Online Companion (Appendix)

This online companion is organized as follows. Section A extends the algorithmic framework to the context of barrier option pricing, Section B collects details of the numerical experiments, and Section C provides details of Remark 8.

Appendix A: Barrier Option Pricing

A.1. Problem Setting This section considers the estimation of probabilities $P(A_n)$ with $A_n = \{\bar{X}_n \in A\}$ and

$$A \stackrel{\text{def}}{=} \{\xi \in \mathbb{D} : \ \xi(1) \leq -b, \ \sup_{t \leq 1} \xi(t) + ct \geq a\},$$

which corresponds to rare-event simulation in the context of down-and-in option. Here, we assume that $a, \ b > 0$ and $c < a$. We consider the two-sided case in Assumption 1. That is, $X(t)$ is a centered Lévy process with Lévy measures ν, and there exists some $\alpha, \ \alpha' > 1$ such that $\nu[x, \infty) \in \mathcal{RV}_{-\alpha}(x)$ and $\nu(-\infty, -x] \in \mathcal{RV}_{-\alpha'}(x)$ as $x \to \infty$. Also, we impose an alternative version of Assumption 2 throughout. Let $X^{(-z,z)}(t)$ be the Lévy process with with generating triplet $(c_X, \sigma, \nu|_{(-z,z)})$. That is, $X^{(-z,z)}(t)$ is a modulated version of X where all jumps with size larger than z are removed.

ASSUMPTION 4. *There exist $z_0, \ C, \ \lambda > 0$ such that*

$$\mathbf{P}\big(X^{(-z,z)}(t) \in [x, x+\delta]\big) \leq \frac{C\delta}{t^\lambda \wedge 1} \qquad \forall z \geq z_0, \ t > 0, \ x \in \mathbb{R}, \ \delta > 0.$$

A.2. Importance Sampling Algorithm Below, we present the design of the importance sampling algorithm. For any $\xi \in \mathbb{D}$ and $t \in (0, 1]$, let $\Delta\xi(t) = \xi(t) - \xi(t-)$ be the discontinuity in ξ at time t, and we set $\Delta\xi(0) \equiv 0$. Let

$$B^\gamma = \Big\{\xi \in \mathbb{D} : \ \#\{t \in [0,1] : \ \Delta\xi(t) \geq \gamma\} \geq 1, \ \#\{t \in [0,1] : \ \Delta\xi(t) \leq -\gamma\} \geq 1\Big\}$$

and let $B_n^\gamma = \{\bar{X}_n \in B_n^\gamma\}$. Intuitively speaking, on event B_n^γ there is at least one upward and one downward "large" jump in $\bar{X}_n$, where $\gamma > 0$ is understood as the threshold for jump sizes to be considered "large".

Fix some $w \in (0, 1)$, and let

$$\mathbf{Q}_n(\cdot) = w\mathbf{P}(\cdot) + (1-w)\mathbf{P}(\ \cdot \ | B_n^\gamma).$$

The algorithm samples

$$L_n = Z_n \frac{d\mathbf{P}}{d\mathbf{Q}_n} = \frac{Z_n}{w + \frac{1-w}{\mathbf{P}(B_n^\gamma)}\mathbf{I}_{B_n^\gamma}}$$

under $\mathbf{Q}_n$. Now, we discuss the design of Z_n to ensure the strong efficiency of L_n. Analogous to the decomposition in (18), let

$$J_n(t) = \sum_{s\in[0,t]} \Delta X(s)\mathbf{I}\big(|\Delta X(s)| \geq n\gamma\big),$$

$$\Xi_n(t) = X(t) - J_n(t) = X(t) - \sum_{s\in[0,t]} \Delta X(s)\mathbf{I}\big(|\Delta X(s)| \geq n\gamma\big).$$

Let $\bar{J}_n(t) = \frac{1}{n}J_n(nt)$, $\bar{J}_n = \{\bar{J}_n(t):\ t \in [0,1]\}$, $\bar{\Xi}_n(t) = \frac{1}{n}\Xi_n(nt)$, and $\bar{\Xi}_n = \{\bar{\Xi}_n(t):\ t\in[0,1]\}$. Meanwhile, set

$$M_c(t) \stackrel{\text{def}}{=} \sup_{s\leq t} X(s) + cs, \qquad Y^*_{n;c} \stackrel{\text{def}}{=} \mathbf{I}\big(M_c(n) \geq na,\ X(n) \leq -nb\big),$$

We have $\mathbf{I}_{A_n} = Y^*_{n;c}$. Under the convention $\hat{Y}_n^{-1} \equiv 0$, consider estimators Z_n of form

$$Z_n = \sum_{m=0}^{\tau} \frac{\hat{Y}^m_{n;c} - \hat{Y}^{m-1}_{n;c}}{\mathbf{P}(\tau \geq m)} \tag{81}$$

where τ is Geom(ρ) for some $\rho\in(0,1)$ and is independent of everything else. Analogous to Proposition 1, the following result provides sufficient conditions on $\hat{Y}^m_{n;c}$ for L_n to attain strong efficiency.

PROPOSITION 4. *Let $C_0 > 0$, $\rho_0 \in (0,1)$, $\mu > \alpha + \alpha' - 2$, and $\bar{m} \in \mathbb{N}$. Suppose that*

$$\mathbf{P}\Big(Y^*_{n;c} \neq \hat{Y}^m_{n;c}\ \Big|\ \mathcal{D}^+(\bar{J}_n) = k,\ \mathcal{D}^-(\bar{J}_n) = k'\Big) \leq C_0\rho_0^m\cdot(k+k'+1) \qquad \forall k,\ k'\geq 0,\ n\geq 1,\ m\geq\bar{m} \tag{82}$$

where $\mathcal{D}^+(\xi)$ and $\mathcal{D}^-(\xi)$ count the number of discontinuities of positive and negative sizes in ξ, respectively. Besides, suppose that for all $\Delta\in(0,1)$,

$$\mathbf{P}\Big(Y^*_{n;c} \neq \hat{Y}^m_{n;c},\ \bar{X}_n \notin A^\Delta\ \Big|\ \mathcal{D}^+(\bar{J}_n) = 0 \text{ or } \mathcal{D}^-(\bar{J}_n) = 0\Big) \leq \frac{C_0\rho_0^m}{\Delta^2 n^\mu} \qquad \forall n\geq 1,\ m\geq 0 \tag{83}$$

*where $A^\Delta = \big\{\xi\in\mathbb{D} : \sup_{t\in[0,1]}\xi(t) + ct \geq a - \Delta,\ \xi(1)\leq -b\big\}$. Then given $\rho\in(\rho_0,1)$, there exists some $\bar{\gamma} = \bar{\gamma}(\rho)\in(0,b)$ such that for all $\gamma\in(0,\bar{\gamma})$, the estimators $(L_n)_{n\geq1}$ are **unbiased and strongly efficient** for $\mathbf{P}(A_n) = \mathbf{P}(\bar{X}_n\in A)$ under the importance sampling distribution $\mathbf{Q}_n$.*

The proof is almost identical to that of Proposition 1. In particular, the proof requires that

$$\mathbf{P}(A_n) = \boldsymbol{O}\big(n\nu[n,\infty)\cdot n\nu(-\infty,-n]\big)$$

and that, for any $\beta > 0$, it holds for all γ small enough that

$$\mathbf{P}(A_n^\Delta \setminus B_n^\gamma) = \boldsymbol{o}(n^\beta)$$

where $A_n^\Delta = \{\bar{X}_n \in A^\Delta\}$. These can be obtained directly using sample path large deviations for heavy-tailed Lévy processes in Result 2. The Proposition 4 is then established by repeating the arguments in Proposition 1 using Result 4 for the debiased multilevel Monte Carlo technique.

A.3. Construction of $\hat{Y}^m_{n;c}$ Next, we describe the construction of $\hat{Y}^m_{n;c}$ that can satisfy the conditions in Proposition 4. Specifically, we consider the case where ARA is involved. Let

$$\Xi_{n;c}(t) \stackrel{\text{def}}{=} \Xi_n(t) + ct.$$

Under both $\mathbf{P}$ and $\mathbf{Q}_n$, $\Xi_{n;c}(t)$ admits the law of a Lévy process with generating triplet $(c_X + c, \sigma, \nu|_{(-n\gamma,n\gamma)})$. This leads to the Lévy-Ito decomposition

$$\begin{aligned}\Xi_{n,c}(t) \stackrel{d}{=} (c_X + c)t + \sigma B(t) &+ \underbrace{\sum_{s\le t}\Delta X(s)\mathbf{I}\Big(\Delta X(s) \in (-n\gamma,-1]\cup[1,n\gamma)\Big)}_{\stackrel{\text{def}}{=} J_{n,-1}(t)} \\ &+ \sum_{m\ge 0}\Bigg[\underbrace{\sum_{s\le t}\Delta X(s)\mathbf{I}\Big(|\Delta X(s)| \in [\kappa_{n,m},\kappa_{n,m-1})\Big) - t\cdot\nu\Big((-\kappa_{n,m-1},-\kappa_{n,m}]\cup[\kappa_{n,m},\kappa_{n,m-1})\Big)}_{\stackrel{\text{def}}{=} J_{n,m}(t)}\Bigg]\end{aligned}$$

with $\kappa_{n,m}$ defined in (32). Besides, let $\bar{\sigma}^2(\cdot)$ be defined as in (34). For each $n \ge 1$ and $m \ge 0$, consider the approximation

$$\breve{\Xi}^m_{n;c}(t) \stackrel{\text{def}}{=} (c_X + c)t + \sigma B(t) + \sum_{q=-1}^{m} J_{n,q}(t) + \sum_{q\ge m+1}\sqrt{\bar{\sigma}^2(\kappa_{n,q-1}) - \bar{\sigma}^2(\kappa_{n,q})}\cdot W^q(t)$$

where $(W^m)_{m\ge 1}$ is a sequence of iid copies of standard Brownian motions independent of everything else.

Next, we discuss how to apply SBA and construct approximators $\hat{Y}^m_{n;c}$'s in (81). Let $\zeta_k(t) = \sum_{i=1}^k z_i \mathbf{I}_{[u_i,n]}(t)$ be a piece-wise step function with k jumps over $(0,n]$, where $0 < u_1 < u_2 < \ldots < u_k \le n$, and $z_i \neq 0$ for each $i \in [k]$. Recall that the jump times in ζ_k leads to a partition of $[0,n]$ of $(I_i)_{i\in[k+1]}$ defined in (24). For any I_i, let the sequence $l^{(i)}_j$'s be defined as in (27)–(28). Conditioning on $(l^{(i)}_j)_{j\geq 1}$, one can then sample $\xi^{(i),m}_{j;c}, \xi^{(i)}_{j;c}$ using

$$(\xi^{(i)}_{j;c}, \xi^{(i),0}_{j;c}, \xi^{(i),1}_{j;c}, \xi^{(i),2}_{j;c}, \dots) \stackrel{d}{=} \Big(\Xi_{n;c}(l^{(i)}_j),\ \breve{\Xi}^0_{n;c}(l^{(i)}_j),\ \breve{\Xi}^1_{n;c}(l^{(i)}_j),\ \breve{\Xi}^2_{n;c}(l^{(i)}_j), \dots\Big).$$

The coupling in (10) then implies

$$\begin{aligned}
&\Big(\Xi_{n;c}(u_i) - \Xi_{n;c}(u_{i-1}),\ \sup_{t\in I_i}\Xi_{n;c}(t) - \Xi_{n;c}(u_{i-1}),\ \breve{\Xi}^0_{n;c}(u_i) - \breve{\Xi}^0_{n;c}(u_{i-1}),\ \sup_{t\in I_i}\breve{\Xi}^0_{n;c}(t) - \breve{\Xi}^0_{n;c}(u_{i-1}),\\
&\qquad\qquad \breve{\Xi}^1_{n;c}(u_i) - \breve{\Xi}^1_{n;c}(u_{i-1}),\ \sup_{t\in I_i}\breve{\Xi}^1_{n;c}(t) - \breve{\Xi}^1_{n;c}(u_{i-1}), \dots\Big)\\
&\stackrel{d}{=} \Big(\sum_{j\geq 1}\xi^{(i)}_{j;c},\ \sum_{j\geq 1}(\xi^{(i)}_{j;c})^+,\ \sum_{j\geq 1}\xi^{(i),0}_{j;c},\ \sum_{j\geq 1}(\xi^{(i),0}_{j;c})^+,\ \sum_{j\geq 1}\xi^{(i),1}_{j;c},\ \sum_{j\geq 1}(\xi^{(i),1}_{j;c})^+, \dots\Big).
\end{aligned}$$

Now, we define

$$\hat{M}^{(i),m}_{n;c}(\zeta_k) = \sum_{j=1}^{m+\lceil \log_2(n^d)\rceil} (\xi^{(i),m}_{j;c})^+$$

as an approximation to $M^{(i),*}_{n;c}(\zeta_k) = \sup_{t\in I_i}\Xi_{n;c}(t) - \Xi_{n;c}(u_{i-1}) = \sum_{j\geq 1}(\xi^{(i)}_{j;c})^+$. Now, set

$$\hat{Y}^m_{n;c}(\zeta_k) = \Big[\max_{i\in[k+1]} \mathbf{I}\Big(\sum_{q=1}^{i-1}\sum_{j\geq 0}\xi^{(q),m}_{j;c} + \sum_{q=1}^{i-1} z_q + \hat{M}^{(i),m}_{n;c}(\zeta_k) \geq na\Big)\Big]\cdot \mathbf{I}\Big(\sum_{q=1}^{k+1}\sum_{j\geq 0}\xi^{(q),m}_{j;c} + \sum_{q=1}^{k} z_q - cn \leq -nb\Big).$$

In (81), we plug in $\hat{Y}^m_{n;c} = \hat{Y}^m_{n;c}(J_n)$.

The proof of the strong efficiency is almost identical to that of Theorem 2. The only major difference is that in Lemma 7, we apply Assumption 4 instead of Assumption 2.

Appendix B: Details of Numerical Experiments This section provides the numerical details for the experiment results shown in Figures 1 and 2.

Appendix C: Details of Remark 8: Biased estimators with Prescribed Relative Accuracy While this paper focuses on achieving *unbiasedness* using the debiased multilevel Monte Carlo estimators Z_n (20), here we touch upon a different perspective that can be viewed as an analogous of strong efficiency for biased Monte Carlo estimators: we

TABLE 1. Relative errors in Figure 1: Algorithm 2 (first row) versus crude Monte Carlo (second row). Underline: crude MC with early termination (threshold: 24 hours); numbers estimated by the corresponding IS results

n	100	200	300	400	500	600	700	800	900	1000
$\alpha = 1.45$	10.00	11.91	12.69	12.99	13.52	14.38	15.49	14.92	15.31	16.30
	66.90	103.51	135.01	156.87	170.26	205.83	219.65	242.41	254.37	273.86
$\alpha = 1.6$	13.33	15.48	15.78	17.63	19.09	19.78	21.73	19.53	19.62	20.54
	131.33	237.31	309.35	431.55	461.54	$\underline{498.28}$	$\underline{652.77}$	$\underline{705.62}$	$\underline{755.92}$	$\underline{804.06}$
$\alpha = 1.75$	16.48	18.66	22.77	22.92	23.24	24.79	25.44	25.73	28.59	26.89
	276.25	530.92	767.48	$\underline{896.33}$	$\underline{1278.52}$	$\underline{1419.70}$	$\underline{1635.84}$	$\underline{1860.28}$	$\underline{2029.07}$	$\underline{2193.27}$

TABLE 2. Runtime (sec) in Figure 1: Algorithm 2 (first row) versus crude Monte Carlo (second row). Underline: crude MC with early termination (threshold: 24 hours); numbers estimated by the corresponding IS results.

n	100	200	300	400	500	600	700	800	900	1000
$\alpha = 1.45$	20.48	26.72	40.21	44.92	45.24	55.84	73.87	49.42	72.51	86.81
	2.75e+02	1.21e+03	2.96e+03	5.50e+03	8.70e+03	1.30e+04	1.79e+04	2.54e+04	3.29e+04	3.97e+04
$\alpha = 1.6$	40.44	72.08	70.67	87.55	144.28	88.8	131.62	137.37	144.67	148.8
	1.12e+03	6.13e+03	1.73e+04	3.49e+04	6.67e+04	$\underline{8.17e+04}$	$\underline{1.63e+05}$	$\underline{2.22e+05}$	$\underline{2.85e+05}$	$\underline{3.49e+05}$
$\alpha = 1.75$	65.08	106.37	105.08	110.66	174.43	173.06	191.4	213.72	221.63	229.39
	4.02e+03	3.37e+04	8.64e+04	$\underline{1.76e+05}$	$\underline{4.50e+05}$	$\underline{6.63e+05}$	$\underline{1.02e+06}$	$\underline{1.54e+06}$	$\underline{2.06e+06}$	$\underline{2.60e+06}$

TABLE 3. Relative errors in Figure 2: Algorithm 3 (first row) versus crude Monte Carlo (second row). Underline: crude MC with early termination (threshold: 24 hours); numbers estimated by the corresponding IS results.

n	100	200	300	400	500
$\alpha = 1.45$	9.64	12.21	12.0	13.68	14.74
	60.5	104.39	129.9	153.77	193.81
$\alpha = 1.6$	13.23	14.11	16.61	15.78	17.81
	134.67	$\underline{214.77}$	$\underline{356.19}$	$\underline{309.39}$	$\underline{599.62}$
$\alpha = 1.75$	16.27	19.92	22.53	21.58	23.82
	$\underline{229.17}$	$\underline{517.39}$	$\underline{610.4}$	$\underline{899.18}$	$\underline{1162.56}$

Table 4. Runtime (sec) in Figure 2: Algorithm 3 (first row) versus crude Monte Carlo (second row). Underline: crude MC with early termination (threshold: 24 hours); numbers estimated by the corresponding IS results.

n	100	200	300	400	500
$\alpha = 1.45$	4.43e+03	5.02e+03	9.85e+03	7.44e+03	8.32e+03
	9.78e+03	2.09e+04	4.73e+04	5.19e+04	7.32e+04
$\alpha = 1.6$	7.63e+03	8.99e+03	1.69e+04	9.48e+03	2.67e+04
	3.76e+04	<u>1.14e+05</u>	<u>3.18e+05</u>	<u>2.47e+05</u>	<u>8.98e+05</u>
$\alpha = 1.75$	9.97e+03	1.91e+04	1.56e+04	2.16e+04	2.69e+04
	<u>1.25e+05</u>	<u>6.6e+05</u>	<u>9.33e+05</u>	<u>2.09e+06</u>	<u>3.38e+06</u>

discuss the cost of achieving a prescribed level of relative accuracy using the biased estimators $\hat{Y}^m_n \cdot \mathbf{I}_{E_n}$ in (31) or (39) (still generated under the importance sampling distributions $\mathbf{Q}_n$). For concreteness of our discussion, we specialize the criterion of *fully polynomial randomized approximation scheme* (FPRAS) (see, e.g., Definition 2.2 of Adler et al. (2012)) to our setting, and say that a Monte Carlo procedure is FPRAS if there exist $q, q_1, q_2 \in [0,\infty)$ such that, for estimating any $\mathbf{P}(A_n)$, the procedure outputs an estimator that is guaranteed to have at most $\epsilon > 0$ relative error with confidence at least $1-\delta \in (0,1)$, under computational costs of order $\boldsymbol{O}\big(\epsilon^{-q_1}\delta^{-q_2}|\log \mathbf{P}(A_n)|^q\big)$.

- **Non-ARA case (i.e., with exact simulation for marginal distributions of** $X^{<n\gamma}(t)$**).** We start by considering the cost of generating $\hat{Y}^m_n(\zeta_k)$ in (31) given a step function $\zeta_k = \sum_{i=1}^k z_i \mathbf{I}_{[u_i,n]}$. By (30), the cost of generating $\big(\hat{M}^{(i),m}_n(\zeta_k)\big)_{i\in[k+1]}$ is $(k+1)\cdot \boldsymbol{O}\big(m+\log n\big)$. Meanwhile, because $\sum_{j\geq 0}\xi^{(q)}_j \stackrel{d}{=} X^{<n\gamma}(u_q - u_{q-1})$ and generating (31) requires us to sample $\sum_{j\geq 0}\xi^{(q)}_j$ for all $q\in[k+1]$, the related computational cost is $(k+1)$. Next, repeating the calculations in (52), we see that unconditionally (i.e., when considering the randomness of sampling the big jump process ζ_k under $\mathbf{Q}_n$), the (expected) computational cost of generating $\hat{Y}^m_n$ is $\boldsymbol{O}\big(m+\log n\big)$. On the other hand, by the last inequality in (52), we know the existence of some $\rho_0\in(0,1)$ such that (recall that Y^*_n and $\hat{Y}^m_n$ are indicator functions)

$$\left|\mathbf{E}^{\mathbf{Q}_n}[\hat{Y}^m_n \mathbf{I}_{E_n}] - \mathbf{P}(A_n)\right| = \left|\mathbf{E}^{\mathbf{Q}_n}[\hat{Y}^m_n \mathbf{I}_{E_n}] - \mathbf{E}^{\mathbf{Q}_n}[Y^*_n \mathbf{I}_{E_n}]\right| \leq \mathbf{E}^{\mathbf{Q}_n}\left|\hat{Y}^m_n - Y^*_n\right| \leq \frac{1}{w}\mathbf{P}(\hat{Y}^m_n \neq Y^*_n) = \boldsymbol{O}(\rho_0^m) \tag{84}$$

  holds for all m,n, where $w\in(0,1)$ is the fixed constant in the design of IS distribution $\mathbf{Q}_n$ in (16). Also, by the large deviations results in Section 2.2, we have $\mathbf{P}(A_n) = \boldsymbol{\Theta}\big(n^{-l^*\cdot(\alpha-1)}L(n)\big)$ where $L(n)$ is some slowly varying function, $\alpha\in(1,\infty)$ is the heavy-tailed index for the increments in $X(t)$, and l^* is defined in (16). Then, to ensure that $\left|\mathbf{E}^{\mathbf{Q}_n}[\hat{Y}_n^m] - \mathbf{P}(A_n)\right| \leq \epsilon \mathbf{P}(A_n)$, one can pick $m = \mathbf{\Theta}\big((|\log\epsilon| + |\log \mathbf{P}(A_n)|)/\log\rho_0\big) = \mathbf{\Theta}(|\log\epsilon| + \log n)$. Under such m, the cost of generating one copy of $\hat{Y}_n^m \mathbf{I}_{E_n}$ is $\boldsymbol{O}(m + \log n) = \boldsymbol{O}(|\log\epsilon| + \log n)$. Furthermore, the proof of Proposition 1 in Section 6.1 essentially confirms that $\sup_n var^{\mathbf{Q}_n}(\hat{Y}_n^m \mathbf{I}_{E_n}) \big/ \big(\mathbf{P}(A_n)\big)^2 < \infty$ under the choice of $m = \mathbf{\Theta}\big(|\log\epsilon| + \log n\big)$. Then, by repeating the derivation prior to Definition 2.2 of Adler et al. (2012) (which boils down to an application of Chebyshev's inequality), an error smaller than $\epsilon\mathbf{P}(A_n)$ can be ensured with confidence at least $1-\delta$ by averaging R iid samples of $\hat{Y}_n^m \mathbf{I}_{E_n}$, where $R = \mathbf{\Theta}(\epsilon^{-2}\delta^{-1})$. In other words, such an averaged estimator is FPRAS for $\mathbf{P}(A_n)$ with a computational cost $R \cdot \boldsymbol{O}(|\log\epsilon| + \log n) = \boldsymbol{O}(\epsilon^{-2}\delta^{-1}\log n) = \boldsymbol{O}\big(\epsilon^{-2}\delta^{-1}|\log \mathbf{P}(A_n)|\big)$.

- **Practical Challenges when Implementing FPRAS Estimators**. While the order of m and R (w.r.t. n) is clear from the discussion above, determining the exact value of m (and hence the implementation of the FPRAS estimators) can be challenging in practice, as it involves instance-specific quantities that can be hard to estimate. For instance, determining the exact value of the sample size R requires the knowledge of $\sup_n var^{\mathbf{Q}_n}(\hat{Y}_n^m \mathbf{I}_{E_n}) \big/ \big(\mathbf{P}(A_n)\big)^2 < \infty$, which is hard to estimate even for a fixed n: in particular, note that the estimation for the denominator $\mathbf{P}(A_n)$ is hard to justify unless we resort to unbiased estimators proposed in the main paper, which then defeats the purposes of identifying (biased but) FPRAS estimators. Besides, determining m relies on the knowledge of the leading coefficient for the $\boldsymbol{O}(\rho_0^m)$ term in (84), which further requires knowing the exact value of the constant C in Assumption 2 (see, e.g., the proof of Propositions 2 and 3). However, Assumption 2 and the sufficient conditions stated in Section 4 are only able to confirm the existence of such C.

- **ARA case**. Next, we consider the case where the exact simulation of $X^{<n\gamma}(t)$ is not available, and we resort to ARA for approximating the small-jump martingales in the underlying Lévy process with infinite activities (as in our Algorithm 3). That is, by $\hat{Y}_n^m(\zeta_k)$ we refer to the construction in (39), where the truncation threshold for ARA is set as κ^m/n^r for some fixed $r > 0$. As noted above, our proofs in Section 6 simultaneously address the non-ARA and ARA cases. Therefore, many of the bounds stated in the previous bullet point still apply to the ARA case. In particular, we still have: (i) $\left|\mathbf{E}^{\mathbf{Q}_n}[\hat{Y}_n^m \mathbf{I}_{E_n}] - \mathbf{P}(A_n)\right| \leq \epsilon\mathbf{P}(A_n)$ and $\sup_n var^{\mathbf{Q}_n}(\hat{Y}_n^m \mathbf{I}_{E_n}) \big/ \big(\mathbf{P}(A_n)\big)^2 < \infty$ under the choice of $m = \mathbf{\Theta}\big(|\log\epsilon| + \log n\big)$ (more precisely, suppose that for some fixed and large enough $C \in (0,\infty)$, the choice of $m = C\big(|\log\epsilon| + \log n\big)$ would suffice); and (ii) an error smaller than $\epsilon\mathbf{P}(A_n)$ can be ensured with confidence at least $1-\delta$ by averaging $R = \mathbf{\Theta}(\epsilon^{-2}\delta^{-1})$ iid samples of such $\hat{Y}_n^m \mathbf{I}_{E_n}$. However, due to the incorporation of ARA, the computational cost is now dominated by the efforts of generating all jumps in $X(t)$

above the level of κ^m/n^r. Analogous to the computations in Remark 7, we see that the computational cost of generating $\hat{Y}^m_n$, with ARA incorporated, is of order $\boldsymbol{\Omega}\big(n\cdot(\kappa^m/n^r)^{-\beta_-}\big)=\boldsymbol{O}\big(n\cdot(\kappa^m/n^r)^{-\beta_+}\big)$ for any $\beta_+\in(\beta,2)$, $\beta_-\in(0,\beta)$, where $\beta\in(0,2)$ is the Blumenthal-Getoor index in Assumption 1. Then, the total cost of achieving ϵ relative error with confidence $1-\delta$ is

$$\begin{aligned}R\cdot\boldsymbol{O}\Big(n^{1+r\beta_+}\cdot\kappa^{-\beta_+\cdot C(|\log\epsilon|+\log n)}\Big)&=\boldsymbol{O}\Big(\epsilon^{-2}\delta^{-1}\cdot n^{1+r\beta_+}\cdot\kappa^{-\beta_+\cdot C(|\log\epsilon|+\log n)}\Big)\\&=\boldsymbol{O}\Big(\epsilon^{-2-C\beta_+|\log\kappa|}\cdot\delta^{-1}\cdot n^{1+r\beta_++C\beta_+|\log\kappa|}\Big).\end{aligned}\tag{85}$$

- Therefore, the FPRAS criterion is *not satisfied* in the ARA case. In particular, Adler et al. (2012) focuses on light-tailed cases where the counterparts of our term $\log\mathbf{P}(A_n)$ scale polynomially w.r.t. n. In contrast, our heavy-tailed setting dictates that $\log\mathbf{P}(A_n)$ scales logarithmically w.r.t. n; this results in an $\boldsymbol{O}\big(\epsilon^{-q_1}\delta^{-q_2}|\log\mathbf{P}(A_n)|^q\big)=\boldsymbol{O}\big(\epsilon^{-q_1}\delta^{-q_2}\cdot(\log n)^q\big)$ bound on the computational cost in the definition of the fully polynomial randomized approximation scheme that, despite its name, is not even polynomial in n. Indeed, by comparing with (85), we see that this criterion is indeed too stringent when ARA is needed for the simulation of the underlying Lévy process.

Appendix D: Proof for Technical Lemmas in Section 6.2 This section collects the proof for technical Lemmas in Section 6.2. First, we review a useful result regarding the Lévy measure.

Result 5 (Lemma 1 of González Cázares et al. (2022)) *Let ν be the Lévy measure of a Lévy process X. Let $I^p_0(\nu)\stackrel{def}{=}\int_{(-1,1)}|x|^p\nu(dx)$. Suppose that $\beta<2$ for the Blumenthal-Getoor index $\beta\stackrel{def}{=}\inf\{p>0:\ I^p_0(\nu)<\infty\}$. Then*

$$\int_{(-\kappa,\kappa)}x^2\nu(dx)\leq\kappa^{2-\beta_+}I^{\beta_+}_0(\nu)\qquad\forall\kappa\in(0,1],\ \beta_+\in(\beta,2).$$

Next, we provide the proof of Lemmas 3–7.

Proof of Lemma 3 Recall that the generating triplet of X is (c_X,σ,ν) and for the Blumenthal-Getoor index $\beta\stackrel{\text{def}}{=}\inf\{p>0:\int_{(-1,1)}|x|^p\nu(dx)<\infty\}$ we have $\beta<2$; see Assumption 1. Fix some $\beta_+\in(1\vee\beta,2)$ in this proof. We prove the lemma for $C_X\stackrel{\text{def}}{=}\max\Big\{|\sigma|\sqrt{\frac{2}{\pi}}+2\sqrt{I^{\beta_+}_0(\nu)},\ (c_X)^++I^1_+(\nu)+2I^{\beta_+}_0(\nu)\Big\}$, where $(x)^+=x\vee 0$, $I^1_+(\nu)=\int_{[1,\infty)}x\nu(dx)$, and $I^p_0(\nu)=\int_{(-1,1)}|x|^p\nu(dx)$.

Recall that Ξ_n is a Lévy process with generating triplet $(c_X,\sigma,\nu|_{(-\infty,n\gamma)})$. Let $\nu_n\stackrel{\text{def}}{=}\nu|_{(-\infty,n\gamma)}$. It follows from Lemma 2 of González Cázares et al. (2022) (specifically, by setting $t=T$ in equation (26)) that, for all $t>0$ and $n\geq 1$,

$$\mathbf{E}\sup_{s\in[0,t]}\Xi_n(t)\leq\Bigg(|\sigma|\sqrt{\frac{2}{\pi}}+2\sqrt{I^{\beta_+}_0(\nu_n)}\Bigg)\sqrt{t}+\Big((c_X)^++I^1_+(\nu_n)+2I^{\beta_+}_0(\nu_n)\Big)t.\tag{86}$$

Note that $I_0^{\beta_+}(\nu_n) = \int_{(-1,1)} |x|^p \nu_n(dx) = \int_{(-1,1)\cap(-\infty,n\gamma)} |x|^p \nu(dx) \leq I_0^{\beta_+}(\nu)$ and $I_+^1(\nu_n) = \int_{[1,\infty)} x\nu_n(dx) = \int_{[1,\infty)\cap(-\infty,n\gamma)} x\nu(dx) \leq I_+^1(\nu)$. Plugging these two bounds into (86), we conclude the proof. $\square$

Proof of Lemma 4 From the definitions of Ξ_n and $\breve{\Xi}_n^m$ in (33) and (35), respectively, we have $\Xi_n(t) - \breve{\Xi}_n^m(t) \stackrel{d}{=} X^{(-\kappa_{n,m},\kappa_{n,m})}(t) - \bar{\sigma}(\kappa_{n,m})B(t)$, where $X^{(-c,c)}$ is the Lévy process with generating triplet $(0,0,\nu|_{(-c,c)})$, $\kappa_{n,m} = \kappa^m/n^r$, and B is a standard Brownian motion independent of $X^{(-\kappa_{n,m},\kappa_{n,m})}$. In particular, $X^{(-\kappa_{n,m},\kappa_{n,m})}$ is a martingale with variance $var\big[X^{(-\kappa_{n,m},\kappa_{n,m})}(1)\big] = \bar{\sigma}^2(\kappa_{n,m})$; see (34) for the definition of $\bar{\sigma}^2(\cdot)$. Therefore,

$$\begin{aligned}
&\mathbf{P}\Big(\sup_{t\in[0,n]}\Big|\Xi_n(t) - \breve{\Xi}_n^m(t)\Big| > x\Big) \\
&\leq \frac{1}{x^2}\mathbf{E}\Big|X^{(-\kappa_{n,m},\kappa_{n,m})}(n) - \bar{\sigma}(\kappa_{n,m})B(n)\Big|^2 \qquad \text{using Doob's inequality} \\
&= \frac{2n}{x^2}\bar{\sigma}^2(\kappa_{n,m}) \qquad \text{due to the independence between } X^{(-\kappa_{n,m},\kappa_{n,m})} \text{ and } B \\
&\leq \frac{2n}{x^2}\cdot \kappa_{n,m}^{2-\beta_+} I_0^{\beta_+}(\nu) \qquad \text{using Result 5} \\
&= \frac{2I_0^{\beta_+}(\nu)}{x^2}\cdot\frac{n\kappa^{m(2-\beta_+)}}{n^{r(2-\beta_+)}} = \frac{2I_0^{\beta_+}(\nu)}{x^2}\cdot\frac{\kappa^{m(2-\beta_+)}}{n^{r(2-\beta_+)-1}} \qquad \text{due to } \kappa_{n,m} = \kappa^m/n^r.
\end{aligned}$$

To conclude the proof, we set $C = 2I_0^{\beta_+}(\nu) = 2\int_{(-1,1)}\int |x|^{\beta_+}\nu(dx)$. $\square$

Proof of Lemma 5 For notational simplicity, set $k(n) = \lceil \log_2(n^d)\rceil$. Due to $|(x)^+ - (y)^+| \leq |x-y|$,

$$\begin{aligned}
&\mathbf{P}\Big(\Big|\sum_{j=1}^{m+k(n)}\big(\xi_j^{[n]}(l)\big)^+ - \sum_{j=1}^{m+k(n)}\big(\xi_j^{[n],m}(l)\big)^+\Big| > y\Big) \\
&\leq \mathbf{P}\Big(\sum_{j=1}^{m+k(n)}\Big|\big(\xi_j^{[n]}(l)\big)^+ - \big(\xi_j^{[n],m}(l)\big)^+\Big| > y\Big) \leq \mathbf{P}\Big(\sum_{j=1}^{m+k(n)}\Big|\underbrace{\xi_j^{[n]}(l) - \xi_j^{[n],m}(l)}_{\stackrel{\text{def}}{=} q_j}\Big| > y\Big).
\end{aligned} \tag{87}$$

Furthermore, we claim the existence of some constant $\tilde{C} \in (0,\infty)$ such that (for any $y,d>0$, $l\in[0.n]$, and any $n\geq 1$, $m\in\mathbb{N}$)

$$\mathbf{P}\Big(\sum_{j=1}^{m+k(n)}|q_j| > y\Big) \leq \tilde{C}\cdot\frac{n\bar{\sigma}^2(\kappa_{n,m})}{y^2}. \tag{88}$$

Then using Result 5, we yield

$$n\bar{\sigma}^2(\kappa_{n,m}) \leq n\cdot\kappa_{n,m}^{2-\beta_+}I_0^{\beta_+}(\nu) = \frac{\kappa^{m(2-\beta_+)}}{n^{r(2-\beta_+)-1}}\cdot I_0^{\beta_+}(\nu)$$

where $I_0^{\beta_+}(\nu) = \int_{(-1,1)} \int |x|^{\beta_+}\nu(dx)$. Setting $C = \tilde{C} I_0^{\beta_+}(\nu)$, we conclude the proof.

Now, it only remains to prove claim (88). Let $\chi = 2^{1/4}$. Note that $1 = (\chi - 1)\sum_{j\geq 1}\frac{1}{\chi^j} \geq (\chi-1)\Big(\frac{1}{\chi}+\frac{1}{\chi^2}+\cdots+\frac{1}{\chi^{m+k(n)}}\Big)$. As a result,

$$\mathbf{P}\Big(\sum_{j=1}^{k(n)+m}|q_j| > y\Big) \leq \mathbf{P}\Big(\sum_{j=1}^{k(n)+m}|q_j| > y(\chi-1)\sum_{j=1}^{k(n)+m}\frac{1}{\chi^j}\Big) \leq \sum_{j=1}^{k(n)+m}\mathbf{P}\Big(|q_j| > y\cdot\frac{\chi-1}{\chi^j}\Big) \tag{89}$$

Next, we bound each $\mathbf{P}(|q_j| > y\frac{\chi-1}{\chi^j})$. Conditioning on $l_j(l) = t$ (for any $t \in [0,l]$), we get

$$\begin{aligned}
\mathbf{P}\Big(|q_j| > y\frac{\chi-1}{\chi^j}\ \Big|\ l_j(l) = t\Big) &= \mathbf{P}\Big(\Big|\Xi_n(t) - \breve{\Xi}_n^m(t)\Big| > y\frac{\chi-1}{\chi^j}\Big) && \text{due to (75)}\\
&\leq \frac{\chi^{2j}}{y^2(\chi-1)^2}\mathbf{E}\Big|X^{(-\kappa_{n,m},\kappa_{n,m})}(t) - \bar{\sigma}(\kappa_{n,m})B(t)\Big|^2\\
&= \frac{\chi^{2j}}{y^2(\chi-1)^2}\cdot 2\bar{\sigma}^2(\kappa_{n,m})t\\
\Longrightarrow \mathbf{P}\Big(|q_j| > y\frac{\chi-1}{\chi^j}\Big) &\leq \frac{\chi^{2j}}{y^2(\chi-1)^2}\cdot 2\bar{\sigma}^2(\kappa_{n,m})\cdot\mathbf{E}[l_j(l)]\\
&= \frac{\sqrt{2^j}}{y^2(2^{1/4}-1)^2}\cdot 2\bar{\sigma}^2(\kappa_{m,n})\cdot\mathbf{E}[l_j(l)] && \text{due to } \chi = 2^{1/4}\\
&= \frac{\sqrt{2^j}}{y^2(2^{1/4}-1)^2}\cdot 2\bar{\sigma}^2(\kappa_{m,n})\cdot\frac{l}{2^j} && \text{by definition of } l_j(l) \text{ in (74)}\\
&\leq \frac{2}{(2^{1/4}-1)^2\sqrt{2^j}}\cdot\frac{n\bar{\sigma}^2(\kappa_{m,n})}{y^2} && \text{due to } l \leq n.
\end{aligned}$$

Therefore, in (89), we get

$$\mathbf{P}\Big(\sum_{j=1}^{k(n)+m}|q_j| > y\Big) \leq \frac{n\bar{\sigma}^2(\kappa_{m,n})}{y^2}\sum_{j\geq 1}\frac{2}{(2^{1/4}-1)^2\sqrt{2^j}} = \frac{n\bar{\sigma}^2(\kappa_{m,n})}{y^2}\cdot\underbrace{\frac{2\sqrt{2}}{(2^{1/4}-1)^2(\sqrt{2}-1)}}_{\stackrel{\text{def}}{=}\tilde{C}},$$

thus establishing claim (88). □

Proof of Lemma 6 For this proof, we adopt the notation $\breve{l}_k(l) \stackrel{\text{def}}{=} l - l_1(l) - l_2(l) - \ldots - l_k(l)$ for the remaining stick length after the first k sticks. Conditioning on $\breve{l}_{m+\lceil\log_2(n^d)\rceil}(l) = t$,

$$\begin{aligned}
\mathbf{P}\Big(\sum_{j>m+\lceil\log_2(n^d)\rceil}\big(\xi_j^{[n]}(l)\big)^+ > x\ \Big|\ \breve{l}_{m+\lceil\log_2(n^d)\rceil}(l) = t\Big) &= \mathbf{P}\Big(\sup_{s\in[0,t]}\Xi_n(s) > x\Big) && \text{by Result 3}\\
&\leq \frac{C_X}{x}(\sqrt{t}+t) && \text{using Lemma 3.}
\end{aligned}$$

Therefore, unconditionally,

$$\begin{aligned}\mathbf{P}\Big(\sum_{j>m+\lceil \log_2(n^d)\rceil}\big(\xi_j^{[n]}(l)\big)^+>x\Big) &\leq \frac{C_X}{x}\mathbf{E}\Big[\sqrt{\breve{l}_{m+\lceil \log_2(n^d)\rceil}(l)}+\mathbf{E}\breve{l}_{m+\lceil \log_2(n^d)\rceil}(l)\Big]\\ &\leq \frac{C_X}{x}\Big[\sqrt{\mathbf{E}\breve{l}_{m+\lceil \log_2(n^d)\rceil}(l)}+\mathbf{E}\breve{l}_{m+\lceil \log_2(n^d)\rceil}(l)\Big]\end{aligned}$$

The last line follows from Jensen's inequality. Lastly, by definition of $l_j(l)$'s in (74), we have

$$\mathbf{E}\breve{l}_{m+\lceil \log_2(n^d)\rceil}(l)=\frac{l}{2^{m+\lceil \log_2(n^d)\rceil}}\leq\frac{l}{2^m\cdot n^d}\leq\frac{n}{2^m\cdot n^d}=\frac{1}{n^{d-1}\cdot 2^m}\qquad\text{due to } l\in[0,n].$$

This concludes the proof. □

Proof of Lemma 7 To simplify notations, in this proof we set $k(n)=\lceil \log_2(n^d)\rceil$ and write $l_j=l_j(l)$ when there is no ambiguity. For the sequence of random variables $(l_1,\cdots,l_{m+k(n)})$, let $\tilde{l}_1\geq\tilde{l}_2\geq\cdots\geq\tilde{l}_{m+k(n)}$ be its order statistics. Given any ordered positive real sequence $t_1\geq t_2\geq\cdots\geq t_{m+k(n)}>0$, by conditioning on $\tilde{l}_j=t_j\ \forall j\in[m+k(n)]$, it follows from (75) that

$$\mathbf{P}\Big(\sum_{j=1}^{m+k(n)}\big(\xi_j^{[n]}(l)\big)^+\in[y,y+c]\ \Big|\ \tilde{l}_j=t_j\ \forall j\in[m+k(n)]\Big)=\mathbf{P}\Big(\sum_{j=1}^{m+k(n)}\big(\Xi_n^{(j)}(t_j)\big)^+\in[y,y+c]\Big)\tag{90}$$

where $\Xi_n^{(j)}$'s are iid copies of the Lévy processes $\Xi_n=X^{<n\gamma}$. Next, fix $\eta=\delta^{m\alpha_3}/n^{\alpha_4}$. Given the sequence of real numbers t_j's, we define $J\stackrel{\text{def}}{=}\#\{j\in[m+k(n)]:\ t_j>\eta\}$ as the number of elements in the sequence that are larger than η. In case that $t_1\leq\eta$, we set $J=0$. With J defined, we consider a decomposition of events in (90) based on the first $j\in[m+k(n)]$ such that $\Xi_n^{(j)}(t_j)>0$ (and hence $\big(\Xi_n^{(j)}(t_j)\big)^+>0$), especially if such t_j is larger than η or not. To be specific,

$$\begin{aligned}&\mathbf{P}\Big(\sum_{j=1}^{m+k(n)}\big(\Xi_n^{(j)}(t_j)\big)^+\in[y,y+c]\Big)\\ &=\sum_{j=1}^{J}\underbrace{\mathbf{P}\Big(\Xi_n^{(i)}(t_i)\leq 0\ \forall i\in[j-1];\ \Xi_n^{(j)}(t_j)>0;\ \sum_{i=j}^{m+k(n)}\big(\Xi_n^{(i)}(t_i)\big)^+\in[y,y+c]\Big)}_{\stackrel{\text{def}}{=}p_j}\\ &+\underbrace{\mathbf{P}\Big(\Xi_n^{(i)}(t_i)\leq 0\ \forall i\in[J];\ \sum_{j=J+1}^{m+k(n)}\big(\Xi_n^{(j)}(t_j)\big)^+\in[y,y+c]\Big)}_{\stackrel{\text{def}}{=}p_*}.\end{aligned}\tag{91}$$

We first bound terms p_j's. For any $j \in [J]$, observe that

$$\begin{aligned}
p_j &\leq \mathbf{P}\Big(\Xi_n^{(j)}(t_j) > 0;\ \sum_{i=j}^{m+k(n)} \big(\Xi_n^{(i)}(t_i)\big)^+ \in [y, y+c]\Big) \\
&= \int_{\mathbb{R}} \mathbf{P}\Big(\Xi_n^{(j)}(t_j) \in [y-x, y-x+c] \cap (0,\infty)\Big)\mathbf{P}\Big(\sum_{i=j+1}^{m+k(n)} \big(\Xi_n^{(i)}(t_i)\big)^+ \in dx\Big) \\
&\leq \frac{Cc}{t_j^{\lambda} \wedge 1} \qquad \text{by Assumption 2} \\
&\leq \frac{Cn^{\alpha_4\lambda}}{\delta^{m\alpha_3\lambda}} \cdot c \qquad \text{due to } j \leq J \text{, and hence } t_j > \eta = \delta^{m\alpha_3}/n^{\alpha_4}.
\end{aligned} \tag{92}$$

On the other hand,

$$\begin{aligned}
p_* &\leq \mathbf{P}\Big(\sum_{j=J+1}^{m+k(n)} \big(\Xi_n^{(j)}(t_j)\big)^+ \in [y, y+c]\Big) \leq \mathbf{P}\Big(\sum_{j=J+1}^{m+k(n)} \big(\Xi_n^{(j)}(t_j)\big)^+ \geq y_0\Big) \qquad \text{due to } y \geq y_0 > 0 \\
&\leq \sum_{j=J+1}^{m+k(n)} \mathbf{P}\Big(\Xi_n^{(j)}(t_j) \geq y_0/N\Big) \qquad \text{with } N \stackrel{\text{def}}{=} m + k(n) - J \\
&\leq \sum_{j=J+1}^{m+k(n)} \frac{C_X(\sqrt{t_j} + t_j)\cdot N}{y_0} \qquad \text{by Lemma 3} \\
&\leq \sum_{j=J+1}^{m+k(n)} \frac{C_X(\sqrt{\eta} + \eta)\cdot N}{y_0} \qquad \text{due to } j > J \text{, and hence } t_j \leq \eta = \delta^{m\alpha_3}/n^{\alpha_4} \\
&= N^2 \cdot \frac{C_X(\sqrt{\eta}+\eta)}{y_0} \leq \big(m + k(n)\big)^2 \cdot \frac{C_X(\sqrt{\eta}+\eta)}{y_0} \qquad \text{due to } N \leq m + k(n) \\
&\leq 2C_X\big(m^2 + (\lceil \log_2(n^d) \rceil)^2\big)\frac{\sqrt{\eta}+\eta}{y_0} \qquad \text{using } (u+v)^2 \leq 2(u^2+v^2) \\
&\leq 4C_X\big(m^2 + (\lceil \log_2(n^d) \rceil)^2\big)\frac{\sqrt{\eta}}{y_0} = 4C_X\big(m^2 + (\lceil \log_2(n^d) \rceil)^2\big)\frac{\delta^{m\alpha_3/2}}{y_0 \cdot n^{\alpha_4/2}}.
\end{aligned} \tag{93}$$

Plugging (92) and (93) into (91), we yield

$$\begin{aligned}
&\mathbf{P}\Big(\sum_{j=1}^{m+k(n)} \big(\xi_j^{[n]}(l)\big)^+ \in [y, y+c] \ \Big|\ \tilde{l}_j = t_j\ \forall j \in [m+k(n)]\Big) \\
&\leq J \cdot \frac{Cn^{\alpha_4\lambda}}{\delta^{m\alpha_3\lambda}} c + 4C_X\big(m^2 + (\lceil \log_2(n^d) \rceil)^2\big)\frac{\delta^{m\alpha_3/2}}{y_0 \cdot n^{\alpha_4/2}} \\
&\leq C\frac{(m + (\lceil \log_2(n^d) \rceil) n^{\alpha_4\lambda}}{\delta^{m\alpha_3\lambda}} c + 4C_X\big(m^2 + (\lceil \log_2(n^d) \rceil)^2\big)\frac{m\delta^{\alpha_3/2}}{y_0 \cdot n^{\alpha_4/2}} \qquad \text{due to } J \leq m + \lceil \log_2(n^d) \rceil.
\end{aligned}$$

To conclude the proof, just note that the inequality above holds when conditioning on any sequence of $t_1 \geq t_2 \geq \cdots \geq t_{m+k(n)} > 0$, so it would also hold unconditionally. $\square$