cdot)$ specified above, the vectors $\boldsymbol{S}_j$ solve the distributional fixed-point equations (2), under the offspring distributions $(B_{i\leftarrow j})_{i\in[d]}$ stated in (11).

Meanwhile, using notations in (100), we denote the size vector of cluster $\boldsymbol{N}^O_{(T_{j;k}^C,j)}$ by

$$S_{i\leftarrow j}^{(k)} \stackrel{\text{def}}{=} \sum_{m\geq 0}\mathbb{I}\Big\{A_{j;k}^O(m) = i\Big\}, \qquad \boldsymbol{S}_j^{(k)} \stackrel{\text{def}}{=} \Big(S_{1\leftarrow j}^{(k)}, S_{2\leftarrow j}^{(k)}, \ldots, S_{d\leftarrow j}^{(k)}\Big)^\top. \tag{106}$$

By definition, $\big(\boldsymbol{S}_j^{(k)}\big)_{k\geq 1}$ are i.i.d. copies of the $\boldsymbol{S}_j$ defined in (105), and hence exhibit the $\mathcal{MHRV}^*$ tail asymptotics characterized in Theorem 2.6.

In the construction of $\boldsymbol{L}(t)$ in (52), let $\boldsymbol{S}_j^{(k)}$ be defined as in (106). Now, Proposition 3.4 is an immediate consequence of the construction of the compound Poisson process $\boldsymbol{L}(t)$ in (52) and the tail asymptotics of Hawkes process clusters established in Theorem 2.6.

Proof of Proposition 3.4. Note that $\boldsymbol{L}(t)$ is the superposition of a sequence of independent compound Poisson processes. That is, $\boldsymbol{L}(t) = \sum_{j\in[d]}\boldsymbol{L}_{\cdot\leftarrow j}(t)$, where

$$\boldsymbol{L}_{\cdot\leftarrow j}(t) \stackrel{\text{def}}{=} \sum_{k\geq 0}\boldsymbol{S}_j^{(k)}\mathbb{I}_{[T_{j;k}^C,\infty)}(t), \qquad \forall j \in [d],$$

the $(T^C_{j;k})_{k\geq 1}$ is a sequence generated by a Poisson process on $(0,\infty)$ with a constant rate c_j, and $(\boldsymbol{S}_j^{(k)})_{k\geq 1}$ are i.i.d. copies of $\boldsymbol{S}_j$. In other words, $\boldsymbol{L}_{\cdot\leftarrow j}(t)$ is a Lévy process with generating triplet $(\mathbf{0},\mathbf{0},\nu_j)$, where the Lévy measure is $\nu_j(\cdot) = c_j \cdot \mathbf{P}(\boldsymbol{S}_j \in \ \cdot\ )$. Therefore, Claim (53) follows from $\boldsymbol{L}(t) = \sum_{j\in[d]} \boldsymbol{L}_{\cdot\leftarrow j}(t)$. Next, by Theorem 2.6,

$$\nu_i \in \mathcal{MHRV}^*\left((\bar{\boldsymbol{s}}_j)_{j\in[d]},\ \big(\alpha(\boldsymbol{j})\big)_{\boldsymbol{j}\subseteq[d]},\ (\lambda_{\boldsymbol{j}})_{\boldsymbol{j}\in\mathcal{P}_d},\ \left(c_i\cdot \mathbf{C}_i^{\boldsymbol{j}}\right)_{\boldsymbol{j}\in\mathcal{P}_d}\right), \qquad \forall i\in[d].$$

Then, the $\mathcal{MHRV}^*$ tail condition of $\nu = \sum_{i\in[d]}\nu_i$ claimed in (54) follows from Definition 2.4 for $\mathcal{MHRV}^*$, as well as the definition of $\mathbf{C}_{\boldsymbol{j}} = \sum_{i\in[d]} c_i\cdot \mathbf{C}_i^{\boldsymbol{j}}$ in (45). □

D.3. Proof of Proposition 3.5

To prove Proposition 3.5, we prepare a few lemmas. First, we decompose $\boldsymbol{L}(t)$ in (52) into processes constructed by "large" clusters and "small" clusters separately. For each $\delta, T>0$ and $n\geq 1$, let

$$\bar{\boldsymbol{L}}_n^{>\delta}(t) \stackrel{\text{def}}{=} \frac{1}{n}\sum_{j\in[d]}\sum_{k\geq 0}\boldsymbol{S}_j^{(k)}\mathbb{I}\left\{\left\|\boldsymbol{S}_j^{(k)}\right\| > n\delta\right\}\cdot \mathbb{I}_{[T^C_{j;k},\infty)}(nt), \tag{107}$$

$$\bar{\boldsymbol{L}}_n^{\leqslant\delta}(t) \stackrel{\text{def}}{=} \bar{\boldsymbol{L}}_n(t) - \bar{\boldsymbol{L}}_n^{>\delta}(t) = \frac{1}{n}\sum_{j\in[d]}\sum_{k\geq 0}\boldsymbol{S}_j^{(k)}\mathbb{I}\left\{\left\|\boldsymbol{S}_j^{(k)}\right\| \leq n\delta\right\}\cdot \mathbb{I}_{[T^C_{j;k},\infty)}(nt). \tag{108}$$

We also define the (scaled) sample paths

$$\bar{\boldsymbol{L}}_n^{>\delta;[0,T]} \stackrel{\text{def}}{=} \left\{\bar{\boldsymbol{L}}_n^{>\delta}(t):\ t\in[0,T]\right\}, \qquad \bar{\boldsymbol{L}}_n^{>\delta;[0,\infty)} \stackrel{\text{def}}{=} \left\{\bar{\boldsymbol{L}}_n^{>\delta}(t):\ t\geq 0\right\}, \tag{109}$$

and adopt notations $\bar{\boldsymbol{L}}_n^{\leqslant\delta;[0,T]}$ and $\bar{\boldsymbol{L}}_n^{\leqslant\delta;[0,\infty)}$ analogously. Recall that $\mathbf{1}(t)=t$ is the linear function with slope 1, and that $\mathbf{E}\boldsymbol{L}(1) = \boldsymbol{\mu}_{\boldsymbol{N}}$ (see (47)). The next lemma establishes the asymptotic equivalence between $\bar{\boldsymbol{L}}_n^{[0,\infty)}$ and $\bar{\boldsymbol{L}}_n^{>\delta;[0,\infty)} + \boldsymbol{\mu}_{\boldsymbol{N}}\mathbf{1}$ under $\boldsymbol{d}_{\mathcal{P}}^{[0,\infty)}$.

Lemma D.5. *Let Assumptions 1–4 hold. For any $\Delta,\beta>0$,*

$$\lim_{n\to\infty} n^{\beta}\cdot\mathbf{P}\left(\boldsymbol{d}_{\mathcal{P}}^{[0,\infty)}\left(\bar{\boldsymbol{L}}_n^{[0,\infty)},\ \bar{\boldsymbol{L}}_n^{>\delta;[0,\infty)} + \boldsymbol{\mu}_{\boldsymbol{N}}\mathbf{1}\right) > \Delta\right) = 0, \qquad \forall \delta>0 \text{ small enough.}$$

Proof. Fix $T>0$ large enough such that $e^{-T} < \Delta/2$. Note that

$$\int_{t>T} e^{-t}\cdot\left[\boldsymbol{d}_{\mathcal{P}}^{[0,t]}\left(\phi_t(\xi^{(1)}),\ \phi_t(\xi^{(2)})\right)\wedge 1\right]dt \leq e^{-T} < \Delta/2, \qquad \forall \xi^{(1)},\xi^{(2)}\in\mathbb{D}[0,\infty).$$

Besides , for each $t\in[0,T]$, the metric $\boldsymbol{d}_{\mathcal{P}}^{[0,t]}$ is bounded by the uniform metric on $\mathbb{D}[0,t]$. Therefore, it suffices to show that

$$\mathbf{P}\left(\sup_{t\in[0,T]}\left\|\bar{\boldsymbol{L}}_n(t) - \left(\bar{\boldsymbol{L}}_n^{>\delta}(t) + \boldsymbol{\mu}_{\boldsymbol{N}}t\right)\right\| > \Delta/2\right) = o(n^{-\beta}), \qquad \forall\delta>0 \text{ small enough.} \tag{110}$$

To proceed, let (for each $k \geq 1$)

$$\tau_n^{>\delta}(k) \stackrel{\text{def}}{=} \left\{t > \tau_n^{>\delta}(k-1):\ \Delta\bar{\boldsymbol{L}}^{>\delta}(t) \neq 0\right\} = \left\{t > \tau_n^{>\delta}(k-1):\ \left\|\Delta\bar{\boldsymbol{L}}(t)\right\| > \delta\right\} \tag{111}$$

be the arrival time of the k^{th} large jump in $\bar{\boldsymbol{L}}_n(t)$, under the convention that $\tau_n^{>\delta}(0) = 0$. Meanwhile, for any non-negative integer K, on the event $\left\{\tau_n^{>\delta}(K+1) > T\right\}$ (meaning that there are at most K large jumps in $\bar{\boldsymbol{L}}_n(t)$ during $t \in [0,T]$), note that

$$\begin{aligned}
&\sup_{t\in[0,T]} \left\|\bar{\boldsymbol{L}}_n(t) - \left(\bar{\boldsymbol{L}}_n^{>\delta}(t) + \boldsymbol{\mu}_{\boldsymbol{N}} t\right)\right\| \\
&= \sup_{t\in[0,T]} \left\|\bar{\boldsymbol{L}}^{\leqslant\delta}(t) - \mu_N t\right\| \quad \text{see (108)} \\
&\leq \sum_{k=1}^{K+1} \underbrace{\sup_{t\in[\tau_n^{>\delta}(k-1),\,\tau_n^{>\delta}(k))\cap[0,T]} \left\|\bar{\boldsymbol{L}}_n^{\leqslant\delta}(t) - \boldsymbol{\mu}_{\boldsymbol{N}}\cdot\left(t - \tau_n^{>\delta}(k-1)\right) - \bar{\boldsymbol{L}}_n^{\leqslant\delta}\left(\tau_n^{>\delta}(k-1)\right)\right\|}_{\stackrel{\text{def}}{=} R_k(\delta)}.
\end{aligned}$$

In the last line of the display above, we applied $\tau_n^{>\delta}(K+1) > T$, as well as the fact that $\bar{\boldsymbol{L}}_n^{\leqslant\delta}(t)$ makes no jumps at any $t = \tau_n^{>\delta}(k)$; see the definitions in (108). Therefore, to prove Claim (110), it suffices to show the existence of some $K \in \mathbb{N}$ such that

$$\mathbf{P}\left(\tau_n^{>\delta}(K+1) \leq T\right) = o(n^{-\beta}), \qquad \forall\delta > 0 \text{ sufficiently small}, \tag{112}$$

and that for each $k \in [K+1]$,

$$\mathbf{P}\left(R_k(\delta) > \frac{\Delta}{2(K+1)}\right) = o(n^{-\beta}), \qquad \forall\delta > 0 \text{ sufficiently small}. \tag{113}$$

Proof of Claim (112). We prove that, for K large enough, the claim holds for any $\delta > 0$. In light of Proposition 3.4 and the definitions in (107)–(108), $\bar{\boldsymbol{L}}^{>\delta}(t)$ is a Lévy process with generating triplet $\left(\mathbf{0}, \mathbf{0}, n\cdot\nu_n(\ \cdot\ \cap\{\boldsymbol{x}\in\mathbb{R}_+^d:\ \|\boldsymbol{x}\|>\delta\})\right)$, where $\nu_n(A) = \nu\{n\boldsymbol{x}:\ \boldsymbol{x}\in A\}$, and ν is defined in (53), satisfying the $\mathcal{MHRV}^*$ tail condition in (54). Next, let $\alpha^*(\cdot)$ and $\alpha(\cdot)$ be defined as in (15)–(16), and let $j^* \stackrel{\text{def}}{=} \arg\min_{j\in[d]} \boldsymbol{c}(\{j\}) = \arg\min_{j\in[d]} \alpha^*(j)$. Note that this argument minimum is unique and $\alpha^*(j^*) > 1$ under Assumptions 2 and 4. As a result, $\bar{\mathbb{R}}^d_{\leqslant}(\{j^*\},\epsilon) = \{\mathbf{0}\}$ for any $\epsilon > 0$. By (9),

$$\limsup_{n\to\infty} \frac{n\nu_n\{\boldsymbol{x}\in\mathbb{R}_+^d:\ \|\boldsymbol{x}\|>\delta\}}{n\lambda_{\{j^*\}}(n)} < \infty, \qquad \forall\delta > 0, \tag{114}$$

where $\lambda_j(n)$ is defined in (17). This implies $n\nu_n\{\boldsymbol{x}\in\mathbb{R}_+^d:\ \|\boldsymbol{x}\|>\delta\} = O\left(n\lambda_{\{j^*\}}(n)\right)$ as $n\to\infty$. Then, for any $K\in\mathbb{N}$,

$$\begin{aligned}
\mathbf{P}\left(\tau_n^{>\delta}(K+1) \leq T\right) &= \mathbf{P}\left(\text{Poisson}\left(n\nu_n\{\boldsymbol{x}\in\mathbb{R}_+^d:\ \|\boldsymbol{x}\|>\delta\}\cdot T\right) \geq K+1\right) \\
&\leq T^{K+1}\cdot\left(n\nu_n\{\boldsymbol{x}\in\mathbb{R}_+^d:\ \|\boldsymbol{x}\|>\delta\}\right)^{K+1} \quad \text{due to } \mathbf{P}\left(\text{Poisson}(\lambda)\geq k\right) \leq \lambda^k
\end{aligned}$$

$$= O\Big(\big(n\lambda_{\{j^*\}}(n)\big)^{K+1}\Big) \qquad \text{by (114)}.$$

Lastly, due to $n\lambda_{\{j^*\}}(n) \in \mathcal{RV}_{-(\alpha^*(j^*)-1)}(n)$ (see (17)) and $\alpha^*(j^*) > 1$, it holds for all K large enough that $(K+1)\cdot\big(\alpha^*(j^*)-1\big) > \beta$, and hence $\big(n\lambda_{\{j^*\}}(n)\big)^{K+1} = o(n^{-\beta})$. This concludes the proof of Claim (112).

Proof of Claim (113). By the Markov property at each stopping time $\tau_n^{>\delta}(k)$, it suffices to show that (for any $\delta > 0$ small enough),

$$\begin{aligned}
&\mathbf{P}\left(\sup_{t\in[0,T]:\, t<\tau_n^{>\delta}(1)} \left\|\bar{\boldsymbol{L}}_n^{\leqslant\delta}(t) - \boldsymbol{\mu}_{\boldsymbol{N}} t\right\| > \frac{\Delta}{2(K+1)}\right)\\
&= \mathbf{P}\left(\sup_{t\in[0,T]:\, t<\tau_n^{>\delta}(1)} \left\|\bar{\boldsymbol{L}}_n(t) - \boldsymbol{\mu}_{\boldsymbol{N}} t\right\| > \frac{\Delta}{2(K+1)}\right) = o(n^{-\beta}),
\end{aligned}$$

where we applied the fact that $\bar{\boldsymbol{L}}_n^{\leqslant\delta}(t) = \bar{\boldsymbol{L}}_n(t)$ for any $t < \tau_n^{>\delta}(1)$; see (107)–(108). Besides, recall that $\mathbf{E}\boldsymbol{L}(1) = \boldsymbol{\mu}_{\boldsymbol{N}}$. Applying Lemma C.4, we conclude the proof of claim (113). □

Analogous to Lemma D.5, we decompose the Hawkes process $\boldsymbol{N}(t)$ by considering the small or large clusters therein. More precisely, for the scaled process $\bar{\boldsymbol{N}}_n(t) = \boldsymbol{N}(nt)/n$, it follows from the cluster representation (104) that

$$\bar{N}_{n,i}(t) = \frac{1}{n}\sum_{j\in[d]}\sum_{k\geq 0}\sum_{m=0}^{K^O(T_j^C(k),j)} \mathbb{I}\Big\{T_{j;k}^C + T_{j;k}^O(m) \leq nt,\ A_{j;k}^O(m) = i\Big\}, \quad \forall t \geq 0,\ i \in [d],$$

and $\bar{\boldsymbol{N}}_n(t) = \big(\bar{N}_{n,1}(t), \bar{N}_{n,2}(t), \ldots, \bar{N}_{n,d}(t)\big)^\top$. Meanwhile, recall that $\boldsymbol{S}_j^{(k)}$ is size vector of the cluster induced by the k^{th} type-j immigrant (see (106)) and $T_{j;k}^C$ is the arrival time of that immigrant. By incorporating the cluster size, for each $n \geq 1$ and $\delta > 0$ we define

$$\bar{N}_{n,i}^{>\delta}(t) \stackrel{\text{def}}{=} \frac{1}{n}\sum_{j\in[d]}\sum_{k\geq 0}\sum_{m=0}^{K^O(T_j^C(k),j)} \mathbb{I}\Big\{\big\|\boldsymbol{S}_j^{(k)}\big\| > n\delta\Big\}\cdot\mathbb{I}\Big\{T_{j;k}^C + T_{j;k}^O(m) \leq nt,\ A_{j;k}^O(m) = i\Big\}, \tag{115}$$

$$\begin{aligned}
\bar{N}_{n,i}^{\leqslant\delta}(t) &\stackrel{\text{def}}{=} \bar{N}_{n,i}(t) - \bar{N}_{n,i}^{>\delta}(t) \qquad (116)\\
&= \frac{1}{n}\sum_{j\in[d]}\sum_{k\geq 0}\sum_{m=0}^{K^O(T_j^C(k),j)} \mathbb{I}\Big\{\big\|\boldsymbol{S}_j^{(k)}\big\| \leq n\delta\Big\}\cdot\mathbb{I}\Big\{T_{j;k}^C + T_{j;k}^O(m) \leq nt,\ A_{j;k}^O(m) = i\Big\},
\end{aligned}$$

and $\bar{\boldsymbol{N}}_n^{>\delta}(t) \stackrel{\text{def}}{=} \big(\bar{N}_{n,1}^{>\delta}(t), \ldots, \bar{N}_{n,d}^{>\delta}(t)\big)^\top$, $\bar{\boldsymbol{N}}_n^{\leqslant\delta}(t) \stackrel{\text{def}}{=} \big(\bar{N}_{n,1}^{\leqslant\delta}(t), \ldots, \bar{N}_{n,d}^{\leqslant\delta}(t)\big)^\top$. Analogous to (109), we define

$$\bar{\boldsymbol{N}}_n^{>\delta;[0,T]} \stackrel{\text{def}}{=} \big\{\bar{\boldsymbol{N}}_n^{>\delta}(t):\ t\in[0,T]\big\}, \qquad \bar{\boldsymbol{N}}_n^{>\delta;[0,\infty)} \stackrel{\text{def}}{=} \big\{\bar{\boldsymbol{N}}_n^{>\delta}(t):\ t\geq 0\big\}. \tag{117}$$

The proof of the next lemma involves the lifetime of each cluster. Adopting notations in (100), we use

$$H_j^{(k)} \stackrel{\text{def}}{=} \max_{m\geq 0} T_{j;k}^O(m) \tag{118}$$

to denote the lifetime of the cluster process $\boldsymbol{N}^{O}_{(T^{C}_{j}(k),j)}$ (i.e., the gap in the birth times between the ancestor and the last descendant in this cluster). Similarly, in (102) we use

$$H_j \stackrel{\text{def}}{=} \max_{m\geq 0,\ i\in[d]} T_i^{(n);j}(m) \tag{119}$$

to denote the lifetime of the point process C^{O}_{j}. Since $\boldsymbol{N}^{O}_{(T^{C}_{j}(k),j)}$ are i.i.d. copies of C^{O}_{j}, the sequence $H^{(k)}_{j}$ are also independent copies of H_j. Lemma D.6 establishes the asymptotic equivalence between $\bar{\boldsymbol{N}}^{[0,\infty)}_{n}$ and $\bar{\boldsymbol{N}}^{>\delta;[0,\infty)}_{n} + \boldsymbol{\mu}_{\boldsymbol{N}}\mathbf{1}$ under $\boldsymbol{d}^{[0,\infty)}_{\mathcal{P}}$.

Lemma D.6. *Let Assumptions 1–4 hold. For any $\Delta, \beta > 0$,*

$$\lim_{n\to\infty} n^{\beta}\cdot \mathbf{P}\left(\boldsymbol{d}^{[0,\infty)}_{\mathcal{P}}\left(\bar{\boldsymbol{N}}^{[0,\infty)}_{n},\ \bar{\boldsymbol{N}}^{>\delta;[0,\infty)}_{n} + \boldsymbol{\mu}_{\boldsymbol{N}}\mathbf{1}\right) > \Delta\right) = 0, \qquad \forall \delta > 0 \text{ small enough.}$$

Proof. The proof is similar to that of Lemma D.5. In particular, by fixing T large enough such that $e^{-T} < \Delta/2$, it suffices to prove that

$$\mathbf{P}\left(\sup_{t\in[0,T]}\left\|\bar{\boldsymbol{N}}_n(t) - \left(\bar{\boldsymbol{N}}^{>\delta}_{n}(t) + \boldsymbol{\mu}_{\boldsymbol{N}}t\right)\right\| > \Delta/2\right) = o(n^{-\beta}), \qquad \forall \delta > 0 \text{ small enough.} \tag{120}$$

First, by comparing the definitions in (108) and (116),

$$\bar{\boldsymbol{N}}^{\leqslant\delta}_{n}(t) \leq \bar{\boldsymbol{L}}^{\leqslant\delta}_{n}(t), \qquad \forall \delta > 0,\ n \geq 1,\ t \geq 0, \tag{121}$$

since in $\bar{\boldsymbol{L}}^{\leqslant\delta}_{n}(t)$ we lump all descendants in the same cluster at the arrival time of the immigrant of the cluster. Here, the order $\boldsymbol{x} \leq \boldsymbol{y}$ between two vectors in $\mathbb{R}^d$ means that $x_i \leq y_i$ for each $i \in [d]$. Next, for each $M > 0$, let

$$\boldsymbol{L}^{|M}(t) \stackrel{\text{def}}{=} \sum_{j\in[d]}\sum_{k\geq 0} \boldsymbol{S}^{(k)}_{j}\mathbb{I}\left\{H^{(k)}_{j} \leq M\right\}\cdot \mathbb{I}_{[T^{C}_{j;k},\infty)}(t).$$

That is, $\boldsymbol{L}^{|M}(t)$ is a modification of $\boldsymbol{L}(t)$ in (52) by only considering clusters with lifetime below M. Note that

$$\mathbf{E}\left[\boldsymbol{L}^{|M}(1)\right] = \sum_{j\in[d]} c_j\cdot\mathbf{E}\left[\boldsymbol{S}_j\mathbb{I}\{H_j \leq M\}\right] \to \sum_{j\in[d]} c_j\cdot\mathbf{E}\boldsymbol{S}_j = \boldsymbol{\mu}_{\boldsymbol{N}}, \qquad \text{as } M\to\infty, \tag{122}$$

by the monotone convergence theorem. Besides, for each $n \geq 1$ and $\delta > 0$, we define processes

$$\bar{\boldsymbol{L}}^{|M}_{n}(t) \stackrel{\text{def}}{=} \frac{1}{n}\sum_{j\in[d]}\sum_{k\geq 0} \boldsymbol{S}^{(k)}_{j}\mathbb{I}\left\{H^{(k)}_{j} \leq M\right\}\cdot \mathbb{I}_{[T^{C}_{j;k},\infty)}(nt),$$

$$\bar{\boldsymbol{L}}^{\leqslant\delta|M}_{n}(t) \stackrel{\text{def}}{=} \frac{1}{n}\sum_{j\in[d]}\sum_{k\geq 0} \boldsymbol{S}^{(k)}_{j}\mathbb{I}\left\{\left\|\boldsymbol{S}^{(k)}_{j}\right\| \leqslant n\delta,\ H^{(k)}_{j} \leq M\right\}\cdot \mathbb{I}_{[T^{C}_{j;k},\infty)}(nt),$$

which scale both time and space by n. For any $\epsilon, \delta > 0$, and any n large enough such that $n\epsilon > M$,

$$\bar{\boldsymbol{N}}^{\leqslant\delta}_{n}(t) \geq \bar{\boldsymbol{L}}^{\leqslant\delta|M}_{n}\big((t-\epsilon)\vee 0\big), \qquad \forall \delta > 0,\ n \geq 1,\ t \geq 0. \tag{123}$$

Indeed, for any cluster with its immigrant arriving at $T^C_{j;k} \le n(t-\epsilon)$ and lifetime $H^{(k)}_j \le M$, all the descendants in this cluster must have arrived by the time $n(t-\epsilon)+M < nt$. Combining (121) and (123), for any $\epsilon, M, \delta > 0$, it holds for all n large enough with $n\epsilon > M$ that

$$\begin{aligned}
&\bar{\boldsymbol{L}}_n^{\leqslant \delta|M}\big((t-\epsilon)\vee 0\big) - \boldsymbol{\mu}_{\boldsymbol{N}} t \le \bar{\boldsymbol{N}}_n^{\leqslant \delta}(t) - \boldsymbol{\mu}_{\boldsymbol{N}} t \le \bar{\boldsymbol{L}}_n^{\leqslant \delta}(t) - \boldsymbol{\mu}_{\boldsymbol{N}} t, \qquad \forall t \ge 0\\
\Longrightarrow \sup_{t\in[0,T]} &\left\|\bar{\boldsymbol{N}}_n^{\leqslant \delta}(t) - \boldsymbol{\mu}_{\boldsymbol{N}} t\right\| \le \sup_{t\in[0,T]} \left\|\bar{\boldsymbol{L}}_n^{\leqslant \delta}(t) - \boldsymbol{\mu}_{\boldsymbol{N}} t\right\| + \sup_{t\in[0,T]} \left\|\bar{\boldsymbol{L}}_n^{\leqslant \delta|M}\big((t-\epsilon)\vee 0\big) - \mu_{\boldsymbol{N}} t\right\|.
\end{aligned}$$

The last step follows from our choice of the L_1 norm and the preliminary bound that $y \le x \le z \implies |x| \le |y| + |z|$. Furthermore, note that

$$\begin{aligned}
&\sup_{t\in[0,T]} \left\|\bar{\boldsymbol{L}}_n^{\leqslant \delta|M}\big((t-\epsilon)\vee 0\big) - \boldsymbol{\mu}_{\boldsymbol{N}} t\right\|\\
&\le \sup_{t\in[0,T]} \left\|\bar{\boldsymbol{L}}_n^{\leqslant \delta|M}(t) - t\cdot \mathbf{E}\big[\boldsymbol{L}^{|M}(1)\big]\right\| + \epsilon\left\|\boldsymbol{\mu}_{\boldsymbol{N}}\right\| + T\cdot\left\|\boldsymbol{\mu}_{\boldsymbol{N}} - \mathbf{E}\big[\boldsymbol{L}^{|M}(1)\big]\right\|\\
&\le \sup_{t\in[0,T]} \left\|\bar{\boldsymbol{L}}_n^{\leqslant \delta|M}(t) - t\cdot \mathbf{E}\big[\boldsymbol{L}^{|M}(1)\big]\right\| + \Delta/6 \qquad \text{for any } \epsilon \text{ small enough and } M \text{ large enough.}
\end{aligned}$$

The last inequality follows from the limit in (122). Therefore, to prove Claim (120), it suffices to show that given $M > 0$, the bounds $\mathbf{P}\Big(\sup_{t\in[0,T]} \left\|\bar{\boldsymbol{L}}_n^{\leqslant \delta|M}(t) - t\cdot \mathbf{E}\big[\boldsymbol{L}^{|M}(1)\big]\right\| > \Delta/6\Big) = o(n^{-\beta})$ and $\mathbf{P}\Big(\sup_{t\in[0,T]} \left\|\bar{\boldsymbol{L}}_n^{\leqslant \delta}(t) - \boldsymbol{\mu}_{\boldsymbol{N}} t\right\| > \Delta/6\Big) = o(n^{-\beta})$ for any $\delta > 0$ small enough. By repeating the arguments in Lemma D.5 for (110), one can establish both claims above. □

To proceed, we prepare Lemma D.7 and characterize the tail asymptotics for H_j in (119) (i.e., the lifetime of clusters).

Lemma D.7. *Let Assumptions 1 and 2 hold. Let $\beta > 0$ and $a : (0,\infty) \to (0,\infty)$ be a regularly varying function with $a(x) \in \mathcal{RV}_{-\beta}(x)$ as $x \to \infty$. Let $f_{p\leftarrow q}(\cdot)$ be the decay functions in (1). Suppose that*

$$\int_{x/\log x}^{\infty} f_{p\leftarrow q}(t)dt = o\big(a(x)\big) \ \text{ as } x\to\infty, \qquad \forall p,q\in[d]. \tag{124}$$

Then,

$$\mathbf{P}(H_j > n\epsilon) = o\big(a(n)\big) \ \text{ as } n\to\infty, \qquad \forall j\in[d],\ \epsilon>0. \tag{125}$$

Proof. In this proof, we fix some $\epsilon > 0$ and $j \in [d]$. Recall the cluster representation of Hawkes processes in (99)–(104). The key of this proof is to establish a stochastic comparison of H_j, the lifetime of a cluster C^O_j, which refines the bounds for the lifetime of clusters in [8] by taking the maximum instead of the sum of the birth times within each generation. To be more specific, we introduce a few notations. Independently for each pair $(i,l) \in [d]^2$, let $\big(B^{(n,m);j}_{i\leftarrow l}\big)_{n,m\ge 1}$ be the i.i.d. copies of $B_{i\leftarrow l}$ in (11), i.e., with law

$$B_{i\leftarrow l} \sim \text{Poisson}\left(\tilde{B}_{i\leftarrow l}\int_0^{\infty} f_{i\leftarrow l}(t)dt\right).$$

Here, recall that by Poisson(X) for a non-negative variable X, we mean the law $\mathbf{P}\big(\text{Poisson}(X) > y\big) = \int_0^\infty \mathbf{P}\big(\text{Poisson}(x) > y\big)\mathbf{P}(X \in dx)$. This agrees with the notations in Definition D.4, where we use $B_{i\leftarrow l}^{(n,m);j}$ to denote, in the cluster process $C_j^O(\cdot)$, the count of type-i children (in the n^{th} generation) born by the m^{th} type-l individual in the $(n-1)^{\text{th}}$ generation. Besides, independent from the sequence $\big(B_{i\leftarrow l}^{(n,m);j}\big)_{n,m\geq 1}$, let $\big(X_{i\leftarrow l}^{(n,k,k')}\big)_{n,k,k'\geq 1}$ be i.i.d. copies of $X_{i\leftarrow l}$ with law

$$\mathbf{P}(X_{i\leftarrow l} \in E) = \frac{\int_{t\in E} f_{i\leftarrow l}(t)dt}{\int_0^\infty f_{i\leftarrow l}(t)dt}, \qquad \forall \text{ Borel } E \subseteq (0,\infty). \tag{126}$$

By step (ii) in Definition D.4, we interpret $X_{i\leftarrow l}^{(n,k,k')}$ as the time that the k^{th} type-l parent in the $(n-1)^{\text{th}}$ generation waited to give birth to its $(k')^{\text{th}}$ type-i child (here, the children are not ordered by age).

As in Definition D.4, we use $R_i^{(n);j}$ to denote, in the cluster C_j^O, the count of type-i descendants in the $(n-1)^{\text{th}}$ generation, and use the sequence $0 < T_i^{(n);j}(1) < T_i^{(n);j}(2) < \dots$ to denote the birth times of type-i descendants in the n^{th} generation (provided that the cluster does have at least n generations and there are type-i individuals in the n^{th} generation). In particular, note that each $T_i^{(n);j}(r)$ admits the expression

$$T_i^{(n);j}(r) = X_{j_1\leftarrow j}^{(1,1,k_1')} + X_{j_2\leftarrow j_1}^{(2,k_2,k_2')} + \dots + X_{i\leftarrow j_{n-1}}^{(n,k_n,k_n')},$$

where, for the (random) indices, we have $j_l \in [d]$, $k_l \in \big[R_{j_{l-1}}^{(l-1);j}\big]$, and $k_l' \in \big[B_{j_l\leftarrow j_{l-1}}^{(n,k_{l-1});j}\big]$ for each l; here, note that we must have $j_0 \equiv j$ and $k_1 \equiv 1$, since the only individual in the 0^{th} generation is the type-j ancestor itself. Next, note that

$$\begin{aligned}
T_i^{(n);j}(r) &\leq X_{j_1\leftarrow j}^{(1,1,k_1')} + X_{j_2\leftarrow j_1}^{(2,k_2,k_2')} + \dots + X_{j_{n-1}\leftarrow j_{n-2}}^{(n-1,k_{n-1},k_{n-1}')} + \max_{p,q\in[d]} \max_{\substack{k\in[R_p^{(n-1);j}]\\ k'\in[B_{q\leftarrow p}^{(n,k);j}]}} X_{q\leftarrow p}^{(n,k,k')} \\
&\underset{\text{s.t.}}{\leq} X_{j_1\leftarrow j}^{(1,1,k_1')} + X_{j_2\leftarrow j_1}^{(2,k_2,k_2')} + \dots + X_{j_{n-1}\leftarrow j_{n-2}}^{(n-1,k_{n-1},k_{n-1}')} + \max_{p,q\in[d]} \max_{m\in[R_q^{(n);j}]} X_{q\leftarrow p}^{(m)} \\
&\leq X_{j_1\leftarrow j}^{(1,1,k_1')} + X_{j_2\leftarrow j_1}^{(2,k_2,k_2')} + \dots + X_{j_{n-1}\leftarrow j_{n-2}}^{(n-1,k_{n-1},k_{n-1}')} + \max_{p,q\in[d]} \max_{m\leq\sum_{l\in[d]} R_l^{(n);j}} X_{q\leftarrow p}^{(m)}.
\end{aligned}$$

In the display above, the stochastic comparison (i.e., the $\underset{\text{s.t.}}{\leq}$ step) holds due to the independence between the $X_{q\leftarrow p}^{(n,k,k')}$'s and $\big(X_{u\leftarrow v}^{(t,k,k')}\big)_{u,v\in[d],k\geq 1,k'\geq 1,t\in[n-1]}$, where we use $X_{q\leftarrow p}^{(m)}$ to denote generic copies of $X_{q\leftarrow p}$ (see (126)) that are independent of the $X_{u\leftarrow v}^{(t,k,k')}$'s. We emphasize that this stochastic comparison holds due to the *independence between the size (i.e., offspring counts) and the birth time distributions* in the cluster. Repeating this argument inductively, we obtain

$$T_i^{(n);j}(r) \underset{\text{s.t.}}{\leq} n \cdot \max_{p,q\in[d]} \max_{1\leq m\leq \|S_j\|} X_{q\leftarrow p}^{(m)}. \tag{127}$$

We note that (127) refines the bounds for the lifetime of clusters in [8].

Adopting the notations in Definition D.4, we let $K_j^O = \max\{n \geq 0 : \ R_i^{(n);j} \geq 1 \text{ for some } i \in [d]\}$ be the total count of generations in this cluster. Besides, we use $R^{(n);j} \overset{\text{def}}{=} \sum_{i\in[d]} R_i^{(n);j}$ to denote the

offspring count in the n^{th} generation of the cluster process $C_j^{\mathcal{O}}(\cdot)$. By (127), for any $c_1 > 0$ we have

$$\begin{aligned}\mathbf{P}(H_j > n\epsilon) &\le \mathbf{P}(K_j^{\mathcal{O}} \ge c_1 \log n) + \mathbf{P}\left(\max_{p,q\in[d]} \max_{1\le m\le \|\boldsymbol{S}_j\|} X_{p\leftarrow q}^{(m)} > \frac{n\epsilon}{c_1 \log n} \right) \\ &\le \mathbf{P}(K_j^{\mathcal{O}} \ge c_1 \log n) + \sum_{p,q\in[d]} \mathbf{P}\left(\max_{1\le m\le \|\boldsymbol{S}_j\|} X_{p\leftarrow q}^{(m)} > \frac{n\epsilon}{c_1 \log n} \right).\end{aligned}$$

To prove (125), it suffices to find some constant $c_1 > 0$ (depending only on β and j) such that

$$\mathbf{P}(K_j^{\mathcal{O}} \ge c_1 \log n) = o\big(a(n)\big) \quad \text{as } n \to \infty, \tag{128}$$

and that

$$\mathbf{P}\left(\max_{1\le m\le \|\boldsymbol{S}_j\|} X_{p\leftarrow q}^{(m)} > \frac{n\epsilon}{c_1 \log n} \right) = o\big(a(n)\big) \quad \text{as } n \to \infty, \qquad \forall (p,q) \in [d]^2. \tag{129}$$

Proof of Claim (128). We fix some $\tilde{\alpha} \in (1, \min_{p,q} \alpha_{p\leftarrow q})$. By Assumption 2, we have $\mathbf{E}[\tilde{B}_{p\leftarrow q}^{\tilde{\alpha}}] < \infty$. Under this condition and Assumption 1, it has been established in Section 3 of [5] (in particular, see Equation (6) in the paper and the discussion below) that

$$\mathbf{E}\big[(R^{(k);j})^{\tilde{\alpha}}\big] \le c_0 \nu^k, \qquad \forall k \ge 1,$$

for some $c_0 \in (0,\infty)$ and $\nu \in (0,1)$ whose values only depend on the law of the $\tilde{B}_{p,q}$'s. Then, by Markov's inequality,

$$\mathbf{P}(K_j^{\mathcal{O}} \ge k) = \mathbf{P}\big(R^{(k);j} > 0\big) = \mathbf{P}\big(R^{(k);j} \ge 1\big) \le c_0 \nu^k, \qquad \forall k \ge 1.$$

By picking $c_1 > 0$ large enough such that $c_1 \log(\nu) < -\beta$, we get

$$\mathbf{P}(K_j^{\mathcal{O}} \ge c_1 \log n) \le c_0 \cdot \nu^{c_1 \log(n)} = c_0 \cdot n^{c_1 \log(\nu)} = o\big(a(n)\big)$$

due to $a(n) \in \mathcal{RV}_{-\beta}(n)$. This verifies Claim (128).

Proof of Claim (129). Since the $X_{p\leftarrow q}^{(m)}$'s are independent of the $X_{p\leftarrow q}^{(n,k,k')}$ and $R_p^{(n);j}$ (and hence the cluster $C_j^{\mathcal{O}}$ and the cluster size vector $\boldsymbol{S}_j$),

$$\begin{aligned}\mathbf{P}\left(\max_{1\le m\le \|\boldsymbol{S}_j\|} X_{p\leftarrow q}^{(m)} > \frac{n\epsilon}{c_1 \log n} \right) &= \sum_{k\ge 1} \mathbf{P}\left(\max_{m\in[k]} X_{p\leftarrow q}^{(m)} > \frac{n\epsilon}{c_1 \log n} \right) \mathbf{P}(\|\boldsymbol{S}_j\| = k) \\ &\le \sum_{k\ge 1} k \cdot \mathbf{P}\left(X_{p\leftarrow q} > \frac{n\epsilon}{c_1 \log n} \right) \cdot \mathbf{P}(\|\boldsymbol{S}_j\| = k) \\ &= \mathbf{E}\big[\,\|\boldsymbol{S}_j\|\,\big] \cdot \mathbf{P}\left(X_{p\leftarrow q} > \frac{n\epsilon}{c_1 \log n} \right) \\ &= \frac{\mathbf{E}\|\boldsymbol{S}_j\|}{\int_0^\infty f_{p\leftarrow q}(t)dt} \cdot \int_{t > n\epsilon/(c_1 \log n)} f_{p\leftarrow q}(t)dt \qquad \text{by (126)} \\ &= o\big(a(n)\big) \qquad \text{due to (124)}.\end{aligned}$$

This concludes the proof of Claim (129). □

We prepare a lemma studying the distance between step functions under the metric $\boldsymbol{d}_{\mathcal{P}}^{[0,T]}$ in (42).

Lemma D.8. *Let $T \in (0,\infty)$, $k, k' \in \mathbb{N}$, $\epsilon > 0$, and $\xi, \xi' \in \mathbb{D}[0,T] = \mathbb{D}\big([0,T], \mathbb{R}^d\big)$. Suppose that*

- *ξ is a non-decreasing step function with k jumps and vanishes at the origin; i.e., $\xi(t) = \sum_{i=1}^{k} \boldsymbol{w}_i \mathbb{I}_{[t_i,T]}(t)$ $\forall t \in [0,T]$, where $\boldsymbol{w}_i \in \mathbb{R}_+^d \setminus \{\boldsymbol{0}\}$ for each $i \in [k]$ and $0 < t_1 < \ldots < t_k = T$;*
- *ξ' is a non-decreasing step function with k' jumps and vanishes at the origin;*
- *$\xi\big((t-\epsilon)\vee 0\big) \le \xi'(t) \le \xi(t)$ for each $t \in [0,T]$.*

Then, for any $t \in (0,T]$ such that $t_i \notin [t-\epsilon, t]$ $\forall i \in [k]$ (i.e., $\xi(\cdot)$ remains constant over $[(t-\epsilon)\vee 0, t]$),

$$\boldsymbol{d}_{\mathcal{P}}^{[0,t]}(\xi, \xi') \le \epsilon.$$

Proof. We write $\xi(t) = (\xi_1(t), \ldots, \xi_d(t))^\top$ and $\xi'(t) = (\xi'_1(t), \ldots, \xi'_d(t))^\top$. Let $I = \{t \in (0,T] :\ t_i \notin [t, t+\epsilon]\ \forall i \in [k]\}$. By definition of the product M_1 metric in (42), it suffices to show that $\boldsymbol{d}_{M_1}^{[0,t]}(\xi_i, \xi'_i) \le \epsilon$ for each $i \in [d]$ and $t \in I$, where $\boldsymbol{d}_{M_1}^{[0,t]}$ is the M_1 metric for $\mathbb{R}$-valued càdlàg paths; see (41). As a result, it suffices to prove the claim for $\mathbb{R}$-valued càdlàg paths. More specifically, let $T \in (0,\infty)$, $k, k' \in \mathbb{Z}_+$, $\epsilon > 0$, and $x, y \in \mathbb{D}\big([0,T], \mathbb{R}\big)$, and suppose that

- x is a non-decreasing step function with k jumps and vanishes at the origin,
- y is a non-decreasing step function with k' jumps and vanishes at the origin,
- and $x\big((t-\epsilon)\vee 0\big) \le y(t) \le x(t)$ for each $t \in [0,T]$.

It suffices to fix some $t \in (0,T]$ such that $\Delta x(u) = 0$ for each $u \in [(t-\epsilon)\vee 0, t]$, and show that

$$\boldsymbol{d}_{M_1}^{[0,t]}(x, y) \le \epsilon. \tag{130}$$

Since both x and y are non-decreasing step functions over $[0,t]$, there exist some $k^* \in \mathbb{Z}_+$ and a sequence $0 = z_0 < z_1 < z_2 < \ldots < z_{k^*}$ such that

$$\big\{z \in [0,\infty) :\ z = x(u) \text{ or } y(u) \text{ for some } u \in [0,t]\big\} = \{z_0, z_1, \ldots, z_{k^*}\}.$$

For each $l \in [k^*]$, let

$$t_l^x \stackrel{\text{def}}{=} \min\{u \ge 0 :\ x(u) \ge z_l\}, \quad t_l^y \stackrel{\text{def}}{=} \min\{u \ge 0 :\ y(u) \ge z_l\}$$

be the first time $x(u)$ (resp., $y(u)$) crosses the level of z_l. Due to $x\big((u-\epsilon)\vee 0\big) \le y(u) \le x(u)$ for each $u \in [0,t]$, we have

$$t_l^y - \epsilon \le t_l^x \le t_l^y, \qquad \forall l \in [k^*]. \tag{131}$$

Also, since $\Delta x(u) = 0$ for each $u \in [(t-\epsilon)\vee 0, t]$, we know that the value of the step function $x(\cdot)$ remains constant over $[(t-\epsilon)\vee 0, t]$. Then by $x\big((t-\epsilon)\vee 0\big) \le y(t) \le x(t)$, at the right endpoint of the time interval $[0,t]$ we have

$$x(t) = y(t) = z_{k^*}. \tag{132}$$

Now, we prove Claim (130) by constructing suitable parametric representations for paths x and y. First, for the path x, we adopt the convention $t_0^x = 0$, and define $\big(u^x(w), s^x(w)\big)$ such that (for each

$l \in [k^*]$) over the interval $w \in \left[\frac{2(l-1)}{2k^*+1}, \frac{2l}{2k^*+1}\right)$, the parametric representation (u^x, s^x) spends half of the time moving uniformly and horizontally from (t^x_{l-1}, z_{l-1}) to (t^x_l, z_{l-1}) over the connected graph of x, and then spends the other half time moving uniformly and vertically from (t^x_l, z_{l-1}) to (t^x_l, z_l). More precisely, for each $l \in [k^*]$, let

$$u^x(w) \stackrel{\text{def}}{=} \begin{cases} t^x_{l-1} + (t^x_l - t^x_{l-1}) \cdot \Big[(2k^*+1)w - 2(l-1)\Big] & \text{if } w \in \left[\dfrac{2(l-1)}{2k^*+1}, \dfrac{2(l-1)+1}{2k^*+1}\right) \\ t^x_l & \text{if } w \in \left[\dfrac{2(l-1)+1}{2k^*+1}, \dfrac{2l}{2k^*+1}\right) \end{cases},$$

$$s^x(w) \stackrel{\text{def}}{=} \begin{cases} x\big(u^x(w)\big) = z_{l-1} & \text{if } w \in \left[\dfrac{2(l-1)}{2k^*+1}, \dfrac{2(l-1)+1}{2k^*+1}\right) \\ z_{l-1} + (z_l - z_{l-1}) \cdot \Big[(2k^*+1)w - \big(2(l-1)+1\big)\Big] & \text{if } w \in \left[\dfrac{2(l-1)+1}{2k^*+1}, \dfrac{2l}{2k^*+1}\right) \end{cases}.$$

By definition, $x(u)$ remains constant over $u \in [t^x_{k^*}, t]$, so we set

$$u^x(w) \stackrel{\text{def}}{=} t^x_{k^*} + (t - t^x_{k^*}) \cdot \Big[(2k^*+1)w - 2k^*\Big], \;\; s^x(w) \stackrel{\text{def}}{=} z_{k^*}, \qquad \forall w \in \left[\frac{2k^*}{2k^*+1}, 1\right].$$

Here, the choice of $s^x(w) = z_{k^*}$ is valid due to the property (132). Analogously, for the path y we define (for each $l \in [k^*]$)

$$u^y(w) \stackrel{\text{def}}{=} \begin{cases} t^y_{l-1} + (t^y_l - t^y_{l-1}) \cdot \Big[(2k^*+1)w - 2(l-1)\Big] & \text{if } w \in \left[\dfrac{2(l-1)}{2k^*+1}, \dfrac{2(l-1)+1}{2k^*+1}\right) \\ t^y_l & \text{if } w \in \left[\dfrac{2(l-1)+1}{2k^*+1}, \dfrac{2l}{2k^*+1}\right) \end{cases},$$

$$s^y(w) \stackrel{\text{def}}{=} \begin{cases} y\big(u^y(w)\big) = z_{l-1} & \text{if } w \in \left[\dfrac{2(l-1)}{2k^*+1}, \dfrac{2(l-1)+1}{2k^*+1}\right) \\ z_{l-1} + (z_l - z_{l-1}) \cdot \Big[(2k^*+1)w - \big(2(l-1)+1\big)\Big] & \text{if } w \in \left[\dfrac{2(l-1)+1}{2k^*+1}, \dfrac{2l}{2k^*+1}\right) \end{cases},$$

and due to $y(u) \equiv z_{k^*}$ for any $u \in [t^y_{k^*}, t]$ (see (132)), we set

$$u^y(w) \stackrel{\text{def}}{=} t^y_{k^*} + (t - t^y_{k^*}) \cdot \Big[(2k^*+1)w - 2k^*\Big], \;\; s^y(w) \stackrel{\text{def}}{=} z_{k^*}, \qquad \forall w \in \left[\frac{2k^*}{2k^*+1}, 1\right].$$

Lastly, by our construction of the parametric representations (u^x, s^x) and (u^y, s^y), we have $s^y(w) = s^x(w)$ for any $w \in [0,1]$, and $\sup_{w\in[0,1]} |u^x(w) - u^y(w)| \le \epsilon$ due to (131). This concludes the proof of Claim (130). □

Equipped with Lemmas D.7 and D.8, we provide high probability bounds over the distance between $\bar{\boldsymbol{L}}_n^{>\delta;[0,\infty)}$ in (109) and $\bar{\boldsymbol{N}}_n^{>\delta;[0,\infty)}$ in (117) under $\boldsymbol{d}_{\mathcal{P}}^{[0,\infty)}$.

Lemma D.9. *Let Assumptions 1–4 hold. Let $\beta > 1$ and $a : (0,\infty) \to (0,\infty)$ be a regularly varying function with $a(x) \in \mathcal{RV}_{-\beta}(x)$ as $x \to \infty$. Suppose that*

$$\int_{x/\log x}^{\infty} f_{p\leftarrow q}(t)dt = o\big(a(x)/x\big) \ \text{ as } x \to \infty, \qquad \forall p, q \in [d]. \tag{133}$$

Then, for each $\Delta > 0$,

$$\lim_{n\to\infty} \big(a(n)\big)^{-1} \cdot \mathbf{P}\Big(\boldsymbol{d}_{\mathcal{P}}^{[0,\infty)}\Big(\bar{\boldsymbol{L}}_n^{>\delta;[0,\infty)},\ \bar{\boldsymbol{N}}_n^{>\delta;[0,\infty)}\Big) > \Delta\Big) = 0, \qquad \forall \delta > 0 \text{ small enough.}$$

Proof. Analogous to the proof of Lemma D.5, we fix some T large enough such that $e^{-T} < \Delta/2$, and note that it suffices to prove

$$\mathbf{P}\left(\int_0^T e^{-t} \cdot \left[\boldsymbol{d}_{\mathcal{P}}^{[0,t]}\Big(\bar{\boldsymbol{L}}_n^{>\delta;[0,t]},\ \bar{\boldsymbol{N}}_n^{>\delta;[0,t]}\Big) \wedge 1\right] dt > \frac{\Delta}{2}\right) = o\big(a(n)\big), \ \ \forall \delta > 0 \text{ small enough.} \tag{134}$$

On the one hand, by comparing the definitions in (108) and (116), we must have

$$\bar{\boldsymbol{N}}_n^{>\delta}(t) \le \bar{\boldsymbol{L}}_n^{>\delta}(t), \qquad \forall \delta > 0,\ n \ge 1,\ t \ge 0. \tag{135}$$

On the other hand, recall that for some cluster induced by a type-j immigrant, we use $\boldsymbol{S}_j^{(k)}$ to denote the cluster size, $T_{j;k}^{C}$ for the arrival time of the immigrant inducing the cluster, and $H_j^{(k)}$ for the lifetime of the cluster (see (118)). On the event $\big\{H_j^{(k)} \le n\epsilon\ \forall (k,j) \text{ with } T_{j;k}^{C} \le nT\big\}$, we must have

$$\bar{\boldsymbol{N}}_n^{>\delta}(t) \ge \bar{\boldsymbol{L}}_n^{>\delta}\big((t-\epsilon)\vee 0\big), \quad \forall \delta > 0,\ n \ge 1,\ t \in [0,T]. \tag{136}$$

Indeed, for any cluster induced by an immigrant arriving at time $T_{j;k}^{C} \le n(t-\epsilon)$ and with cluster lifetime $H_j^{(k)} \le n\epsilon$, all descendants in this cluster must have arrived by the time nt. Furthermore, let $\tau_n^{>\delta}(k)$ be defined as in (111), and note that the sequence of stopping times $\big(\tau_n^{>\delta}(k)\big)_{k\ge 1}$ corresponds to the arrival times of jumps in $\bar{\boldsymbol{L}}_n^{>\delta}(t)$. Therefore, for any $K \ge 1$, on the event

$$A_n(K,T,\epsilon,\delta) \stackrel{\text{def}}{=} \{\tau_n^{>\delta}(K+1) > T\} \cap \big\{H_j^{(k)} \le n\epsilon\ \forall (k,j) \text{ with } T_{j;k}^{C} \le nT\big\},$$

we have $\bar{\boldsymbol{L}}_n^{>\delta}\big((t-\epsilon)\vee 0\big) \le \bar{\boldsymbol{N}}_n^{>\delta}(t) \le \bar{\boldsymbol{L}}_n^{>\delta}(t)$ for any $t \in [0,T]$, and $\bar{\boldsymbol{L}}_n^{>\delta;[0,T]}$ is a step function that vanishes at the origin and has at most K jumps (all belonging to $\mathbb{R}_+^d \setminus \{\mathbf{0}\}$). By Lemma D.8, there exists a Borel set $I \subseteq [0,T]$, whose choice is random and depends on the paths $\bar{\boldsymbol{L}}_n^{>\delta;[0,t]}$, $\bar{\boldsymbol{N}}_n^{>\delta;[0,t]}$, such that

- $\mathcal{L}_\infty(I) \le K\epsilon$, where $\mathcal{L}_\infty$ is the Lebesgue measure on $(0,\infty)$;
- $\boldsymbol{d}_{\mathcal{P}}^{[0,t]}\big(\bar{\boldsymbol{L}}_n^{>\delta;[0,t]},\ \bar{\boldsymbol{N}}_n^{>\delta;[0,t]}\big) \le \epsilon$ for any $t \in [0,T] \setminus I$.

Therefore, on the event $A_n(K,T,\epsilon,\delta)$, we have

$$\begin{aligned}
&\int_0^T e^{-t} \cdot \left[\boldsymbol{d}_{\mathcal{P}}^{[0,t]}\Big(\bar{\boldsymbol{L}}_n^{>\delta;[0,t]},\ \bar{\boldsymbol{N}}_n^{>\delta;[0,t]}\Big) \wedge 1\right] dt \\
&\le \int_{t\in[0,T]\setminus I} \boldsymbol{d}_{\mathcal{P}}^{[0,t]}\Big(\bar{\boldsymbol{L}}_n^{>\delta;[0,t]},\ \bar{\boldsymbol{N}}_n^{>\delta;[0,t]}\Big) dt + \int_{t\in I} dt \le T\epsilon + K\epsilon.
\end{aligned}$$

Furthermore, by picking $\epsilon > 0$ small enough we have $(T + K)\epsilon < \Delta/2$. In summary, to prove Claim (134), it suffices to show the existence of some $K \geq 1$ such that

$$\mathbf{P}\Big(\tau_n^{>\delta}(K+1) \leq T\Big) = o\big(a(n)\big), \qquad \forall \delta > 0 \text{ small enough}, \tag{137}$$

and that for any $\epsilon > 0$,

$$\mathbf{P}\Big(H_j^{(k)} > n\epsilon \text{ for some } (k,j) \text{ with } T_{j;k}^C \leq nT\Big) = o\big(a(n)\big). \tag{138}$$

Proof of Claim (137). This has been established in the proof of Claim (112) in Lemma D.5.

Proof of Claim (138). It suffices to fix some $j \in [d]$, $\epsilon > 0$ and show that

$$\mathbf{P}\Big(H_j^{(k)} > n\epsilon \text{ for some } k \geq 1 \text{ such that } T_{j;k}^C \leq nT\Big) = o\big(a(n)\big).$$

Recall that $(T_{j;k}^C)_{k\geq 1}$ are generated by a Poisson process with constant rate c_j. Therefore, by picking C large enough (whose value only depends on T and the constants c_j), it follows from Cramer's Theorem that $\mathbf{P}\big(T_{j;\lfloor nC\rfloor}^C \leq nT\big) = o\big(a(n)\big)$. Fixing such C, we only need to show that $\mathbf{P}\big(H_j^{(k)} > n\epsilon \text{ for some } k \leq \lfloor nC\rfloor\big) = o\big(a(n)\big)$ holds for any $\delta > 0$ small enough. However, since $H_j^{(k)}$ are i.i.d. copies of H_j (see (118) and (119)),

$$\mathbf{P}\big(H_j^{(k)} > n\epsilon \text{ for some } k \leq \lfloor nC\rfloor\big) \leq nC \cdot \mathbf{P}(H_j > n\epsilon) = nC \cdot o\big(a(n)/n\big) = o\big(a(n)\big).$$

The second to last equality follows from the condition (133) and Lemma D.7. This concludes the proof of Claim (138). □

Lastly, we prove Proposition 3.5.

Proof of Proposition 3.5. Due to

$$\begin{aligned}
&\boldsymbol{d}_{\mathcal{P}}^{[0,\infty)}\left(\bar{\boldsymbol{N}}_n^{[0,\infty)}, \bar{\boldsymbol{L}}_n^{[0,\infty)}\right) \\
&\leq \boldsymbol{d}_{\mathcal{P}}^{[0,\infty)}\left(\bar{\boldsymbol{N}}_n^{[0,\infty)}, \bar{\boldsymbol{N}}_n^{>\delta,[0,\infty)} + \mu_N \mathbf{1}\right) + \boldsymbol{d}_{\mathcal{P}}^{[0,\infty)}\left(\bar{\boldsymbol{N}}_n^{>\delta,[0,\infty)} + \mu_N \mathbf{1}, \bar{\boldsymbol{L}}_n^{>\delta,[0,\infty)} + \mu_N \mathbf{1}\right) \\
&\qquad + \boldsymbol{d}_{\mathcal{P}}^{[0,\infty)}\left(\bar{\boldsymbol{L}}_n^{[0,\infty)}, \bar{\boldsymbol{L}}_n^{>\delta,[0,\infty)} + \mu_N \mathbf{1}\right) \\
&= \boldsymbol{d}_{\mathcal{P}}^{[0,\infty)}\left(\bar{\boldsymbol{N}}_n^{[0,\infty)}, \bar{\boldsymbol{N}}_n^{>\delta,[0,\infty)} + \mu_N \mathbf{1}\right) + \boldsymbol{d}_{\mathcal{P}}^{[0,\infty)}\left(\bar{\boldsymbol{N}}_n^{>\delta,[0,\infty)}, \bar{\boldsymbol{L}}_n^{>\delta,[0,\infty)}\right) + \boldsymbol{d}_{\mathcal{P}}^{[0,\infty)}\left(\bar{\boldsymbol{L}}_n^{[0,\infty)}, \bar{\boldsymbol{L}}_n^{>\delta,[0,\infty)} + \mu_N \mathbf{1}\right),
\end{aligned}$$

Proposition 3.5 follows directly from Lemmas D.5, D.6, and D.9. □

D.4. Univariate Results for M_1 Topology of $\mathbb{D}[0,T]$

We show that in the univariate setting, Theorem 3.3 can be uplifted to the M_1 topology of $\mathbb{D}[0,T]$—the càdlàg space with compact domain. Without loss of generality, throughout this section we fix $T = 1$ and adopt the notations

$$\mathbb{D} = \mathbb{D}\big([0,1], \mathbb{R}\big), \quad \boldsymbol{d}_{M_1} = \boldsymbol{d}_{M_1}^{[0,1]}.$$

We start by adapting the notions from the main paper to the univariate setting, as many of them now admit simpler or more explicit expressions. First, the univariate Hawkes process $N(t)$ has initial value $N(0) = 0$ and conditional intensity

$$h(t) = c + \int_0^T \tilde{B}(s) f(t-s) dN(s).$$

Here, the constant $c_N > 0$ is the arrival rate of immigrants, the decay function $f(\cdot)$ is such that $\|f\|_1 < \infty$, and the excitation rates $\big(\tilde{B}(s)\big)_{s>0}$ are independent copies of $\tilde{B}$ with $\mathbf{P}(\tilde{B} > x) \in \mathcal{RV}_{-\alpha}(x)$ for some $\alpha > 1$. Second, we adapt the tail asymptotics characterized in Section 2.3 for Hawkes processes clusters to the unvariate setting. Let S be a generic copy for the size of a cluster. For any $x > 0$, we have

$$\mathbf{P}(S > nx) \sim \frac{(\mathbf{E}S)^{1+\alpha}}{x^\alpha} \cdot \mathbf{P}(B > n), \qquad \text{as } n \to \infty, \tag{139}$$

where $B \overset{\mathcal{D}}{=} \text{Poisson}\big(\tilde{B} \cdot \|f\|_1\big)$. Also, note that the regularly varying law of $\tilde{B}$ implies

$$\mathbf{P}(B > x) \sim \mathbf{P}(\tilde{B} \cdot \|f\|_1 > x) \sim \|f\|_1^\alpha \, \mathbf{P}(\tilde{B} > x).$$

The asymptotics (139) are well-documented in existing literature on branching processes (e.g., Theorem 1 of [1]) and are an immediate consequence of Theorem 2.6 in univariate settings. In particular, using notations in Section 2.3, we have $\mathbf{P}(S > nx) \sim \lambda(n) \cdot \mathbf{C}_S\big((x,\infty)\big)$ as $n \to \infty$, where

$$\lambda(n) = \mathbf{P}(B > n); \qquad \mathbf{C}_S\big((x,\infty)\big) = (\mathbf{E}S)^{1+\alpha} \big/ x^\alpha \ \forall x > 0. \tag{140}$$

Next, we adapt the notions introduced in Section 3 to the univariate setting. By (32), for each $k \geq 1$ we set

$$\check{\lambda}_k(n) = \big(n\mathbf{P}(B > n)\big)^k. \tag{141}$$

Analogous to (27), for each $k \geq 0$ and $x \in \mathbb{R}$, we define

$$\check{\mathbb{D}}_{k;x} \overset{\text{def}}{=} \left\{ x\mathbf{1} + \sum_{i=1}^k w_i \mathbb{I}_{[t_i,1]} : \ w_i \geq 0 \text{ and } t_i \in (0,1] \ \forall i \in [k] \right\},$$

where $\mathbf{1}(t) = t$ is the identity mapping of $\mathbb{R}^1$. That is, $\check{\mathbb{D}}_{k;x}$ is the subset of $\mathbb{D} = \mathbb{D}\big([0,1],\mathbb{R}\big)$ containing any piece-wise linear function with slope x that vanishes at the origin and has at most k jumps (all upward). Adapting (45) and (47), we define

$$\mathbf{C} = c \cdot \mathbf{C}_S, \qquad \mu_N = c \cdot \mathbf{E}S. \tag{142}$$

We also adapt (30) and define Borel measures on $\check{\mathbb{D}}_{k;x}$ by (for each $k \geq 1$ and $x \in \mathbb{R}$)

$$\begin{aligned}
\check{\mathbf{C}}_{k;x}(\,\cdot\,) &= \frac{1}{k!}\mathbf{E}\left[\int_{w_i>0 \ \forall i\in[k]} \mathbb{I}\left\{ x\mathbf{1} + \sum_{i=1}^k w_i \mathbb{I}_{[U_i,1]} \in \ \cdot \ \right\} \bigtimes_{i=1}^k \mathbf{C}(dw_i) \right] \\
&= \frac{(\alpha c_N)^k (\mathbf{E}S)^{k(1+\alpha)}}{k!} \mathbf{E}\left[\int_{w_i>0 \ \forall i\in[k]} \mathbb{I}\left\{ x\mathbf{1} + \sum_{i=1}^k w_i \mathbb{I}_{[U_i,1]} \in \ \cdot \ \right\} \bigtimes_{i=1}^k \frac{dw_i}{(w_i)^{1+\alpha}} \right],
\end{aligned}$$

where the sequence $(U_i)_{i\geq 1}$ are independent copies of Unif$(0,1)$, and the last line of the display above follows from the definition of $\mathbf{C}_S(\cdot)$ in (140). Now, let

$$\bar{N}_n(t) = N(nt)/n, \qquad \bar{\boldsymbol{N}}_n = \{\bar{N}_n(t) : \ t \in [0,1]\}.$$

We are ready to state the strengthened univariate results.

Theorem D.10. *Let $k \in \mathbb{N}$. Let A be a Borel set of $\mathbb{D}$ equipped with the Skorokhod M_1 topology. Suppose that*

$$\int_{x/\log x}^{\infty} f(t)dt = o\big(\breve{\lambda}_k(x)/x\big), \tag{143}$$

and the set A is bounded away from $\breve{\mathbb{D}}_{k-1;\mu_N}$ under $\boldsymbol{d}_{M_1}$. Then,

$$\breve{\mathbf{C}}_{k;\mu_N}(A^\circ) \leq \liminf_{n\to\infty} \frac{\mathbf{P}(\bar{\boldsymbol{N}}_n \in A)}{\breve{\lambda}_k(n)} \leq \limsup_{n\to\infty} \frac{\mathbf{P}(\bar{\boldsymbol{N}}_n \in A)}{\breve{\lambda}_k(n)} \leq \breve{\mathbf{C}}_{k;\mu_N}(A^-),$$

where the interior and exterior sets A° and A^- are determined under $\boldsymbol{d}_{M_1}$ (i.e., w.r.t. the M_1 topology of $\mathbb{D} = \mathbb{D}\big([0,1],\mathbb{R}\big)$).

Since Theorem D.10 is established w.r.t. the M_1 topology of $\mathbb{D}[0,T]$ (i.e., the càdlàg space with compact domain) instead of $\mathbb{D}[0,\infty)$, our results now address a broader class of events or functional of $\bar{\boldsymbol{N}}_n$ compared to only applying Theorem 3.3 directly to univariate Hawkes processes. We conclude this subsection with the proof of Theorem D.10.

Proof of Theorem D.10. The proof is in the same spirit as Theorem 3.3. In particular, like in (52) and (55), we set $L(t)$ as the compound Poisson process with c_N as arrival rates and S (the cluster size) as increments, and let $\bar{\boldsymbol{L}}_n = \{L(nt)/n : \ t \in [0,1]\}$ be the scaled version of the sample path of $L(t)$ embedded in $\mathbb{D}$. By Proposition 3.4 and Lemma 2.3, it suffices to fix some positive integer k and verify the following: Let A be a Borel set of $\mathbb{D}$ equipped with the M_1 topology such that A is bounded away from $\breve{\mathbb{D}}_{k-1;\mu_N}$ under $\boldsymbol{d}_{M_1}$; for each $\Delta > 0$ we have

$$\mathbf{P}\Big(\boldsymbol{d}_{M_1}(\bar{\boldsymbol{N}}_n, \bar{\boldsymbol{L}}_n)\mathbb{I}\big\{\bar{\boldsymbol{N}}_n \text{ or } \bar{\boldsymbol{L}}_n \in A\big\} > \Delta\Big) = o\big(\breve{\lambda}_k(n)\big), \quad \text{as } n \to \infty. \tag{144}$$

To this end, we consider the definitions in (107)–(109) and (115)–(117) in the unvariate setting. That is, we use $\bar{\boldsymbol{N}}_n^{>\delta}$ (resp., $\bar{\boldsymbol{N}}_n^{\leqslant\delta}$) to denote a modified version of $\bar{\boldsymbol{N}}_n$ where we only include clusters in $N(t)$ whose sizes are larger than (resp., bounded by) $n\delta$. We adopt similar notations $\bar{\boldsymbol{L}}_n^{>\delta}$ and $\bar{\boldsymbol{L}}_n^{\leqslant\delta}$ for $\bar{\boldsymbol{L}}_n$. First, the proof of Claim (110) in Lemma D.5 implies that for any $\Delta > 0$,

$$\mathbf{P}\bigg(\sup_{t\in[0,1]}\Big|\bar{L}_n(t) - \Big(\bar{L}_n^{>\delta}(t) + \mu_N t\Big)\Big| > \Delta\bigg) = o\big(\breve{\lambda}_k(n)\big), \qquad \forall \delta > 0 \text{ small enough.} \tag{145}$$

Similarly, the proof of Claim (120) in Lemma D.6 implies that for any $\Delta > 0$,

$$\mathbf{P}\bigg(\sup_{t\in[0,1]}\Big|\bar{N}_n(t) - \Big(\bar{N}_n^{>\delta}(t) + \mu_N t\Big)\Big| > \Delta\bigg) = o\big(\breve{\lambda}_k(n)\big), \qquad \forall \delta > 0 \text{ small enough.} \tag{146}$$

Meanwhile, in (144) note that: (i) $\boldsymbol{d}_{M_1}$ is upper bounded by the uniform metric; (ii) whenever $\bar{\boldsymbol{L}}_n \in A$ and $\boldsymbol{d}_{M_1}\big(\bar{\boldsymbol{L}}_n, \bar{\boldsymbol{L}}_n^{>\delta} + \mu_N \mathbf{1}\big) \le \Delta$, we must have $\bar{\boldsymbol{L}}_n^{>\delta} + \mu_N \mathbf{1} \in A^\Delta$; the same can be argued for $\bar{\boldsymbol{N}}_n$; and (iii) for any $\Delta > 0$ small enough the set A^Δ is still bounded away from $\breve{\mathbb{D}}_{k-1;\mu_N}$. Therefore, due to (145)–(146) and the arbitrariness of the choice of A and $\Delta > 0$ in (144), it suffices to show the following: Let A be a Borel set of $(\mathbb{D}, \boldsymbol{d}_{M_1})$ that is bounded away from $\breve{\mathbb{D}}_{k-1;\mu_N}$ under $\boldsymbol{d}_{M_1}$ and let $\Delta > 0$; for any $\delta > 0$ small enough we have

$$\mathbf{P}\Big(\boldsymbol{d}_{M_1}\big(\bar{\boldsymbol{N}}_n^{>\delta}, \bar{\boldsymbol{L}}_n^{>\delta}\big)\underbrace{\mathbb{I}\big\{\bar{\boldsymbol{N}}_n^{>\delta} + \mu_N \mathbf{1} \in A \text{ or } \bar{\boldsymbol{L}}_n^{>\delta} + \mu_N \mathbf{1} \in A\big\}}_{\stackrel{\text{def}}{=} A_n(\delta)} > \Delta\Big) = o\big(\breve{\lambda}_k(n)\big). \tag{147}$$

Without loss of generality, in (147) we fix $\delta > 0$, and take some $\Delta > 0$ small enough such that A^Δ is still bounded away from $\breve{\mathbb{D}}_{k-1;\mu_N}$ under $\boldsymbol{d}_{M_1}$. To proceed, we define a few events. Let H be the lifetime of a generic copy of the cluster S, and the sequence (S_j, H_j) be independent copies of (S, H). Generate the sequence $0 < T_1 < T_2 < \dots$ by a Poisson process with rate c_N, which denote the arrival times of each immigrant in $N(t)$. By definition, the associated compound Poisson process admits the form $L(t) = \sum_{j\ge 1} S_j \mathbb{I}\{T_j \le t\}$. For each $\epsilon > 0$, define events

$$F_n(\epsilon) \stackrel{\text{def}}{=} \Big\{H_j \le n\epsilon \text{ for any } j \ge 1 \text{ with } T_j \le n\Big\}.$$

That is, on the event $F_n(\epsilon)$, any immigrant that arrives by time n induces a cluster with lifetime shorter than $n\epsilon$. Next, let $K_n^{>\delta}$ be the count of jumps in $\bar{\boldsymbol{L}}_n^{>\delta}$. Due to the $O(n)$ time scaling in $\bar{\boldsymbol{L}}_n^{>\delta}$, we have $K_n^{>\delta} = \#\{j \ge 1 :\ T_j \le n,\ S_j > n\delta\}$. Now, define events

$$E_n^{(1)}(\delta) \stackrel{\text{def}}{=} \Big\{K_n^{>\delta} = k\Big\}, \qquad E_n^{(2)}(\delta, \epsilon) \stackrel{\text{def}}{=} \Big\{\bar{\boldsymbol{L}}_n^{>\delta}(t) - \bar{\boldsymbol{L}}_n^{>\delta}(t-) = 0\ \forall t \in [1-\epsilon, 1]\Big\}.$$

Also, recall the definition of events $A_n(\delta)$ in (147). Suppose that (i) for any $\epsilon > 0$, we have

$$\mathbf{P}\Big(\big(F_n(\epsilon)\big)^c\Big) = o\big(\breve{\lambda}_k(n)\big); \tag{148}$$

(ii) next,

$$\mathbf{P}\Big(\Big(A_n(\delta) \cap F_n(\epsilon)\Big) \setminus E_n^{(1)}(\delta)\Big) = o\big(\breve{\lambda}_k(n)\big), \quad \forall \epsilon \in (0, \Delta), \tag{149}$$

$$\lim_{\epsilon \downarrow 0} \lim_{n\to\infty} \mathbf{P}\Big(\Big(A_n(\delta) \cap F_n(\epsilon) \cap E_n^{(1)}(\delta)\Big) \setminus E_n^{(2)}(\delta, \epsilon)\Big) \Big/ \breve{\lambda}_k(n) = 0; \tag{150}$$

(iii) furthermore, for any $\epsilon \in (0, \Delta)$,

$$\text{it holds on the event } A_n(\delta) \cap F_n(\epsilon) \cap E_n^{(1)}(\delta) \cap E_n^{(2)}(\delta, \epsilon) \text{ that } \boldsymbol{d}_{M_1}\big(\bar{\boldsymbol{N}}_n^{>\delta}, \bar{\boldsymbol{L}}_n^{>\delta}\big) \le \Delta. \tag{151}$$

Then, the bound (147) follows directly from Claims (148)–(151). Now, it only remains to verify these claims.

Proof of Claim (148). This is an immediate consequence of Lemma D.7 under condition (143). See also the proof of Claim (138) in Lemma D.9.

Proof of Claim (149). By the law of the compound Poisson process $L(t)$,

$$\begin{aligned}\mathbf{P}\big(K_n^{>\delta} \geq k+1\big) &= \mathbf{P}\Big(\text{Poisson}\big(n\mathbf{P}(S>n\delta)\big) \geq k+1\Big)\\ &\leq \big(n\mathbf{P}(S>n\delta)\big)^{k+1} \qquad \text{due to } \mathbf{P}\big(\text{Poisson}(\lambda)\geq m\big)\leq \lambda^m\\ &= O\Big(\big(n\mathbf{P}(B>n)\big)^{k+1}\Big) = o\big(\breve{\lambda}_k(n)\big) \qquad \text{by (139) and (141).}\end{aligned}$$

Therefore, it suffices to show that for any $\epsilon \in (0,\Delta)$,

$$\Big\{\bar{\boldsymbol{N}}_n^{>\delta} + \mu_N\mathbf{1} \in A \text{ or } \bar{\boldsymbol{L}}_n^{>\delta} + \mu_N\mathbf{1} \in A\Big\} \cap F_n(\epsilon) \cap \Big\{K_n^{>\delta} \leq k-1\Big\} = \emptyset.$$

First, on $\big\{K_n^{>\delta} \leq k-1\big\}$, we have $\bar{\boldsymbol{L}}_n^{>\delta} + \mu_N\mathbf{1} \in \breve{\mathbb{D}}_{k-1;\mu_N}$. Since A is bounded away from $\breve{\mathbb{D}}_{k-1;\mu_N}$, this confirms that $\big\{\bar{\boldsymbol{L}}_n^{>\delta} + \mu_N\mathbf{1} \in A\big\} \cap \big\{K_n^{>\delta} \leq k-1\big\} = \emptyset$. Next, on the event $F_n(\epsilon)$, the lifetime of any cluster arriving by time n is upper bounded by $n\epsilon$. Note that in $\bar{N}_n^{>\delta}$ and $\bar{L}_n^{>\delta}$, due to the $O(n)$ time scaling, this translates to an ϵ upper bound time-wise. Also, recall that we only consider $\epsilon \in (0,\Delta)$. By the same arguments in Lemma D.8, one can see that, on the event $F_n(\epsilon) \cap \big\{K_n^{>\delta} \leq k-1\big\}$, by setting $\xi(t) = \bar{L}_n^{>\delta}(t) \wedge \bar{N}_n^{>\delta}(1)\ \forall t \in [0,1]$, the path $\xi \in \mathbb{D}$ admits the form

$$\xi(t) = \mu_N t + \sum_{i=1}^{k-1} w_i \mathbb{I}_{[t_i,1]}(t), \qquad \text{where } w_i \geq 0 \text{ and } t_i \in (0,1]\ \forall i \in [k-1],$$

and satisfies $\boldsymbol{d}_{M_1}(\xi, \bar{\boldsymbol{N}}_n^{>\delta}) < \epsilon < \Delta$. This confirms that, on the event $F_n(\epsilon) \cap \big\{K_n^{>\delta} \leq k-1\big\}$, we must have $\bar{\boldsymbol{N}}_n^{>\delta} + \mu_N\mathbf{1} \in (\breve{\mathbb{D}}_{k-1;\mu_N})^{\Delta}$. However, this implies that on the event $F_n(\epsilon) \cap \big\{K_n^{>\delta} \leq k-1\big\}$, we must have $\bar{\boldsymbol{N}}_n^{>\delta} + \mu_N\mathbf{1} \notin A$. Indeed, note that

$$F_n(\epsilon) \cap \big\{K_n^{>\delta} \leq k-1\big\} \cap \big\{\bar{\boldsymbol{N}}_n^{>\delta} + \mu_N\mathbf{1} \in A\big\} \subseteq \big\{\bar{\boldsymbol{N}}_n^{>\delta} + \mu_N\mathbf{1} \in (\breve{\mathbb{D}}_{k-1;\mu_N})^{\Delta} \cap A\big\}.$$

Meanwhile, recall the running assumption that A^{Δ} is bounded away from $\breve{\mathbb{D}}_{k-1;\mu_N}$ under $\boldsymbol{d}_{M_1}$, which implies $(\mathbb{D}_{k-1;\mu_N})^{\Delta} \cap A = \emptyset$, and hence $F_n(\epsilon) \cap \big\{K_n^{>\delta} \leq k-1\big\} \cap \big\{\bar{\boldsymbol{N}}_n^{>\delta} + \mu_N\mathbf{1} \in A\big\} = \emptyset$. This concludes the proof of Claim (149).

Proof of Claim (150). It suffices to show that

$$\lim_{\epsilon \downarrow 0} \mathbf{P}\Big(\underbrace{K_n^{>\delta} = k;\ \exists t \in [1-\epsilon,1] \text{ such that } \bar{L}_n^{>\delta}(t) - \bar{L}_n^{>\delta}(t-) \neq 0}_{= E_n^{(1)}(\delta) \setminus E_n^{(2)}(\delta,\epsilon)}\Big) \Big/ \breve{\lambda}_k(n) = 0.$$

By the law of compound Poisson processes, we have (where the U_j's are i.i.d. copies of $\text{Unif}(0,1)$)

$$\begin{aligned}&\mathbf{P}\big(E_n^{(1)}(\delta) \setminus E_n^{(2)}(\delta,\epsilon)\big)\\ &= \mathbf{P}\Big(\text{Poisson}\big(n\mathbf{P}(S>n\delta)\big) = k\Big) \cdot \mathbf{P}\Big(U_j \in [1-\epsilon,1] \text{ for some } j \in [k]\Big)\\ &\leq \big(n\mathbf{P}(S>n\delta)\big)^k \cdot k\epsilon\end{aligned}$$

$$\sim k\epsilon \cdot \left(\frac{(\mathbf{E}S)^{1+\alpha}}{\delta^{\alpha}}\right)^{k} \cdot \breve{\lambda}_k(n) \qquad \text{by (139) and (141).}$$

Sending $\epsilon \downarrow 0$, we conclude the proof of Claim (150).

Proof of Claim (151). This follows form Lemma D.8 and our choice of $\epsilon \in (0, \Delta)$. In particular, note that in the one-dimensional case, the product M_1 metric $\boldsymbol{d}_{\mathcal{P}}^{[0,1]}$ degenerates to the standard M_1 metric $\boldsymbol{d}_{M_1}$ on $\mathbb{D} = \mathbb{D}([0,1], \mathbb{R})$. □

Appendix E: Sample Path Large Deviations for Lévy Processes with $\mathcal{MHRV}$ Increments (General Case)

In this section, we provide an analog of Theorem 3.2—the sample path LDPs for Lévy Processes with $\mathcal{MHRV}$ Increments—under relaxed conditions. Such developments allow us to extend Theorem 3.2 beyond the context of Hawkes processes considered in the main paper and to cover a broader class of Lévy processes whose increments exhibit a more flexible form of multivariate hidden regular variation.

Specifically, we work with the following condition regarding the Lévy measure $\nu(\cdot)$ of the d-dimensional Lévy process $\boldsymbol{L}(t)$.

Assumption 2 (**$\mathcal{MHRV}$ increments in Lévy Processes (General Case)**). *The Lévy measure $\nu(\cdot)$ of $\boldsymbol{L}(t)$ is supported on $\mathbb{R}^d_+$ and satisfies $\nu \in \mathcal{MHRV}\Big((\bar{\boldsymbol{s}}_j)_{j\in[m]}, \big(\alpha(\boldsymbol{j})\big)_{\boldsymbol{j}\subseteq[m]}, (\lambda_{\boldsymbol{j}})_{\boldsymbol{j}\in\mathcal{P}_m}, (\mathbf{C}_{\boldsymbol{j}})_{\boldsymbol{j}\in\mathcal{P}_m}\Big)$ under*

- *some positive integer $m \in [d]$,*
- *a collection of vectors $(\bar{\boldsymbol{s}}_j)_{j\in[m]}$ in $\mathbb{R}^d_+$ that are linearly independent,*
- *a collection of real numbers $(\alpha(\boldsymbol{j}))_{\boldsymbol{j}\subseteq[m]}$ that are strictly monotone w.r.t. $\boldsymbol{j}$ (see (5)) with $\alpha(\emptyset) = 0$, $\alpha(\boldsymbol{j}) > 1\ \forall \boldsymbol{j} \in \mathcal{P}_m$, and $\alpha(\boldsymbol{j}) \neq \alpha(\boldsymbol{j}')\ \forall \boldsymbol{j}, \boldsymbol{j}' \in \mathcal{P}_m$ with $\boldsymbol{j} \neq \boldsymbol{j}'$;*
- *regularly varying functions $\lambda_{\boldsymbol{j}}(n) \in \mathcal{RV}_{-\alpha(\boldsymbol{j})}(n)$ (for each $\boldsymbol{j} \in \mathcal{P}_m$),*
- *Borel measures $(\mathbf{C}_{\boldsymbol{j}})_{\boldsymbol{j}\subseteq[m]}$ where, for each $\boldsymbol{j} \in \mathcal{P}_m$, the measure $\mathbf{C}_{\boldsymbol{j}}$ is supported on $\mathbb{R}^d(\boldsymbol{j})$, such that $\mathbf{C}_{\boldsymbol{j}}(A) < \infty$ holds for any $\epsilon > 0$ and Borel set $A \subseteq \mathbb{R}^d_+$ bounded away from $\bar{\mathbb{R}}^d(\boldsymbol{j}, \epsilon)$ (see (8)).*

Throughout this section, we denote the basis by $\bar{\mathbf{S}} = \{\bar{\boldsymbol{s}}_j : j \in [m]\}$ and the collection of tail indices by $\boldsymbol{\alpha} = \big(\alpha(\boldsymbol{j})\big)_{\boldsymbol{j}\subseteq[m]}$. Besides, we adopt the simplified notations $\mathbb{R}^d(\boldsymbol{j}) = \mathbb{R}^d(\boldsymbol{j}; \bar{\mathbf{S}})$, $\bar{\mathbb{R}}^d(\boldsymbol{j}, \epsilon) = \bar{\mathbb{R}}^d(\boldsymbol{j}, \epsilon; \bar{\mathbf{S}})$, and $\bar{\mathbb{R}}^d_{\leqslant}(\boldsymbol{j}, \epsilon) = \bar{\mathbb{R}}^d_{\leqslant}(\boldsymbol{j}, \epsilon; \bar{\mathbf{S}}, \boldsymbol{\alpha})$ in Section 2.2. Compared to Assumption 5, the key differences in Assumption 2 are that

- we drop the condition (22) on the rate functions $\lambda_{\boldsymbol{j}}(\cdot)$;
- we now consider a weaker $\mathcal{MHRV}$ condition (instead of the $\mathcal{MHRV}^*$ condition in Assumption 5); also, the number of base vectors $\bar{s}_j$ considered is not fixed as d.

When imposing the more flexible Assumption 2, one potential issue is that we remain agnostic about the tail behaviors outside of the cone $\mathbb{R}^d([m])$. This potential lack of knowledge about the hidden regular variation in certain regions is reflected in Theorem E.1—the main result of this section—through the following definition:

$$\hat{\mathbb{R}}^d(\boldsymbol{j}, \epsilon; \bar{\mathbf{S}}) \stackrel{\text{def}}{=} \begin{cases} \bar{\mathbb{R}}(\boldsymbol{j}, \epsilon; \bar{\mathbf{S}}) & \text{if } \boldsymbol{j} \in \mathcal{P}_m,\ \boldsymbol{j} \neq [m] \\ \mathbb{R}^d_+ & \text{if } \boldsymbol{j} = [m] \end{cases}. \tag{152}$$

In other words, we only modify the definition $\bar{\mathbb{R}}^d(\boldsymbol{j},\epsilon;\bar{\mathbf{S}})$ to include the entire domain $\mathbb{R}^d_+$. We also write $\hat{\mathbb{R}}^d(\boldsymbol{j},\epsilon) \stackrel{\text{def}}{=} \hat{\mathbb{R}}^d(\boldsymbol{j},\epsilon;\bar{\mathbf{S}})$ as there will be no ambiguity about $\bar{\mathbf{S}}$ in this section.

We introduce a few notions that are required for the statement of Theorem E.1. Recall that we use notations like $\boldsymbol{k} = (k_i)_{i\in\mathcal{J}} \in A^{\mathcal{J}}$ to denote vectors of length $|\mathcal{J}|$ with each coordinate taking values in A and indexed by elements in $\mathcal{J}$. Besides, recall that we use $\mathcal{P}_m$ to denote the collection of all *non-empty* subsets of $[m]$. Analogous to (27), given $\boldsymbol{x}\in\mathbb{R}^d$, $\boldsymbol{\mathcal{K}} = (\mathcal{K}_{\boldsymbol{j}})_{\boldsymbol{j}\in\mathcal{P}_m} \in \mathbb{Z}_+^{\mathcal{P}_m}\setminus\{\mathbf{0}\}$, we define

$$\hat{\mathbb{D}}^{\epsilon}_{\boldsymbol{\mathcal{K}};\boldsymbol{x}}[0,T] \stackrel{\text{def}}{=} \Bigg\{ \boldsymbol{x}\mathbf{1} + \sum_{\boldsymbol{j}\in\mathcal{P}_m}\sum_{k=1}^{\mathcal{K}_{\boldsymbol{j}}} \boldsymbol{w}_{\boldsymbol{j},k}\mathbb{I}_{[t_{\boldsymbol{j},k},T]} : \\ t_{\boldsymbol{j},k}\in(0,T] \text{ and } \boldsymbol{w}_{\boldsymbol{j},k}\in\hat{\mathbb{R}}^d(\boldsymbol{j},\epsilon)\ \forall \boldsymbol{j}\in\mathcal{P}_m,\ k\in[\mathcal{K}_{\boldsymbol{j}}];\ t_{\boldsymbol{j},k}\neq t_{\boldsymbol{j}',k'}\ \forall(\boldsymbol{j},k)\neq(\boldsymbol{j}',k') \Bigg\}, \tag{153}$$

i.e., the set of all paths ξ of the form

$$\xi(t) = \boldsymbol{x}t + \sum_{\boldsymbol{j}\in\mathcal{P}_m}\sum_{k=1}^{\mathcal{K}_{\boldsymbol{j}}} \boldsymbol{w}_{\boldsymbol{j},k}\mathbb{I}_{[t_{\boldsymbol{j},k},T]}(t), \qquad \forall t\in[0,T], \tag{154}$$

where each $\boldsymbol{w}_{\boldsymbol{j},k}$ belongs to the cone $\hat{\mathbb{R}}^d(\boldsymbol{j},\epsilon)$, and any two numbers in the sequence $(t_{\boldsymbol{j},k})_{\boldsymbol{j}\in\mathcal{P}_m,\ k\in[\mathcal{K}_{\boldsymbol{j}}]}$ are not equal (i.e., the arrival times of jumps would not coincide). Next, define

$$\hat{c}(\boldsymbol{\mathcal{K}}) \stackrel{\text{def}}{=} \sum_{\boldsymbol{j}\in\mathcal{P}_m} \mathcal{K}_{\boldsymbol{j}}\cdot\Big(\alpha(\boldsymbol{j})-1\Big), \qquad \forall \boldsymbol{\mathcal{K}} = (\mathcal{K}_{\boldsymbol{j}})_{\boldsymbol{j}\in\mathcal{P}_m}\in\mathbb{Z}_+^{\mathcal{P}_m}. \tag{155}$$

Note that this agrees with $\check{c}(\cdot)$ in (29), and we repeat the definition here so that $\hat{c}(\boldsymbol{\mathcal{K}})$ is consistent with other notations introduced in this section. Given $\boldsymbol{x}\in\mathbb{R}^d$, $\epsilon\geq 0$, $\boldsymbol{\mathcal{K}} = (\mathcal{K}_{\boldsymbol{j}})_{\boldsymbol{j}\in\mathcal{P}_m}\in\mathbb{Z}_+^{\mathcal{P}_m}\setminus\{\mathbf{0}\}$, let

$$\hat{\mathbb{D}}^{\epsilon}_{\leqslant\boldsymbol{\mathcal{K}};\boldsymbol{x}}[0,T] \stackrel{\text{def}}{=} \bigcup_{\substack{\boldsymbol{\mathcal{K}}'\in\mathbb{Z}_+^{\mathcal{P}_m}:\\ \boldsymbol{\mathcal{K}}'\neq\boldsymbol{\mathcal{K}},\ \hat{c}(\boldsymbol{\mathcal{K}}')\leq\hat{c}(\boldsymbol{\mathcal{K}})}} \hat{\mathbb{D}}^{\epsilon}_{\boldsymbol{\mathcal{K}}';\boldsymbol{x}}[0,T]. \tag{156}$$

Meanwhile, given $\boldsymbol{\mathcal{K}} = (\mathcal{K}_{\boldsymbol{j}})_{\boldsymbol{j}\in\mathcal{P}_m}\in\mathbb{Z}_+^{\mathcal{P}_m}\setminus\{\mathbf{0}\}$, let

$$\hat{\lambda}_{\boldsymbol{\mathcal{K}}}(n) \stackrel{\text{def}}{=} \prod_{\boldsymbol{j}\in\mathcal{P}_m}\Big(n\lambda_{\boldsymbol{j}}(n)\Big)^{\mathcal{K}_{\boldsymbol{j}}}, \tag{157}$$

where $\lambda_{\boldsymbol{j}}(n)\in\mathcal{RV}_{-\alpha(\boldsymbol{j})}(n)$ is the rate function of the $\mathcal{MHRV}$ condition in Assumption 2. Then, by the definition of $\hat{c}(\cdot)$ in (155), we have $\hat{\lambda}_{\boldsymbol{\mathcal{K}}}(n)\in\mathcal{RV}_{-\hat{c}(\boldsymbol{\mathcal{K}})}(n)$. Lastly, analogous to (30), given $\boldsymbol{x}\in\mathbb{R}^d$ and $\boldsymbol{\mathcal{K}} = (\mathcal{K}_{\boldsymbol{j}})_{\boldsymbol{j}\in\mathcal{P}_m}\in\mathbb{Z}_+^{\mathcal{P}_m}\setminus\{\mathbf{0}\}$, we define the Borel measure

$$\hat{\mathbf{C}}^{[0,T]}_{\boldsymbol{\mathcal{K}};\boldsymbol{x}}(\,\cdot\,) \stackrel{\text{def}}{=} \frac{1}{\prod_{\boldsymbol{j}\in\mathcal{P}_m}\mathcal{K}_{\boldsymbol{j}}!} \\ \cdot\int \mathbb{I}\Bigg\{\boldsymbol{x}\mathbf{1} + \sum_{\boldsymbol{j}\in\mathcal{P}_m}\sum_{k\in[\mathcal{K}_{\boldsymbol{j}}]} \boldsymbol{w}_{\boldsymbol{j},k}\mathbb{I}_{[t_{\boldsymbol{j},k},T]}\in\,\cdot\,\Bigg\} \bigtimes_{\boldsymbol{j}\in\mathcal{P}_m}\bigtimes_{k\in[k_{\boldsymbol{j}}]}\Big((\mathbf{C}_{\boldsymbol{j}}\times\mathcal{L}_{(0,T)})\Big(d(\boldsymbol{w}_{\boldsymbol{j},k},t_{\boldsymbol{j},k})\Big)\Big), \tag{158}$$

where the $\mathbf{C}_{\boldsymbol{j}}$'s are the limiting measures for the $\mathcal{MHRV}$ condition of Assumption 2, $\mathcal{L}_{(0,T)}$ is the Lebesgue measure on the interval $(0,T)$, and we use $\nu_1 \times \nu_2$ to denote the product measure of ν_1, ν_2.

Let $\boldsymbol{\mu}_{\boldsymbol{L}}$ and $\bar{\boldsymbol{L}}_n^{[0,T]}$ be defined as in (24)–(25). We are now ready to state Theorem E.1.

Theorem E.1. *Let Assumption 2 hold. Let $T \in (0,\infty)$, $\mathcal{K} = (\mathcal{K}_{\boldsymbol{j}})_{\boldsymbol{j}\in\mathcal{P}_m} \in \mathbb{Z}_+^{\mathcal{P}_m} \setminus \{\mathbf{0}\}$, and let B be a Borel set of $\mathbb{D}[0,T]$ equipped with the Skorokhod J_1 topology. Suppose that B is bounded away from $\hat{\mathbb{D}}^{\epsilon}_{\leqslant\mathcal{K};\boldsymbol{\mu}_{\boldsymbol{L}}}[0,T]$ under $\boldsymbol{d}_{J_1}^{[0,T]}$ for some (and hence all) $\epsilon > 0$ small enough. Then,*

$$\hat{\mathbf{C}}^{[0,T]}_{\mathcal{K};\boldsymbol{\mu}_{\boldsymbol{L}}}(B^{\circ}) \le \liminf_{n\to\infty} \frac{\mathbf{P}\big(\bar{\boldsymbol{L}}_n^{[0,T]} \in B\big)}{\hat{\lambda}_{\mathcal{K}}(n)} \le \limsup_{n\to\infty} \frac{\mathbf{P}\big(\bar{\boldsymbol{L}}_n^{[0,T]} \in B\big)}{\hat{\lambda}_{\mathcal{K}}(n)} \le \hat{\mathbf{C}}^{[0,T]}_{\mathcal{K};\boldsymbol{\mu}_{\boldsymbol{L}}}(B^{-}) < \infty.$$

In other words, Theorem E.1 addresses the general case where a more flexible (and hence less detailed) description of the multivariate hidden regular variation (as in Assumption 2) for the Lévy measure $\nu(\cdot)$ is available, at the cost of slightly weaker characterizations of sample path large deviations when compared to Theorem 3.2. In particular, without conditions like (22), we do not know if the rate functions $\hat{\lambda}_{\mathcal{K}}(n)$ are comparable (as $n \to \infty$) for different $\mathcal{K}$ if their tail indices $\hat{c}(\mathcal{K})$ are equal; as a result, Theorem E.1 does not allow the merging of multiple cases corresponding to different $\mathcal{K}$ with equal tail indices $\hat{c}(\mathcal{K})$. Besides, the modification from $\bar{\mathbb{R}}^d(\boldsymbol{j},\epsilon)$ to $\hat{\mathbb{R}}^d(\boldsymbol{j},\epsilon)$ in (152) enlarges the set $\hat{\mathbb{D}}^{\epsilon}_{\mathcal{K};\boldsymbol{\mu}_{\boldsymbol{L}}}[0,T]$ in (153) (when compared to $\check{\mathbb{D}}^{\epsilon}_{\mathcal{K};\boldsymbol{\mu}_{\boldsymbol{L}}}[0,T]$ in (27)). This generally leads to fewer sets B that are bounded away from $\hat{\mathbb{D}}^{\epsilon}_{\leqslant\mathcal{K};\boldsymbol{\mu}_{\boldsymbol{L}}}[0,T]$, and thus fewer combinations of rare events and the rates $\mathcal{K}$ to which Theorem E.1 can be applied.

The proof of Theorem E.1 is almost identical to that of Theorem 3.2 stated in Section C, and we provide a proof sketch below for the sake of completeness.

Proof Sketch for Theorem E.1. Analogous to Theorem 3.2, w.l.o.g. we consider the case where $T = 1$ and $\boldsymbol{\mu}_{\boldsymbol{L}} = \mathbf{0}$, and adopt the notations $\bar{\boldsymbol{L}}_n = \bar{\boldsymbol{L}}_n^{[0,1]}$, $\mathbb{D} = \mathbb{D}\big([0,1],\mathbb{R}^d\big)$, and $\boldsymbol{d}_{J_1} = \boldsymbol{d}_{J_1}^{[0,1]}$. Again, the proof relies on the notion of asymptotic equivalence for $\mathbb{M}$-convergence. By Lemma 2.3, it suffices to verify the following two conditions:

(i) For each $\mathcal{K} = (\mathcal{K}_{\boldsymbol{j}})_{\boldsymbol{j}\in\mathcal{P}_m} \in \mathbb{Z}_+^{\mathcal{P}_m} \setminus \{\mathbf{0}\}$ and $\Delta > 0$,

$$\lim_{n\to\infty} \mathbf{P}\Big(\boldsymbol{d}_{J_1}\big(\hat{\boldsymbol{L}}_n^{>\delta}, \bar{\boldsymbol{L}}_n\big) > \Delta\Big) \Big/ \hat{\lambda}_{\mathcal{K}}(n) = 0, \qquad \forall \delta > 0 \text{ small enough},$$

where $\hat{\boldsymbol{L}}_n^{>\delta}$ is defined in (31)–(33).

(ii) Let $\mathcal{K} = (\mathcal{K}_{\boldsymbol{j}})_{\boldsymbol{j}\in\mathcal{P}_m} \in \mathbb{Z}_+^{\mathcal{P}_m} \setminus \{\mathbf{0}\}$ and $\epsilon > 0$. Let $f : \mathbb{D} \to [0,\infty)$ be bounded and continuous (w.r.t. the J_1 topology of $\mathbb{D}$). Suppose that $B = \mathrm{supp}(f)$ is bounded away from $\hat{\mathbb{D}}^{\epsilon}_{\leqslant\mathcal{K};\mathbf{0}}[0,1]$ under $\boldsymbol{d}_{J_1}$. Then, there exists $\delta_0 > 0$ such that for all but countably many $\delta \in (0,\delta_0)$,

$$\lim_{n\to\infty} \mathbf{E}\big[f(\hat{\boldsymbol{L}}_n^{>\delta})\big] \big/ \hat{\lambda}_{\mathcal{K}}(n) = \hat{\mathbf{C}}^{[0,1]}_{\mathcal{K};\mathbf{0}}(f) < \infty.$$

One can see that these two conditions are similar to the statements in Propositions C.2 and C.3. Likewise, these two conditions can be verified by repeating the arguments in Section C.2. The only key difference stems from the definition of the modified cones $\hat{\mathbb{R}}^d(\boldsymbol{j},\epsilon)$ in (152). In particular, any vector in $\mathbb{R}_+^d \setminus \mathbb{R}^d([m])$ would now fall into $\hat{\mathbb{R}}^d([m],\epsilon)$ (under any $\epsilon \ge 0$). Therefore, when keeping track of the large jumps in $\boldsymbol{L}(t)$, there is no need to specifically count those that lie outside of $\mathbb{R}^d([m])$ (i.e.,

the notions of $\tilde{\tau}_n^{>\delta}(k)$ and $\widetilde{K}_n^{>\delta}$ introduced in (41)–(42) are not needed here). Instead, for each $\delta > 0$, $\boldsymbol{j} \in \mathcal{P}_m$, let

$$\hat{\mathbb{R}}^{>\delta}(\boldsymbol{j}) \stackrel{\text{def}}{=} \left\{ \boldsymbol{x} \in \mathbb{R}_+^d : \|\boldsymbol{x}\| > \delta,\ \boldsymbol{x} \in \hat{\mathbb{R}}^d(\boldsymbol{j}, \delta) \setminus \left(\bigcup_{\boldsymbol{j}' \subseteq [m]:\, \boldsymbol{j}' \neq \boldsymbol{j},\ \alpha(\boldsymbol{j}') \leq \alpha(\boldsymbol{j})} \hat{\mathbb{R}}^d(\boldsymbol{j}', \delta) \right) \right\},$$

which adapts the definition of $\bar{\mathbb{R}}^{>\delta}(\boldsymbol{j})$ in (34) to our current setting. Next, for each $\delta > 0$, $\boldsymbol{j} \in \mathcal{P}_m$, and $k \geq 1$, we define (under the convention that $\hat{\tau}_n^{>\delta}(0; \boldsymbol{j}) \equiv 0$)

$$\hat{\tau}_n^{>\delta}(k; \boldsymbol{j}) \stackrel{\text{def}}{=} \inf \left\{ t > \hat{\tau}_n^{>\delta}(k-1; \boldsymbol{j}) :\ \Delta \bar{\boldsymbol{L}}_n(t) \in \hat{\mathbb{R}}^{>\delta}(\boldsymbol{j}) \right\}, \qquad \hat{\boldsymbol{W}}_n^{>\delta}(k; \boldsymbol{j}) \stackrel{\text{def}}{=} \Delta \bar{\boldsymbol{L}}_n \left(\hat{\tau}_n^{>\delta}(k; \boldsymbol{j}) \right).$$

Besides, for each $\delta > 0$, let

$$\hat{K}_n^{>\delta}(\boldsymbol{j}) \stackrel{\text{def}}{=} \max\{k \geq 0 :\ \hat{\tau}_n^{>\delta}(k; \boldsymbol{j}) \leq 1\}, \quad \forall \boldsymbol{j} \in \mathcal{P}_m; \qquad \hat{\boldsymbol{K}}_n^{>\delta} \stackrel{\text{def}}{=} \left(\hat{K}_n^{>\delta}(\boldsymbol{j}) \right)_{\boldsymbol{j} \in \mathcal{P}_m}.$$

In short, we adapt the definitions in (37)–(40) to count each type of large jumps in $\boldsymbol{L}(t)$ based on their locations w.r.t. the (modified) cones $\hat{\mathbb{R}}^d(\boldsymbol{j})$ in (152), which are generated by the basis described in Assumption 2. Then, to prove conditions (i) and (ii) listed above, it only remains to repeat the proofs in Section C. The changes are minimal and, in fact, the arguments are now even simpler. For instance, when adapting Lemma C.6, as noted above there is no need to consider $\tilde{K}_n^{>\delta}$ anymore. Besides, in Lemmas C.7 and C.8, the notion of allocation $\mathbb{A}(\boldsymbol{k})$ is not needed (as noted right after Theorem E.1, we do not consider merging the cases corresponding to different $\mathcal{K}$ here), so it suffices to fix some $\mathcal{K} = (\mathcal{K}_{\boldsymbol{j}})_{\boldsymbol{j} \in \mathcal{P}_m} \in \mathbb{Z}_+^{\mathcal{P}_m} \setminus \{\boldsymbol{0}\}$ in the proof. The arguments are almost identical, and we omit the details here to avoid repetition. □

Appendix F: Theorem Tree

Theorem Tree of Theorem 3.2

- Theorem 3.2
  - Lemma 2.3
  - Proposition C.2
    - Lemma C.4
    - Lemma C.5
    - Lemma C.6
  - Proposition C.3
    - Lemma C.1
    - Lemma C.6
    - Lemma C.7
    - Lemma C.8
      - Lemma C.7

Theorem Tree of Theorem 3.3

- Theorem 3.3
  - Lemma 2.3
  - Theorem D.3
    - Theorem 3.2
    - Theorem 2.2
    - Lemma D.1
      - Lemma C.8
    - Lemma D.2
  - Proposition 3.4
    - Theorem 2.6
  - Proposition 3.5
    - Lemma D.5
      - Lemma C.4
      - Proposition 3.4
    - Lemma D.6
    - Lemma D.9
      - Lemma D.7
      - Lemma D.8

References

[1] ASMUSSEN, S. and FOSS, S. (2018). Regular variation in a fixed-point problem for single- and multi-class branching processes and queues. *Advances in Applied Probability* **50** 47–61. https://doi.org/10.1017/apr.2018.69

[2] BLANCHET, J., LAEVEN, R. J. A., WANG, X. and ZWART, B. (2025). Tail asymptotics of cluster sizes in multivariate heavy-tailed Hawkes processes. Preprint. Available at arXiv: 2503.01004.

[3] DALEY, D. J. and VERE-JONES, D. (2003). *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods, Volume II: General Theory and Structure*. Springer.

[4] HAWKES, A. G. and OAKES, D. (1974). A cluster process representation of a self-exciting process. *Journal of Applied Probability* **11** 493–503. https://doi.org/10.2307/3212693

[5] KEVEI, P. and WIANDT, P. (2021). Moments of the stationary distribution of subcritical multitype Galton-Watson processes with immigration. *Statistics & Probability Letters* **173** 109067. https://doi.org/10.1016/j.spl.2021.109067

[6] LINIGER, T. (2009). Multivariate Hawkes processes, Doctoral Thesis, ETH Zurich, Zürich Diss., Eidgenössische Technische Hochschule ETH Zürich, Nr. 18403, 2009. https://doi.org/10.3929/ethz-a-006037599

[7] PROTTER, P. E. (2005). *Stochastic Integration and Differential Equations*. Springer Berlin.

[8] REYNAUD-BOURET, P. and ROY, E. (2007). Some non asymptotic tail estimates for Hawkes processes. *Bulletin of the Belgian Mathematical Society - Simon Stevin* **13** 883–896. https://doi.org/10.36045/bbms/1170347811

[9] WANG, X. and RHEE, C.-H. (2024). Large deviations and metastability analysis for heavy-tailed dynamical systems. Preprint. Available at arXiv: 2307.03479.