Rare event simulation for processes generated via stochastic fixed point equations

Collamore, Jeffrey F.; Diao, Guoqing; Vidyashankar, Anand N.

Published in:
Annals of Applied Probability

DOI:
10.1214/13-AAP974

Publication date:
2014

Document version
Publisher's PDF, also known as Version of record

The Annals of Applied Probability
2014, Vol. 24, No. 5, 2143–2175
DOI: 10.1214/13-AAP974

RARE EVENT SIMULATION FOR PROCESSES GENERATED VIA STOCHASTIC FIXED POINT EQUATIONS

BY JEFFREY F. COLLAMORE[1], GUOQING DIAO AND ANAND N. VIDYASHANKAR[2]

University of Copenhagen, George Mason University and George Mason University

In a number of applications, particularly in financial and actuarial mathematics, it is of interest to characterize the tail distribution of a random variable V satisfying the distributional equation $V \stackrel{\mathcal{D}}{=} f(V)$, where $f(v) = A\max\{v, D\} + B$ for $(A, B, D) \in (0, \infty) \times \mathbb{R}^2$. This paper is concerned with computational methods for evaluating these tail probabilities. We introduce a novel importance sampling algorithm, involving an exponential shift over a random time interval, for estimating these rare event probabilities. We prove that the proposed estimator is: (i) consistent, (ii) strongly efficient and (iii) optimal within a wide class of dynamic importance sampling estimators. Moreover, using extensions of ideas from nonlinear renewal theory, we provide a precise description of the running time of the algorithm. To establish these results, we develop new techniques concerning the convergence of moments of stopped perpetuity sequences, and the first entrance and last exit times of associated Markov chains on $\mathbb{R}$. We illustrate our methods with a variety of numerical examples which demonstrate the ease and scope of the implementation.

1. Introduction. This paper introduces a rare event simulation algorithm for estimating the tail probabilities of the *stochastic fixed point equation* (SFPE)

$$V \stackrel{\mathcal{D}}{=} f(V) \qquad \text{where } f(v) \equiv A\max\{v, D\} + B \tag{1.1}$$

for $(A, B, D) \in (0, \infty) \times \mathbb{R}^2$. SFPEs of this general form arise in a wide variety of applications, such as extremal estimates for financial time series models and ruin estimates in actuarial mathematics. Other related applications arise in branching

Received July 2011; revised September 2013.

[1]Supported in part by Danish Research Council (SNF) Grant "Point Process Modelling and Statistical Inference," No. 09-092331.

[2]Supported by Grant NSF DMS 000-03-07057.

MSC2010 subject classifications. Primary 65C05, 91G60, 68W40, 60H25; secondary 60F10, 60G40, 60J05, 60J10, 60J22, 60K15, 60K20, 60G70, 68U20, 91B30, 91B70, 91G70.

Key words and phrases. Monte Carlo methods, importance sampling, perpetuities, large deviations, nonlinear renewal theory, Harris recurrent Markov chains, first entrance times, last exit times, regeneration times, financial time series, GARCH processes, ARCH processes, risk theory, ruin theory with stochastic investments.

processes in random environments and the study of algorithms in computer science. See Collamore (2009), Collamore and Vidyashankar (2013b), or Section 4 below for a more detailed description of some of these applications.

In a series of papers [e.g., Kesten (1973), Vervaat (1979), Goldie (1991)], the tail probabilities for the SFPE (1.1) have been asymptotically characterized. Under appropriate moment and regularity conditions, it is known that

$$\lim_{u\to\infty} u^{\xi}\mathbf{P}\{V > u\} = C \tag{1.2}$$

for finite positive constants C and ξ, where ξ is identified as the nonzero solution to the equation $\mathbf{E}[A^{\alpha}] = 1$. Recently, in Collamore and Vidyashankar (2013b), the constant C has been identified as the ξth moment of the difference of a perpetuity sequence and a conjugate sequence.

The purpose of this article is to introduce a rigorous computational approach, based on importance sampling, for Monte Carlo estimation of the rare event probability $\mathbf{P}\{V > u\}$. While importance sampling methods have been developed for numerous large deviation problems involving i.i.d. and Markov-dependent random walks [cf. Asmussen and Glynn (2007)], the adaptation of these methods to (1.1) is distinct and requires new techniques. In this paper, we propose a nonstandard approach involving a *dual* change of measure of a process $\{V_n\}$ performed over two random time intervals: namely, the excursion of $\{V_n\}$ to (u,∞) followed by the return of this process to a given set $\mathcal{C} \subset \mathbb{R}$.

The motivation for our algorithm stems from the observation that the SFPE (1.1) induces a forward recursive sequence, namely,

$$V_n = A_n \max\{D_n, V_{n-1}\} + B_n, \qquad n = 1, 2, \ldots, V_0 = v, \tag{1.3}$$

where $\{(A_n, B_n, D_n) : n \in \mathbb{Z}_+\}$ is an i.i.d. sequence with the same law as (A, B, D). It is important to observe that in many applications, the mathematical process under study is obtained through the *backward* iterates of the given SFPE [as described by Letac (1986) or Collamore and Vidyashankar (2013b), Section 2.1]. For example, the linear recursion $f(v) = Av + B$ induces the backward recursive sequence or *perpetuity sequence*

$$Z_n := V_0 + \frac{B_1}{A_1} + \frac{B_2}{A_1A_2} + \cdots + \frac{B_n}{A_1\cdots A_n}, \qquad n = 1, 2, \ldots. \tag{1.4}$$

However, since $\{Z_n\}$ is not Markovian, it is less natural to simulate $\{Z_n\}$ than the corresponding forward sequence $\{V_n\}$. Thus, a central aspect of our approach is the *conversion* of the given perpetuity sequence, via its SFPE, into a forward recursive sequence which we then simulate. Because $\{V_n\}$ is Markovian, we can then study this process over excursions emanating from, and then returning to, a given set $\mathcal{C} \subset \mathbb{R}$.

In the special case of the perpetuity sequence in (1.4), simulation methods for estimating $\mathbf{P}\{\lim_{n\to\infty} Z_n > u\}$ have recently been studied in Blanchet, Lam and Zwart (2012) under the strong assumption that $\{B_n\}$ is nonnegative. Their method is very different from ours, involving the simulation of $\{Z_n\}$ *directly* until the first passage time to a level cu, where $c \in (0, 1)$, and a rough analytical approximation to relate this probability to the first passage probability at level u. Their methods do not generalize to the other processes studied in this paper, such as the ruin problem with investments or related extensions. In contrast, our goal here is to develop a *general* algorithm which is flexible and can be applied to the wider class of processes governed by (1.1) and some of its extensions. While we focus on (1.1), it is worthwhile to mention here that our algorithm provides an important ingredient for addressing a larger class of problems, including nonhomogeneous recursions on trees, which are analyzed in Collamore, Vidyashankar and Xu (2013). Also, it seems plausible that the method should extend to the class of random maps which can be approximated by (1.1) in the sense of Collamore and Vidyashankar (2013b), Section 2.4. This extension would encompass several other problems of applied interest, such as the AR(1) process with ARCH(1) errors. Yet another feasible generalization is to Markov-dependent recursions under Harris recurrence, utilizing the reduction to i.i.d. recursions described in Collamore (2009) and Collamore and Vidyashankar (2013a), Section 3.

In this paper, we present an algorithm and establish that it is consistent and efficient; that is, it displays the bounded relative error property. It is interesting to note that in the proof of efficiency, certain new issues arise concerning the convergence of the perpetuity sequence (1.4). Specifically, while it is known that (1.4) converges to a finite limit under minimal conditions, the necessary and sufficient condition for the L_β convergence of $\{Z_n\}$ in (1.4) is that $\mathbf{E}[A^{-\beta}] < 1$; cf. Alsmeyer, Iksanov and Rösler (2009). However, our analysis will involve moments of quantities similar to $\{Z_n\}$, but where $\mathbf{E}[A^{-\beta}]$ is *greater* than one, and hence our perpetuity sequences will necessarily be divergent in L_β. To circumvent this difficulty, we study these perpetuity sequences over randomly stopped intervals, namely, over cycles emanating from, and returning to, a given subset $\mathcal{C}$ of $\mathbb{R}$. As a technical point, it is worth noting that if the return time, K, were replaced by the more commonly studied regeneration time τ of the chain $\{V_n\}$, then the existing literature on Markov chain theory would still not shed much light on the tails of τ and hence the convergence of V_τ. Thus, the fact that K has sufficient exponential tails for the convergence of V_K is due to the recursive structure of the particular class of Markov chains we consider and seems to be a general property for this class of Markov chains. These results concerning the moments of L_β-divergent perpetuity sequences complement the known literature on perpetuities and appear to be of some independent interest.

Next, we go beyond the current literature by establishing a *sharp* asymptotic estimate for the running time of the algorithm, thereby showing that our algorithm is, in fact, strongly efficient; cf. Remark 2.2 below. To this end, we introduce methods from nonlinear renewal theory, as well as methods from Markov chain theory involving the first entrance and last exit times of the process $\{V_n\}$. Finally, motivated by the Wentzell–Freidlin theory of large deviations, we provide an optimality result; specifically, we consider other possible level-dependent changes of measure for the process $\{V_n\}$ selected from a wide class of dynamic importance sampling algorithms [in the sense of Dupuis and Wang (2005)]. We show that our algorithm is the *unique* choice which attains bounded relative error, thus establishing the validity of our method amongst a natural class of possible algorithms.

2. The algorithm and a statement of the main results.

2.1. *Background*: *The forward and backward recursive sequences*. We start with a general SFPE of the form

$$V \stackrel{\mathcal{D}}{=} f(V) \equiv F_Y(V), \tag{2.1}$$

where $F_Y : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ is deterministic, measurable and continuous in its first component. Let v be an element of the range of F_Y, and let $\{Y_n\}$ be an i.i.d. sequence of r.v.'s such that $Y_n \stackrel{\mathcal{D}}{=} Y$ for all n. Then the *forward sequence* generated by the SFPE (2.1) is defined by

$$V_n(v) = F_{Y_n} \circ F_{Y_{n-1}} \circ \cdots \circ F_{Y_1}(v), \qquad n = 1, 2, \ldots, \; V_0 = v, \tag{2.2}$$

whereas the *backward sequence* generated by this SFPE is defined by

$$Z_n(v) = F_{Y_1} \circ F_{Y_2} \circ \cdots \circ F_{Y_n}(v), \qquad n = 1, 2, \ldots, \; Z_0 = v. \tag{2.3}$$

While the forward sequence is always Markovian, the backward equation need not be Markovian; however, for every v and n, $V_n(v)$ and $Z_n(v)$ are identically distributed. This observation is critical since it suggests that—regardless of whether the SFPE was originally obtained via forward or backward iteration—a natural approach to analyzing the process is through its forward iterates.

2.2. *Background*: *Asymptotic estimates*. We now specialize to the recursion (1.1). This recursion is often referred to as "Letac's model E."

Let $\mathfrak{F}_n$ denote the σ-field generated by $\{(A_i, B_i, D_i) : 1 \le i \le n\}$, and let

$$\lambda(\alpha) = \mathbf{E}\big[A^\alpha\big] \quad \text{and} \quad \Lambda(\alpha) = \log \lambda(\alpha), \qquad \alpha \in \mathbb{R}.$$

Let μ denote the distribution of $Y = (\log A, B, D)$ and μ_α denote the α-shifted distribution with respect to the first variable; that is,

$$\mu_\alpha(E) := \frac{1}{\lambda(\alpha)} \int_E e^{\alpha x}\, d\mu(x, y, z), \qquad E \in \mathcal{B}\big(\mathbb{R}^3\big), \alpha \in \mathbb{R}, \tag{2.4}$$

where, here and in the following, $\mathcal{B}(E)$ denotes the Borel sets of E. Let $\mathbf{E}_\alpha[\cdot]$ denote expectation with respect to this α-shifted measure.

For any r.v. X, let $\mathfrak{L}(X)$ denote the probability law of X, and let $\operatorname{supp}(X)$ denote the support of X. Also, write $X \sim \mathfrak{L}(X)$ to denote that X has this probability law. Given an i.i.d. sequence $\{X_n\}$, we will often write X for a "generic" element of this sequence. Finally, for any function f, let $\operatorname{dom}(f)$ denote the domain of f, and let f', f'', etc. denote the successive derivatives of f.

We now state the main hypotheses needed to establish the asymptotic decay of $\mathbf{P}\{V > u\}$ in (1.2); note that (H_0) is only needed to obtain the explicit representation of C, as given in Collamore and Vidyashankar (2013b). These conditions will form the starting point of our study.

HYPOTHESES.

(H_0) The r.v. A has an absolutely continuous component with respect to Lebesgue measure with a nontrivial continuous density in a neighborhood of $\mathbb{R}$.

(H_1) $\Lambda(\xi) = 0$ for some $\xi \in (0, \infty) \cap \mathrm{dom}(\Lambda')$.

(H_2) $\mathbf{E}[|B|^{\xi}] < \infty$ and $\mathbf{E}[(A|D|)^{\xi}] < \infty$.

(H_3) $\mathbf{P}\{A > 1, B > 0\} > 0$ or $\mathbf{P}\{A > 1, B \geq 0, D > 0\} > 0$.

Note that (H_3) implies that the process $\{V_n\}$ is nondegenerate (i.e., it is not concentrated at a single point).

Under these hypotheses, it can be shown that the forward sequence $\{V_n\}$ generated by the SFPE (1.1) is a Markov chain which is φ-irreducible and geometrically ergodic [Collamore and Vidyashankar (2013b), Lemma 5.1]. Thus $\{V_n\}$ converges to a r.v. V *which itself satisfies the SFPE* (1.1). Moreover, with respect to its α-shifted measure, the process $\{V_n\}$ is *transient* [Collamore and Vidyashankar (2013b), Lemma 5.2].

Our present goal is to develop an efficient Monte Carlo algorithm for evaluating $\mathbf{P}\{V > u\}$, for *fixed* u, which remains efficient in the asymptotic limit as $u \to \infty$.

2.3. *The algorithm.* Since the forward process $V_n = A_n \max\{D_n, V_{n-1}\} + B_n$ satisfies $V_n \approx A_n V_{n-1}$ for large V_{n-1}, and since $\{V_n\}$ is transient in its ξ-shifted measure, large deviation theory suggests that we consider shifted distributions and, in particular, the shifted measure μ_ξ, where ξ is given as in (H_1). To relate $\mathbf{P}\{V > u\}$ under its original measure to the paths of $\{V_n\}$ under μ_ξ-measure, let $\mathcal{C} := [-M, M]$ for some $M \geq 0$, and let π denote the stationary distribution of $\{V_n\}$. Now define a probability measure γ on $\mathcal{C}$ by setting

$$\gamma(E) = \frac{\pi(E)}{\pi(\mathcal{C})}, \qquad E \in \mathcal{B}(\mathcal{C}). \tag{2.5}$$

Let $K := \inf\{n \in \mathbb{Z}_+ : V_n \in \mathcal{C}\}$. Then in Section 3, we will establish the following representation formula:

$$\mathbf{P}\{V > u\} = \pi(\mathcal{C})\mathbf{E}_\gamma[N_u], \qquad N_u := \sum_{n=0}^{K-1} \mathbf{1}_{\{V_n > u\}}, \tag{2.6}$$

where $\mathbf{E}_\gamma[\cdot]$ denotes the expectation when the initial state $V_0 \sim \gamma$. Thus motivated by large deviation theory and the previous formula, we simulate $\{V_n\}$ *over a cycle* emanating from the set $\mathcal{C}$ (with initial state $V_0 \sim \gamma$), and then returning to $\mathcal{C}$, where simulation is performed in the dual measure, which we now describe.

Set $T_u = \inf\{n : V_n > u\}$, and let

$$\mathfrak{L}(\log A_n, B_n, D_n) = \begin{cases} \mu_\xi, & \text{for } n = 1, \dots, T_u, \\ \mu, & \text{for } n > T_u, \end{cases} \tag{$\mathfrak{D}$}$$

where μ_ξ is defined as in (2.4) and ξ is given as in (H_1). Let $\{V_n\}$ be generated by the forward recursion (1.3), but with a *driving sequence* $\{\mathcal{V}_n\} \equiv \{(\log A_n, B_n, D_n)\}$ which is governed by ($\mathfrak{D}$) rather than by the fixed measure μ. Roughly speaking, the "dual measure" ($\mathfrak{D}$) shifts the distribution of $\log A_n$ on a path of $\{V_n\}$ until this process exceeds the level u, and reverts to the original measure thereafter. Let $\mathbf{E}_{\mathfrak{D}}[\cdot]$ denote expectation with respect to ($\mathfrak{D}$).

To relate the simulated sequence in the dual measure to the required probability in the original measure, we introduce a weighting factor. Specifically, in the proof of Theorem 2.2 below, we will show

$$\mathbf{E}_{\mathfrak{D}}[\mathcal{E}_u] = \pi(\mathcal{C})\mathbf{E}_{\mathfrak{D}}\big[N_u e^{-\xi S_{T_u}} \mathbf{1}_{\{T_u < K\}} | V_0 \sim \gamma\big],$$

where $S_n := \sum_{i=1}^n \log A_i$ and γ is given as in (2.5). Using this identity, it is natural to introduce the *importance sampling estimator*

$$\mathcal{E}_u = N_u e^{-\xi S_{T_u}} \mathbf{1}_{\{T_u < K\}}. \tag{2.7}$$

Then $\pi(\mathcal{C})\mathcal{E}_u$ is an unbiased estimator for $\mathbf{P}\{V > u\}$. However, since the stationary distribution π and hence the distribution γ is seldom known—even if the underlying distribution of $(\log A, B, D)$ is known—we first run multiple realizations of $\{V_n\}$ according to the known measure μ and thereby estimate $\pi(\mathcal{C})$ and γ. Let $\hat{\pi}_k(\mathcal{C})$, $\hat{\gamma}_k$ denote the estimates obtained for $\pi(\mathcal{C})$, γ, respectively, and let $\widehat{\mathcal{E}}_{u,n}$ denote the estimate obtained upon averaging the realizations of $\mathcal{E}_u$. This yields the estimator $\hat{\pi}_k(\mathcal{C})\widehat{\mathcal{E}}_{u,n}$.

This discussion can be formalized as follows:

Rare event simulation algorithm using forward iterations of the SFPE

$V_0 \sim \hat{\gamma}_k$, $m = 0$
repeat
  $m \leftarrow m + 1$
  $V_m = A_m \max\{D_m, V_{m-1}\} + B_m$, $(\log A_m, B_m, D_m) \sim \mu_\xi$
until $V_m > u$ or $V_m \in \mathcal{C}$
if $V_m > u$ **then**
  repeat
    $m \leftarrow m + 1$
    $V_m = A_m \max\{D_m, V_{m-1}\} + B_m$, $(\log A_m, B_m, D_m) \sim \mu$
  until $V_m \in \mathcal{C}$
  $\mathcal{E}_u = N_u e^{-\xi S_{T_u}} \mathbf{1}_{\{T_u < K\}}$
else
  $\mathcal{E}_u = 0$
end if

The actual estimate is then obtained by letting $\mathcal{E}_{u,j}$ $(j = 1, \dots, n)$ denote the realizations of $\mathcal{E}_u$ produced by the algorithm and setting $\mathbf{P}\{V > u\} = \hat{\pi}_k(\mathcal{C})\widehat{\mathcal{E}}_{u,n}$, where

$$\hat{\pi}_k(\mathcal{C}) = \frac{1}{k}\sum_{j=1}^{k} \mathbf{1}_{\{V^{(j)} \in \mathcal{C}\}} \quad \text{and} \quad \widehat{\mathcal{E}}_{u,n} = \frac{1}{n}\sum_{j=1}^{n} \mathcal{E}_{u,j},$$

where $V^{(1)}, V^{(2)}, \dots, V^{(k)}$ is a sample from the distribution of V (which, we emphasize, is sampled from the *center* of the distribution). In Section 4, we describe how to obtain samples from V from a practical perspective. Finally, note that $\widehat{\mathcal{E}}_{u,n}$ also depends on k.

It is worth observing that in the special case $D = 1$ and $B = 0$, Letac's model E reduces to a multiplicative random walk. Moreover, in that case, one can always take γ to be a point mass at $\{1\}$, at which point the process regenerates. In this much-simplified setting, our algorithm reduces to a standard regenerative importance sampling algorithm, as may be used to evaluate the stationary exceedance probabilities in a GI/G/1 queue.

2.4. *Consistency and efficiency of the algorithm.* We begin by stating our results on consistency and efficiency.

THEOREM 2.1. *Assume Letac's model E, and suppose that* (H_1), (H_2) *and* (H_3) *are satisfied. Then for any* $\mathcal{C}$ *such that* $\mathcal{C} \cap \operatorname{supp}(\pi) \neq \varnothing$ *and any* u *such that* $u \notin \mathcal{C}$, *the algorithm is strongly consistent*; *that is*,

$$\lim_{k\to\infty}\lim_{n\to\infty} \hat{\pi}_k(\mathcal{C})\widehat{\mathcal{E}}_{u,n} = \mathbf{P}\{V > u\} \qquad a.s. \tag{2.8}$$

REMARK 2.1. If the stationary distribution π of $\{V_n\}$ is known on $\mathcal{C}$ (e.g., $\mathcal{C} = \{v\}$ for $v \in \mathbb{R}$), then it will follow from the proof of the theorem that $\pi(\mathcal{C})\widehat{\mathcal{E}}_{u,n}$ is an unbiased estimator for $\mathbf{P}\{V > u\}$.

THEOREM 2.2. *Assume Letac's model E, and suppose that* (H_1) *and* (H_3) *are satisfied. Also, in place of* (H_2), *assume that for some* $\alpha > \xi$,

$$\mathbf{E}\big[\big(A^{-1}|B|^2\big)^{\alpha}\big] < \infty \quad \text{and} \quad \mathbf{E}\big[\big(A|D|^2\big)^{\alpha}\big] < \infty. \tag{2.9}$$

Moreover, assume that one of the following two conditions holds: $\lambda(\alpha) < \infty$ *for some* $\alpha < -\xi$; *or* $\mathbf{E}[(|D| + (A^{-1}|B|))^{\alpha}] < \infty$ *for all* $\alpha > 0$. *Then, there exists an* $M > 0$ *such that*

$$\sup_{u\ge 0}\sup_{k\in\mathbb{Z}_+} u^{2\xi}\mathbf{E}_{\mathfrak{D}}\big[\mathcal{E}_u^2 | V_0 \sim \hat{\gamma}_k\big] < \infty. \tag{2.10}$$

Equation (2.10) implies that our estimator exhibits bounded relative error. However, a good choice of M is critical for the practical usefulness of the algorithm. A canonical method for choosing M can be based on the drift condition satisfied by $\{V_n\}$ (as given in Lemma 3.1 below), but in practice, a proper choice of M is problem-dependent and only obtained numerically based on the methods we introduce below in Section 4.

2.5. *Running time of the algorithm.* Next we provide precise asymptotics for the running time of the algorithm. In the following theorem, recall that K denotes the first return time to $\mathcal{C}$ (corresponding to the termination of the algorithm), whereas T_u denotes the first passage time to (u, ∞).

THEOREM 2.3. *Assume Letac's model E, and suppose that hypotheses* (H_0)–(H_3) *hold,* Λ''' *is finite on* $\{0, \xi\}$ *and for some* $\varepsilon > 0$,

$$\mathbf{P}_\xi\{V_1 \le 1 | V_0 = v\} = o(v^{-\varepsilon}) \qquad \text{as } v \to \infty. \tag{2.11}$$

Then

$$\mathbf{E}_{\mathfrak{D}}[K \mathbf{1}_{\{K<\infty\}}] < \infty; \tag{2.12}$$

$$\lim_{u\to\infty} \mathbf{E}_{\mathfrak{D}}\left[\frac{T_u}{\log u}\middle| T_u < K\right] = \frac{1}{\Lambda'(\xi)}; \tag{2.13}$$

$$\lim_{u\to\infty} \mathbf{E}_{\mathfrak{D}}\left[\frac{K - T_u}{\log u}\middle| T_u < K\right] = \frac{1}{|\Lambda'(0)|}. \tag{2.14}$$

REMARK 2.2. The ultimate objective of the algorithm is to minimize the simulation cost, that is, the total number of Monte Carlo simulations needed to attain a given accuracy. This grows according to

$$\operatorname{Var}(\mathcal{E}_u)\bigl\{c_1 \mathbf{E}_{\mathfrak{D}}[K | T_u < K] + c_2 \mathbf{E}_{\mathfrak{D}}[K \mathbf{1}_{\{T_u \ge K\}}]\bigr\} \qquad \text{as } u \to \infty \tag{2.15}$$

for appropriate constants c_1 and c_2; cf. Siegmund (1976). However, as a consequence of Theorem 2.4, we have that under the dual measure $(\mathfrak{D})$,

$$\mathbf{E}_{\mathfrak{D}}[K | T_u < K] \sim \Theta \log u \qquad \text{as } u \to \infty$$

for some positive constant Θ, while the last term in (2.15) converges to a finite constant. Thus, by combining Theorems 2.3 and 2.4, we conclude that our algorithm is indeed strongly efficient.

2.6. *Optimality of the algorithm.* We conclude with a comparison of our algorithm to other algorithms obtained through forward iterations involving alternative measure transformations. A natural alternative would be to simulate with some measure μ_α until the time $T_u = \inf\{n : V_n > u\}$ and revert to some other measure μ_β thereafter. More generally, we may consider simulating from a general class of distributions with some form of state dependence, as we now describe.

Let $\nu(\cdot; w, q)$ denote a probability measure on $\mathcal{B}(\mathbb{R}^3)$ indexed by two parameters, $w \in [0,1]$ and $q \in \{0,1\}$, where (w,q) denotes a realization of (W_n', Q_n) for

$$W_n' := \frac{\log V_{n-1}}{\log u} \quad \text{and} \quad Q_n := \mathbf{1}_{\{T_u < n\}}.$$

Set $W_n = W_n' \mathbf{1}_{\{W_n' \in [0,1]\}} + (W_n' \wedge 1)\mathbf{1}_{\{W_n' > 1\}}$. Note that (W_n, Q_n) is $\mathfrak{F}_{n-1}$ measurable. Let $\nu_n(\cdot) = \nu(\cdot; W_n, Q_n)$ be a random measure derived from the measure ν. Observe that, conditioned on $\mathfrak{F}_{n-1}$, ν_n is a probability measure. Now, we assume that the family of random measures $\{\nu_n(\cdot)\} \equiv \{\nu(\cdot; W_n, Q_n)\}$ satisfy the following regularity condition:

Condition (C_0): $\mu \ll \nu$ for each pair $(w,q) \in [0,1] \times \{0,1\}$, and

$$\mathbf{E}_{\mathfrak{D}}\left[\log\left(\frac{d\mu}{d\nu}(Y_n; W_n, Q_n)\right)\middle| W_n = w, Q_n = q\right]$$

is piecewise continuous as a function of w.

Let $\mathfrak{M}$ denote the class of measures $\{\nu_n\}$ where ν satisfies (C_0). Thus, we consider a class of distributions where we shift *all three* members of the driving sequence $Y_n = (\log A_n, B_n, D_n)$ in some way, allowing dependence on the history of the process through the parameters (w,q).

Now suppose that simulation is performed using a modification of our main algorithm, where $Y_n \sim \nu_n$ for some collection $\boldsymbol{\nu} := \{\nu_1, \nu_2, \ldots\} \in \mathfrak{M}$. Let $\mathcal{E}_u^{(\boldsymbol{\nu})}$ denote the corresponding importance sampling estimator. Let $\hat{\pi}_k$ denote an empirical estimate for π, as described in the discussion of our main algorithm, and let $\mathcal{E}_{u,1}^{(\boldsymbol{\nu})}, \ldots, \mathcal{E}_{u,n}^{(\boldsymbol{\nu})}$ denote simulated estimates for $\mathcal{E}_u^{(\boldsymbol{\nu})}$ obtained by repeating this algorithm, but with $\{\nu_n\}$ in place of the dual measure ($\mathfrak{D}$). Then it is easy to see, using the arguments of Theorem 2.2, that

$$\lim_{k\to\infty} \lim_{n\to\infty} \hat{\pi}_k(\mathcal{C}) \widehat{\mathcal{E}}_{u,n}^{(\boldsymbol{\nu})} = \mathbf{P}\{V > u\}, \tag{2.16}$$

where $\widehat{\mathcal{E}}_{u,n}^{(\boldsymbol{\nu})}$ denotes the average of n simulated samples of $\mathcal{E}_u^{(\boldsymbol{\nu})}$ (and depends on k); cf. (2.8). It remains to compare the variance of these estimators, which is the subject of the next theorem.

THEOREM 2.4. *Assume that the conditions of Theorems* 2.2 *and* 2.3 *hold. Let ν be a probability measure on $\mathcal{B}(\mathbb{R}^3)$ indexed by parameters $w \in [0,1]$ and $q \in \{0,1\}$, and assume that $\boldsymbol{\nu} \in \mathfrak{M}$. Then for any initial state $v \in \mathcal{C}$,*

$$\liminf_{u\to\infty} \frac{1}{\log u} \log\left(u^{2\xi} \mathbf{E}_\nu\left[\left(\mathcal{E}_u^{(\boldsymbol{\nu})}\right)^2 \middle| V_0 = v\right]\right) \ge 0. \tag{2.17}$$

Moreover, equality holds in (2.17) *if and only if* $\nu(\cdot; w, 0) = \mu_\xi$ *and* $\nu(\cdot; w, 1) = \mu$ *for all* $w \in [0, 1]$. *Thus, the dual measure in* ($\mathfrak{D}$) *is the unique optimal simulation strategy within the class* $\mathfrak{M}$.

3. Proofs of consistency and efficiency.

We start with consistency.

PROOF OF THEOREM 2.1. Let $K_0 := 0$, $K_n := \inf\{i > K_{n-1} : V_i \in \mathcal{C}\}$, $n \in \mathbb{Z}_+$, denote the successive return times of $\{V_n\}$ to $\mathcal{C}$. Set

$$X_n = V_{K_n}, \qquad n = 0, 1, \ldots.$$

Then we claim that the stationary distribution of $\{X_n\}$ is given by $\gamma(E) = \pi(E)/\pi(\mathcal{C})$, where π is the stationary distribution of $\{V_n\}$.

Notice that $\{X_n\}$ is φ-irreducible and geometrically ergodic [cf. Collamore and Vidyashankar (2013b), Lemma 5.1]. Now set $\mathfrak{N}_n := \sum_{i=1}^{n} \mathbf{1}_{\{V_i \in \mathcal{C}\}}$. Then by the law of large numbers for Markov chains,

$$\pi(E) = \lim_{n\to\infty} \frac{\mathfrak{N}_n}{n} \left(\frac{1}{\mathfrak{N}_n} \sum_{i=1}^{\mathfrak{N}_n} \mathbf{1}_{\{X_i \in E\}} \right) = \pi(\mathcal{C})\gamma(E) \qquad \text{a.s., } E \in \mathcal{B}(\mathcal{C}). \tag{3.1}$$

Hence $\gamma(E) = \pi(E)/\pi(\mathcal{C})$.

Next, we assert that $\mathbf{P}\{V > u\} = \pi(\mathcal{C})\mathbf{E}_\gamma[N_u]$. To establish this equality, again apply the law of large numbers for Markov chains to obtain that

$$\begin{aligned} \mathbf{P}\{V > u\} &:= \pi\big((u, \infty)\big) \\ &= \lim_{n\to\infty} \frac{1}{n} \left\{ \sum_{i=0}^{K_{\mathfrak{N}_n}-1} \mathbf{1}_{\{V_i > u\}} + \sum_{i=K_{\mathfrak{N}_n}}^{n} \mathbf{1}_{\{V_i > u\}} \right\} \qquad \text{a.s.} \end{aligned} \tag{3.2}$$

By the Markov renewal theorem [Iscoe, Ney and Nummelin (1985), Lemma 6.2], we claim that the last term on the right-hand side (RHS) of this equation converges to zero a.s. To see this, let $I(n)$ denote the last regeneration time occurring in the interval $[0, n]$, let $J(n)$ denote the first regeneration time occurring after time n, let τ denote a typical regeneration time. Then by Lemma 6.2 of Iscoe, Ney and Nummelin (1985) and the geometric ergodicity of $\{V_n\}$,

$$\lim_{n\to\infty} \mathbf{E}\big[e^{\varepsilon(J(n)-I(n))}\big] = \frac{1}{\mathbf{E}[\tau]} \mathbf{E}\big[\tau e^{\varepsilon\tau}\big] < \infty, \qquad \text{some } \varepsilon > 0. \tag{3.3}$$

Now by Nummelin's split-chain construction [Nummelin (1984), Section 4.4] and by the definition of $K_{\mathfrak{N}_n}$, $I(n) \le K_{\mathfrak{N}_n} \le n \le J(n) - 1$. Hence by a Borel–Cantelli argument,

$$\frac{1}{n} \sum_{i=K_{\mathfrak{N}_n}}^{n} \mathbf{1}_{\{V_i > u\}} \to 0 \qquad \text{a.s. as } n \to \infty. \tag{3.4}$$

Next consider the first term on the RHS of (3.2). Assume V_0 has distribution γ. For any $n \in \mathbb{Z}_+$, set $N_{u,n} = \sum_{i=K_{n-1}}^{K_n-1} \mathbf{1}_{\{V_i>u\}}$ (namely, the number of exceedances above level u which occur over the successive cycles starting from $\mathcal{C}$). Let $S_n^N = N_{u,1} + \cdots + N_{u,n}$, $n \in \mathbb{Z}_+$. It can be seen that $\{(X_n, N_{u,n})\}$ is a positive Harris chain and, hence, by another application of the law of large numbers for Markov chains,

$$\mathbf{E}_\gamma[N_u] = \lim_{n\to\infty} \frac{S_n^N}{n} := \lim_{n\to\infty} \frac{1}{n} \sum_{n=0}^{K_n-1} \mathbf{1}_{\{V_i>u\}} \qquad \text{a.s.} \tag{3.5}$$

Since $\mathfrak{N}_n/n \to \pi(\mathcal{C})$ as $n \to \infty$, it follows from (3.2), (3.4) and (3.5) that

$$\mathbf{P}\{V > u\} = \lim_{n\to\infty} \frac{\mathfrak{N}_n}{n} \left(\frac{1}{\mathfrak{N}_n} \sum_{i=0}^{K_{\mathfrak{N}_n}-1} \mathbf{1}_{\{V_i>u\}} \right) = \pi(\mathcal{C}) \mathbf{E}_\gamma[N_u]. \tag{3.6}$$

Finally recall $\mathcal{E}_u := N_u e^{-\xi S_{T_u}} \mathbf{1}_{\{T_u<K\}}$ and hence by an elementary change-of-measure argument [as in (3.18) below], we have $\mathbf{E}_\gamma[N_u] = \mathbf{E}_{\mathfrak{D}}[\mathcal{E}_u]$.

To complete the proof, it remains to show that

$$\lim_{k\to\infty} \mathbf{E}_{\mathfrak{D}}\big[N_u e^{-\xi S_{T_u}} \mathbf{1}_{\{T_u<K\}} | V_0 \sim \hat{\gamma}_k\big] = \mathbf{E}_{\mathfrak{D}}\big[N_u e^{-\xi S_{T_u}} \mathbf{1}_{\{T_u<K\}} | V_0 \sim \gamma\big], \tag{3.7}$$

where $S_n := \sum_{i=1}^n \log A_i$. Set

$$H(v) = \mathbf{E}_{\mathfrak{D}}\big[\mathbf{E}_{\mathfrak{D}}[N_u|\mathfrak{F}_{T_u}] e^{-\xi S_{T_u}} \mathbf{1}_{\{T_u<K\}} | V_0 = v\big]. \tag{3.8}$$

We now claim that $H(v)$ is uniformly bounded in $v \in \mathcal{C}$. To establish this claim, first apply Proposition 4.1 of Collamore and Vidyashankar (2013b) to obtain that

$$\mathbf{E}_{\mathfrak{D}}[N_u|\mathfrak{F}_{T_u}] \mathbf{1}_{\{T_u<K\}} \le \left(C_1(u) \log\left(\frac{V_{T_u}}{u}\right) + C_2(u) \right) \mathbf{1}_{\{T_u<\tau\}}, \tag{3.9}$$

where $\tau \ge K$ is the first regeneration time and $C_i(u) \to C_i < \infty$ as $u \to \infty$ ($i = 1, 2$). Moreover, for $Z_n := V_n/(A_1 \cdots A_n)$, we clearly have

$$e^{-\xi S_{T_u}} = u^{-\xi} \left(\frac{V_{T_u}}{u}\right)^{-\xi} Z_{T_u}^{\xi}. \tag{3.10}$$

Substituting the last two equations into (3.8) yields

$$|H(v)| \le \Theta \mathbf{E}_{\mathfrak{D}}\big[|Z_{T_u}^{\xi} \mathbf{1}_{\{T_u<\tau\}}| | V_0 = v\big] \le \overline{\Theta} \tag{3.11}$$

for finite constants Θ and $\overline{\Theta}$, where the last step was obtained by Collamore and Vidyashankar (2013b), Lemma 5.5(ii). Consequently, $H(v)$ is bounded uniformly in $v \in \mathcal{C}$.

Since $\hat{\gamma}_k$ and γ are both supported on $\mathcal{C}$, it then follows since $\hat{\gamma}_k \Rightarrow \gamma$ that

$$\lim_{k\to\infty} \int_{\mathcal{C}} H(v)\, d\hat{\gamma}_k(v) = \int_{\mathcal{C}} H(v)\, d\gamma(v),$$

which is (3.7). $\square$

Before turning to the proof of efficiency, it will be helpful to have a characterization of the return times of $\{V_n\}$ to the set $\mathcal{C}$ when $Y_n \sim \mu_\beta$ for $\beta \in \operatorname{dom}(\Lambda)$, where $Y_n := (\log A_n, B_n, D_n)$ and μ_β is defined according to (2.4). First let

$$\lambda_\beta(\alpha) = \int_{\mathbb{R}^3} e^{\alpha x}\, d\mu_\beta(x, y, z), \qquad \Lambda_\beta(\alpha) = \log \lambda_\beta(\alpha), \qquad \alpha \in \mathbb{R}$$

and note by the definition of μ_β that

$$\Lambda_\beta(\alpha) = \Lambda(\alpha + \beta) - \Lambda(\beta). \tag{3.12}$$

Recall that if P denotes the transition kernel of $\{V_n\}$, then we say that $\{V_n\}$ satisfies a *drift condition* if there exists a function $h : \mathbb{R} \to [0, \infty)$ such that

$$\int_{\mathbb{S}} h(y) P(x, dy) \le \rho h(x) \qquad \text{for all } x \notin \mathcal{C}, \tag{$\mathcal{D}$}$$

where $\rho \in (0, 1)$ and $\mathcal{C}$ is some Borel subset of $\mathbb{R}$.

LEMMA 3.1. *Assume Letac's model E, and suppose that* (H_1), (H_2) *and* (H_3) *are satisfied. Let $\{V_n\}$ denote the forward recursive sequence generated by this SFPE under the measure μ_β, chosen such that $\inf_{\alpha>0} \lambda_\beta(\alpha) < 1$. Then the drift condition ($\mathcal{D}$) holds with $h(x) = |x|^\alpha$, where $\alpha > 0$ is any constant satisfying the equation $\Lambda_\beta(\alpha) < 0$. Moreover, we may take $\rho = \rho_\beta$ and $\mathcal{C} = [-M_\beta, M_\beta]$, where*

$$\rho_\beta := t\lambda_\beta(\alpha) \qquad \textit{for some } t \in \left(1, \frac{1}{\lambda_\beta(\alpha)}\right) \tag{3.13}$$

and

$$M_\beta := \begin{cases} (\mathbf{E}_\beta[\widetilde{B}^\alpha])^{1/\alpha} (\lambda_\beta(\alpha)(t-1))^{-1/\alpha}, & \textit{if } \alpha \in (0, 1), \\ (\mathbf{E}_\beta[\widetilde{B}^\alpha])^{1/\alpha} ((\lambda_\beta(\alpha))^{1/\alpha}(t^{1/\alpha} - 1))^{-1}, & \textit{if } \alpha \ge 1. \end{cases} \tag{3.14}$$

Furthermore, for any (ρ_β, M_β) satisfying this pair of equations,

$$\sup_{v \in \mathcal{C}} \mathbf{P}_\beta\{K > n | V_0 = v\} \le \rho_\beta^n \qquad \textit{for all } n \in \mathbb{Z}_+. \tag{3.15}$$

PROOF. Let $\widetilde{B}_n := A_n|D_n| + |B_n|$. If $\alpha \ge 1$, then Minkowskii's inequality yields

$$\begin{aligned} &\mathbf{E}_\beta\big[|V_1|^\alpha | V_0 = v\big] \\ &\quad \le \big((\mathbf{E}_\beta[A^\alpha])^{1/\alpha} v + (\mathbf{E}_\beta[\widetilde{B}^\alpha])^{1/\alpha}\big)^\alpha \\ &\quad = \rho_\beta v^\alpha \left(\frac{1}{t^{1/\alpha}} + \frac{(\mathbf{E}_\beta[\widetilde{B}^\alpha])^{1/\alpha}}{\rho_\beta^{1/\alpha} v}\right)^\alpha \qquad \text{where } \rho_\beta := t\lambda_\beta(\alpha). \end{aligned} \tag{3.16}$$

Then $(\mathcal{D})$ is established. For M_β, set $t^{-1/\alpha} + (\mathbf{E}_\beta[\widetilde{B}^\alpha])^{1/\alpha}/(\rho_\beta^{1/\alpha} v) = 1$ and solve for v. Similarly, if $\alpha < 1$, use $|x+y|^\alpha \le |x|^\alpha + |y|^\alpha$, $\alpha \in (0, 1]$, in place of Minkowskii's inequality. Then (3.15) follows by a standard argument, as in Nummelin (1984) or Collamore and Vidyashankar (2013b), Remark 6.2. □

We now introduce some additional notation which will be needed in the proof of Theorem 2.2. Let $A_0 \equiv 1$ and, for any $n = 0, 1, 2, \dots,$ set

$$\mathcal{P}_n = A_0 \cdots A_n, \qquad S_n = \sum_{i=0}^{n} \log A_i,$$

$$Z_n = \frac{V_n}{A_0 \cdots A_n} \quad \text{and} \quad \overline{Z}^{(p)} = \sum_{n=0}^{\infty} \frac{\widetilde{B}_n}{A_0 \cdots A_n} \mathbf{1}_{\{K > n\}},$$

where

$$\widetilde{B}_0 = |V_0| \quad \text{and} \quad \widetilde{B}_n = A_n |D_n| + |B_n|. \tag{3.17}$$

Also introduce the dual measure with respect to an arbitrary measure μ_α, where $\alpha \in \operatorname{dom}(\Lambda)$. Namely, define

$$\mathfrak{L}(\log A_n, B_n, D_n) = \begin{cases} \mu_\alpha, & \text{for } n = 1, \dots, T_u, \\ \mu, & \text{for } n > T_u. \end{cases} \tag{$\mathfrak{D}_\alpha$}$$

Note that it follows easily from this definition that for any r.v. $\mathcal{U}$ which is measurable with respect to $\mathfrak{F}_K$,

$$\mathbf{E}[\mathcal{U} \mathbf{1}_{\{T_u < K\}}] = \mathbf{E}_{\mathfrak{D}}\big[(\lambda(\alpha))^{T_u} e^{-\alpha S_{T_u}} \mathcal{U} \mathbf{1}_{\{T_u < K\}}\big], \tag{3.18}$$

an identity which will be useful in the following.

PROOF OF THEOREM 2.2. Assume $V_0 = v \in \mathcal{C}$. We will show that the result holds uniformly in $v \in \mathcal{C}$.

Case 1: $\lambda(\alpha) < \infty$, for some $\alpha < -\xi$.

To evaluate

$$\mathbf{E}_{\mathfrak{D}}[\mathcal{E}_u^2] := \mathbf{E}_{\mathfrak{D}}[N_u^2 e^{-2\xi S_{T_u}} \mathbf{1}_{\{T_u < K\}}],$$

first note that $V_n e^{-S_n} := V_n / \mathcal{P}_n := Z_n$. Since $V_{T_u} > u$, it follows that $0 \le u e^{-S_{T_u}} \le Z_{T_u}$. Moreover, as in the proof of Lemma 5.5 of Collamore and Vidyashankar (2013b) [cf. (5.27), (5.28)], we obtain

$$Z_n \le \sum_{i=0}^{n} \frac{\widetilde{B}_i}{\mathcal{P}_i} \qquad \text{implying } Z_{T_u} \mathbf{1}_{\{T_u < K\}} \le \sum_{n=0}^{\infty} \frac{\widetilde{B}_n}{\mathcal{P}_n} \mathbf{1}_{\{n \le T_u < K\}}.$$

Consequently,

$$u^{2\xi} \mathbf{E}_{\mathfrak{D}}[\mathcal{E}_u^2] \le \mathbf{E}_{\mathfrak{D}}\left[N_u^2 \left(\sum_{n=0}^{\infty} \frac{\widetilde{B}_n}{\mathcal{P}_n} \mathbf{1}_{\{n \le T_u < K\}} \right)^{2\xi} \right]. \tag{3.19}$$

If $2\xi \geq 1$, apply Minkowskii's inequality to the RHS to obtain

$$\begin{aligned}\left(u^{2\xi}\mathbf{E}_{\mathfrak{D}}[\mathcal{E}_u^2]\right)^{1/2\xi} &\leq \sum_{n=0}^{\infty}\left(\mathbf{E}_{\mathfrak{D}}\left[N_u^2\left(\frac{\widetilde{B}_n}{\mathcal{P}_n}\right)^{2\xi}\mathbf{1}_{\{n\leq T_u<K\}}\right]\right)^{1/2\xi} \\ &= \sum_{n=0}^{\infty}\left(\mathbf{E}\left[N_u^2\mathcal{P}_n^{-\xi}\widetilde{B}_n^{2\xi}\mathbf{1}_{\{n\leq T_u<K\}}\right]\right)^{1/2\xi},\end{aligned} \tag{3.20}$$

where the last step follows from (3.18). Using the independence of $(A_n, \widetilde{B}_n)$ and $\mathbf{1}_{\{n-1<T_u\wedge K\}}$, it follows by an application of Hölder's inequality that the left-hand side (LHS) of (3.20) is bounded above by

$$\sum_{n=0}^{\infty}\left(\mathbf{E}\left[N_u^{2r}\right]\right)^{1/2r\xi}\left(\mathbf{E}\left[(A_n^{-1}\widetilde{B}_n^2)^{s\xi}\right]\right)^{1/2s\xi}\left(\mathbf{E}\left[\mathcal{P}_{n-1}^{-s\xi}\mathbf{1}_{\{n-1<T_u\wedge K\}}\right]\right)^{1/2s\xi},$$

where $r^{-1}+s^{-1}=1$. Set $\zeta = s\xi$ for the remainder of the proof. The last term on the RHS of the previous equation may be expressed in $\mu_{-\zeta}$-measure as

$$\mathbf{E}\left[\mathcal{P}_{n-1}^{-\zeta}\mathbf{1}_{\{n-1<T_u\wedge K\}}\right] = \left(\lambda(-\zeta)\right)^{n-1}\mathbf{P}_{-\zeta}\{n-1<T_u\wedge K\}. \tag{3.21}$$

Substituting this last equation into the upper bound for (3.20), we conclude that

$$\left(u^{2\xi}\mathbf{E}_{\mathfrak{D}}[\mathcal{E}_u^2]\right)^{1/2\xi} \leq \sum_{n=0}^{\infty}\mathcal{J}_n\left(\left(\lambda(-\zeta)\right)^{n-1}\mathbf{P}_{-\zeta}\{n-1<T_u\wedge K\}\right)^{1/2\zeta}, \tag{3.22}$$

where

$$\mathcal{J}_n := \left(\mathbf{E}\left[N_u^{2r}\right]\right)^{1/2r\xi}\left(\mathbf{E}\left[(A_n^{-1}\widetilde{B}_n^2)^{\zeta}\right]\right)^{1/2\zeta}, \qquad n = 0, 1, \ldots.$$

Since $N_u \leq K$, applying Lemma 3.1 with $\beta = 0$ yields

$$\sup_{v\in\mathcal{C}}\mathbf{E}\left[N_u^{2r}|V_0=v\right] < \infty \qquad \text{for any finite constant } r. \tag{3.23}$$

Moreover, for sufficiently small $s > 1$ and $\zeta = s\xi$, it follows by (2.9) that $\mathbf{E}[(A^{-1}\widetilde{B}^2)^{\zeta}] < \infty$. Thus, to show that the quantity on the LHS of (3.22) is finite, it suffices to show for some $\zeta > \xi$ and some $t > 1$,

$$\mathbf{P}_{-\zeta}\{n-1<T_u\wedge K\} \leq \left(t\lambda(-\zeta)\right)^{-n+1} \qquad \text{for all } n \geq N_0, \tag{3.24}$$

where N_0 is a finite positive integer, uniformly in u and uniformly in $v \in \mathcal{C}$.

To this end, note that $\{T_u \wedge K > n-1\} \subset \{K > n-1\}$, and by Lemma 3.1 [using that $\min_\alpha \lambda_{-\zeta}(\alpha) < (\lambda(-\zeta))^{-1}$ by (3.12)],

$$\sup_{v\in\mathcal{C}}\mathbf{P}_{-\zeta}\{K > n-1|V_0=v\} \leq \left(t\lambda(-\zeta)\right)^{-n+1}, \tag{3.25}$$

where $\mathcal{C} := [-M, M]$ and $M > M_{-\xi}$. [Since $\zeta > \xi$ was arbitrary, we have replaced $M_{-\zeta}$ with $M_{-\xi}$ in this last expression. We note that we also require $M > M_0$ for (3.23) to hold.] We have thus established (3.24) for the case $2\xi \geq 1$.

If $2\xi < 1$, then the above argument can be repeated but using the deterministic inequality $|x+y|^\alpha \le |x|^\alpha + |y|^\alpha$, $\alpha \in (0,1]$, in place of Minkowskii's inequality, establishing the theorem for this case.

Case 2: $\lambda(-\zeta) = \infty$ for $\zeta > \xi$, while $\mathbf{E}[(A^{-1}\widetilde{B})^\alpha] < \infty$ for all $\alpha > 0$.

First assume $2\xi \ge 1$. Then, as before, $(u^{2\xi}\mathbf{E}_{\mathfrak{D}}[\mathcal{E}_u^2])^{1/2\xi}$ is bounded above by the RHS of (3.20). In view of the display following (3.20), it is sufficient to show that uniformly in $v \in \mathcal{C}$ (for some set $\mathcal{C} = [-M, M]$),

$$\sup_{n\in\mathbb{Z}_+} \mathbf{E}\big[\mathcal{P}_{n-1}^{-\zeta}\mathbf{1}_{\{n-1<T_u\wedge K\}}\big] < \infty \qquad \text{for some } \zeta > \xi. \tag{3.26}$$

Set $W_n = \mathcal{P}_{n-1}^{-\zeta}\mathbf{1}_{\{n-1<T_u\wedge K\}}$, and first observe that $\mathbf{E}[W_n] < \infty$. Indeed,

$$|V_n| \le A_n|V_{n-1}|\left(1 + \frac{\widetilde{B}_n}{A_n|V_{n-1}|}\right), \qquad n = 1, 2, \dots \tag{3.27}$$

and $n-1 < T_u \wedge K \implies |V_i| \in (M, u)$ for $i = 1, \dots, n-1$. Hence (3.27) implies

$$\begin{aligned} A_i^{-\zeta} \le \left(\frac{u}{M}\right)^\zeta \left(1 + \frac{\widetilde{B}_i}{MA_i}\right)^\zeta, \\ i = 1, \dots, n-1 \text{ on } \{n-1 < T_u \wedge K\}. \end{aligned} \tag{3.28}$$

This equation yields an upper bound for $\mathcal{P}_{n-1}$. Using the assumption that $\mathbf{E}[(A^{-1}\widetilde{B})^\alpha] < \infty$ for all $\alpha > 0$, we conclude by (3.28) that $\mathbf{E}[W_n] < \infty$.

Next let $\{L_k\}$ be a sequence of positive real numbers such that $L_k \downarrow 0$ as $k \to \infty$, and set $F_k = \bigcap_{i=1}^{k-1}\{A_i \ge L_k\}$. Assume that L_k has been chosen sufficiently small such that

$$\mathbf{E}[W_k\mathbf{1}_{F_k^c}] \le \frac{1}{k^2}, \qquad k = 1, 2, \dots. \tag{3.29}$$

Then it suffices to show that

$$\sum_{k=0}^\infty \mathbf{E}[W_k\mathbf{1}_{F_k}] < \infty. \tag{3.30}$$

To verify (3.30), set $\overline{A}_{0,k} = 1$ and introduce the truncation

$$\overline{A}_{n,k} = A_n\mathbf{1}_{\{A_n\ge L_k\}} + L_k\mathbf{1}_{\{A_n<L_k\}}, \qquad n = 1, 2, \dots.$$

Let $\lambda_k(\alpha) = \mathbf{E}[\overline{A}_{1,k}^\alpha]$ and $\overline{W}_k = (\overline{A}_0 \cdots \overline{A}_{k-1})^{-\zeta}\mathbf{1}_{\{k-1<T_u\wedge K\}}$. After a change of measure [as in (3.18), (3.21)], we obtain

$$\mathbf{E}[\overline{W}_k] \le \big(\lambda_k(-\zeta)\big)^{k-1}\mathbf{E}_{-\zeta}[\mathbf{1}_{\{K>k-1\}}\mathbf{1}_{F_k}]. \tag{3.31}$$

To evaluate the expectation on the RHS, start with the inequality

$$|V_{n,k}| \le \overline{A}_{n,k}|V_{n-1,k}|\left(1 + \frac{\widetilde{B}_n}{\overline{A}_{n,k}|V_{n-1,k}|}\right), \qquad n = 1, 2, \dots. \tag{3.32}$$

Write $\mathbf{E}_{-\zeta,w}[\cdot] = \mathbf{E}_{-\zeta}[\cdot | V_{0,k} = w]$. Then for any $\beta > 0$, a change of measure followed by an application of Hölder's inequality yields

$$\begin{aligned} \mathbf{E}_{-\zeta,w}\big[|V_{1,k}|^{\beta}\big] &\le \frac{w^{\beta}}{\lambda_k(-\zeta)} \mathbf{E}\bigg[(\overline{A}_{1,k})^{\beta-\zeta}\bigg(1+\frac{\widetilde{B}_1}{w\overline{A}_{1,k}}\bigg)^{\beta}\bigg] \\ &\le \rho_k w^{\beta}\bigg(t^{-q}\mathbf{E}\bigg[\bigg(1+\frac{\widetilde{B}_1}{w\overline{A}_{1,k}}\bigg)^{q\beta}\bigg]\bigg)^{1/q}, \end{aligned} \tag{3.33}$$

where $\rho_k := (\mathbf{E}[(\overline{A}_{1,k})^{p(\beta-\zeta)}])^{1/p}(t/\lambda_k(-\zeta))$ and $p^{-1} + q^{-1} = 1$.

Set $\hat{\beta} = \arg\min_{\alpha} \lambda(\alpha)$ and choose β such that $p(\beta - \zeta) = \hat{\beta}$, and assume that $p > 1$ is sufficiently small such that $\rho_k < \infty$, $\forall k$. Noting that $\lambda(\hat{\beta}) < 1$, we conclude that for $t \in (1, (\lambda(\hat{\beta}))^{-1/p})$ and for some constant $\rho \in (0, 1)$,

$$\lim_{k\to\infty} \lambda_k(-\zeta)\rho_k := t \lim_{k\to\infty}\big(\mathbf{E}\big[(\overline{A}_{1,k})^{p(\beta-\zeta)}\big]\big)^{1/p} = t\big(\lambda(\hat{\beta})\big)^{1/p} < \rho, \tag{3.34}$$

where the second equality was obtained by observing that as $k \to \infty$, $L_k \downarrow 0$ and hence $\lambda_k(\alpha) \downarrow \lambda(\alpha)$, $\alpha > 0$. Equation (3.34) yields that $\lambda_k(-\zeta)\rho_k \le \rho$ for all $k \ge k_0$, and with this value of ρ, (3.33) yields

$$\mathbf{E}_{-\zeta,w}\big[|V_{1,k}|^{\beta}\big] \le \frac{\rho w^{\beta}}{\lambda_k(-\zeta)} \qquad \text{for all } k \ge k_0, \tag{3.35}$$

provided that

$$t^{-q}\mathbf{E}\bigg[\bigg(1+\frac{\widetilde{B}_1}{w\overline{A}_{1,k}}\bigg)^{q\beta}\bigg] \le 1. \tag{3.36}$$

Our next objective is to find a set $\mathcal{C} = [-M, M]$ such that for all $w \notin \mathcal{C}$, (3.36) holds. First assume $q\beta \ge 1$ and apply Minkowskii's inequality to the LHS of (3.36). Then set this quantity equal to one, solve for w and set $w = M_k$. After some algebra, this yields

$$M_k = \frac{1}{t^{1/\beta} - 1}\bigg(\mathbf{E}\bigg[\bigg(\frac{\widetilde{B}_1}{\overline{A}_{1,k}}\bigg)^{q\beta}\bigg]\bigg)^{1/q\beta}. \tag{3.37}$$

The quantity in parentheses tends to $\mathbf{E}[(A^{-1}\widetilde{B})^{q\beta}]$ as $k \to \infty$. Using the assumption $\mathbf{E}[(A^{-1}\widetilde{B})^{\alpha}] < \infty$ for $\alpha > 0$, we conclude $M := \sup_k M_k < \infty$.

If $q\beta < 1$, then a similar expression is obtained for M by using the deterministic inequality $|x + y|^{\beta} \le |x|^{\beta} + |y|^{\beta}$ in place of Minkowskii's inequality.

To complete the proof, iterate (3.35) with $\mathcal{C} = [-M, M]$ (as in the proof of Lemma 3.1) to obtain that

$$\mathbf{E}_{-\zeta}[\mathbf{1}_{\{K>k-1\}}\mathbf{1}_{F_k}] \le \bigg(\frac{\rho}{\lambda_k(-\zeta)}\bigg)^{-k+1} \qquad \text{for all } k \ge k_0. \tag{3.38}$$

Note that on the set F_k, $\{V_{n,k} : 1 \le n \le k\}$ and $\{V_n : 1 \le n \le k\}$ agree, and thus $\{K > k - 1\}$ coincides for these two sequences. Substituting (3.38) into (3.31) yields (3.30) as required. Finally, the modifications needed when $2\xi < 1$ follow along the lines of those outlined in case 1, so we omit the details. $\square$

4. Examples and simulations. In this section we provide several examples illustrating the implementation of our algorithm.

4.1. *The ruin problem with stochastic investments.* Let the fluctuations in the insurance business be governed by the classical Cramér–Lundberg model,

$$X_t = u + ct - \sum_{n=1}^{N_t} \zeta_n, \tag{4.1}$$

where u denotes the company's initial capital, c its premium income rate, $\{\zeta_n\}$ the claims losses, and N_t the number of Poisson claim arrivals occurring in $[0, t]$. Let $\{\zeta_n\}$ be i.i.d. and independent of $\{N_t\}$. We now depart from this classical model by assuming that at discrete times $n = 1, 2, \dots,$ the surplus capital is invested, earning stochastic returns $\{R_n\}$, assumed to be i.i.d. Let $L_n := -(X_n - X_{n-1})$ denote the losses incurred by the insurance business during the nth discrete time interval. Then the total capital of the insurance company at time n is described by the recursive sequence of equations

$$Y_n = R_n Y_{n-1} - L_n, \qquad n = 1, 2, \dots, \ Y_0 = u, \tag{4.2}$$

where it is typically assumed that $\mathbf{E}[\log R] > 0$ and $\mathbf{E}[L] < 0$.

Our objective is to estimate the probability of ruin,

$$\psi(u) := \mathbf{P}\{Y_n < 0, \text{ for some } n \in \mathbb{Z}_+ | Y_0 = u\}. \tag{4.3}$$

By iterating (4.2), we obtain that $Y_n = (R_1 R_2 \cdots R_n)(Y_0 - \mathcal{L}_n)$, where $\mathcal{L}_n := \sum_{i=1}^n L_i/(R_1 \cdots R_i)$. Thus $\psi(u) = \mathbf{P}\{\mathcal{L}_n > u, \text{ some } n\}$. Setting $\mathcal{L} = (\sup_{n \in \mathbb{Z}_+} \mathcal{L}_n) \vee 0$, then by an elementary argument [as in Collamore and Vidyashankar (2013b), Section 3], we obtain that $\mathcal{L}$ satisfies the SFPE

$$\mathcal{L} \stackrel{\mathcal{D}}{=} (A\mathcal{L} + B)^+ \qquad \text{where } A \stackrel{\mathcal{D}}{=} \frac{1}{R_1} \text{ and } B \stackrel{\mathcal{D}}{=} \frac{L_1}{R_1}. \tag{4.4}$$

This can be viewed as a special case of Letac's model E with $D := -B/A$.

Now take

$$A_n = \exp\left\{-\left(\mu - \frac{\sigma^2}{2}\right) - \sigma Z_n\right\} \qquad \text{for all } n, \tag{4.5}$$

where $\{Z_n\}$ is an i.i.d. sequence of standard Gaussian r.v.'s. It can be seen that $\xi = 2\mu/\sigma^2 - 1$ and $\mu_\xi \sim \text{Normal}(\mu - \sigma^2/2, \sigma^2)$.

We set $\mu = 0.2$, $\sigma^2 = 0.25$, $c = 1$, $\{\zeta_n\} \sim \text{Exp}(1)$ and let $\{N_t\}$ be a Poisson process with parameter $1/2$.

We implemented our algorithm to estimate the probabilities of ruin for $u = 10, 100, 10^3, 10^4, 10^5$. In all of our simulations, the distribution in step 1 was based on $k = 10^4$, and V_{1000} was taken as an approximation to the limit r.v. V. We arrived at this choice using extensive exploratory analysis and two-sample comparisons using Kolmogorov–Smirnov tests between V_{1000} and other values of V_n, where $n = 2000$, 5000, 10,000 (with p-values ≥ 0.185). Also, it is worthwhile to point out here that by Sanov's theorem and Markov chain theory, the difference between the approximating V_{n^*} and V on $\mathcal{C}$ is exponentially small, since $\mathcal{C}$ is in the center of the distribution of V.

In implementing the algorithm, we chose $M = 0$, since, arguing as in the proof of Lemma 3.1, we obtain that $M_\beta = \min_{i=1,2} M_\beta^{(i)}$, where

$$
\begin{aligned}
M_\beta^{(1)} &= \inf_{\alpha \in (0,1) \cap \Phi} \frac{\|B_1^+\|_{\beta,\alpha}}{(1 - \|A_1\|_{\beta,\alpha}^\alpha)^{1/\alpha}}, \\
M_\beta^{(2)} &= \inf_{\alpha \in [1,\infty) \cap \Phi} \frac{\|B_1^+\|_{\beta,\alpha}}{1 - \|A_1\|_{\beta,\alpha}}
\end{aligned}
\tag{4.6}
$$

and $\Phi = \{\alpha \in \mathbb{R} : \mathbf{E}_\beta[A^\alpha] < 1\}$. (Here $\|\cdot\|_{\beta,\alpha}$ denotes the L_α norm under the measure μ_β.) As previously, we consider two cases, $\beta = 0$ and $\beta = -\xi$. For each of these cases, this infimum is computed numerically, yielding $M_0 = 0 = M_{-\xi}$.

Table 1 summarizes the probabilities of ruin (with $M = 0$) and the lower and upper bounds of the 95% confidence intervals (LCL, UCL) based on 10^6 simulations. The confidence intervals in this and other examples in this section are based on the simulations; that is, the lower 2.5% and upper 97.5% quantiles of the simulated values of $\mathbf{P}\{V > u\}$. We also evaluated the true constant $C(u) := \mathbf{P}\{V > u\}u^\xi$ [which would appear in (1.2) if this expression were exact], and the relative error (RE). Even in the extreme tail—far below the probabilities of practical interest in this problem—our algorithm works effectively and is clearly seen to have bounded relative error. For comparison, we also present the crude Monte Carlo estimates of the probabilities of ruin based on 5×10^6 realizations of V_{2000}. We observe that for small values of u, the importance sampling estimates and the crude Monte Carlo estimates are close, *which provides an empirical validation of the algorithm for small values of* u.

4.2. *The* ARCH(1) *process*. Now consider the ARCH(1) process, which models the squared returns on an asset via the recurrence equation

$$R_n^2 = (a + bR_{n-1}^2)\zeta_n^2 = A_n R_{n-1}^2 + B_n, \qquad n = 1, 2, \ldots,$$

TABLE 1
Importance sampling estimation for the ruin probability with investments obtained using $M = 0$

u	**P{V > u}**	**LCL**	**UCL**	**C**	**RE**	**Crude est.**
1.0e+01	5.86e−02	5.65e−02	6.07e−02	2.33e−01	1.84e+01	5.73e−02
1.0e+02	1.33e−02	1.28e−02	1.39e−02	2.11e−01	2.12e+01	1.29e−02
1.0e+03	3.27e−03	3.14e−03	3.41e−03	2.07e−01	2.12e+01	3.21e−03
1.0e+04	8.13e−04	7.78e−04	8.49e−04	2.04e−01	2.24e+01	8.01e−04
1.0e+05	1.98e−04	1.90e−04	2.07e−04	1.98e−01	2.16e+01	2.10e−04

where $A_n = b\zeta_n^2$, $B_n = a\zeta_n^2$, and $\{\zeta_n\}$ is an i.i.d. Gaussian sequence. Setting $V_n = R_n^2$, we see that $V := \lim_{n\to\infty} V_n$ satisfies the SFPE $V \stackrel{\mathcal{D}}{=} AV + B$, and it is easy to verify that the assumptions of our theorems are satisfied. Then it is of interest to determine $\mathbf{P}\{V > u\}$ for large u.

Next we implement our algorithm to estimate these tail probabilities. As in the previous example, we identify V_{1000} as an approximation to V. Turning to identification of M, recall that in the previous example, we worked with a sharpened form of the formulas in Lemma 3.1; however, in other examples, this approach may, like Lemma 3.1, yield a poor choice for M. This is due to the fact that these types of estimate for V_n^α typically use Minkowskii- or Hölder-type inequalities, which are usually not very sharp. We now outline an alternative method for obtaining M and demonstrate that it yields meaningful answers from a practical perspective. In the numerical method, we work directly with the conditional expectation and avoid upper-bound inequalities. We emphasize that this procedure applies to any process governed by Letac's model E.

Numerical procedure for calculating M. The procedure involves a Monte Carlo method for calculating the conditional expectation appearing in the drift condition, that is, for evaluating

$$\mathbf{E}_\beta\left[\left(\frac{V_1}{V_0}\right)^\alpha \middle| V_0 = v\right] = \mathbf{E}_\beta\left[\left(A\max\left\{\frac{D}{v}, 1\right\} + \frac{B}{v}\right)^\alpha\right],$$

when $\beta = 0$ and $\beta = -\xi$. The goal is to find an α such that $M := \max\{M_0, M_{-\xi}\}$ is minimized, where M_β satisfies

$$\mathbf{E}_\beta\left[\left(A\max\left\{\frac{D}{v}, 1\right\} + \frac{B}{v}\right)^\alpha\right] \le \rho_\beta \qquad \text{for all } v > M_\beta \text{ and some } \rho_\beta \in (0, 1).$$

In this expression, α is chosen such that $\mathbf{E}_\beta[A^\alpha] \in (0, 1)$, and hence we expect that $\rho_\beta \in (\mathbf{E}_\beta[A^\alpha], 1)$. Note that M_β depends on the choice of α; thus, we also minimize over all possible α such that $\mathbf{E}_\beta[A^\alpha] \in (0, 1)$.

Let $\{(A_i, B_i, D_i) : 1 \le i \le N\}$ denote a collection of i.i.d. r.v.'s having the same distribution as (A, B, D). Then the numerical method for finding an optimal choice of M proceeds as follows.

First, using a root finding algorithm such as Gauss–Hermite quadrature, solve for ξ in the equation $\mathbf{E}[A^\xi] = 1$. Next, for $\mathbf{E}_\beta[A^\alpha] < 1$, use a Monte Carlo procedure with sample size N to compute $\mathbf{E}_\beta[|V_1|^\alpha | V_0 = v]$ and solve for v in the formula

$$\frac{1}{N}\sum_{i=1}^N \left|A_i \max\left\{\frac{D_i}{v}, 1\right\} + \frac{B_i}{v}\right|^\alpha = \rho_\beta,$$

where this quantity is computed in the β-shifted measure for $\beta \in \{0, -\xi\}$ and where $\rho_\beta < 1$. Then select α so that it provides the smallest possible value of v. Choose $M_\beta > v$ for $\beta = 0$ and $\beta = -\xi$. Finally, set $M = \max\{M_0, M_{-\xi}\}$.

Implementation. We set $b = 4/5$ and considered the values $a : 1.9 \times 10^{-5}, 1$. It can be shown that

$$\mathbf{E}[A_n^{\alpha}] = \frac{(2b)^{\alpha}\Gamma(\alpha + 1/2)}{\Gamma(1/2)}.$$

We solved the equation $\mathbf{E}[A_n^{\xi}] = 1$ using Gauss–Hermite quadrature to obtain $\xi = 1.3438$. Under the ξ-shifted measure, $A_n = bX_n$ and $B_n = aX_n$, where $X_n \sim \Gamma(\xi + 1/2, 2)$. Using the formulas in (4.6) for M, we obtained [upon taking the limit as $\delta \to 0$ and using the Taylor approximation $\Gamma(\delta + 1/2) = \Gamma(1/2) + \delta\Gamma'(1/2) + O(\delta^2)$] that $M_0 = 0.362, 6.879 \times 10^{-6}$ when $a = 1$, 1.9×10^{-5}, respectively. Moreover, by applying the numerical method we have just outlined, it can be seen that $M_{-\xi} = 0$. [In contrast, by applying Lemma 3.1 directly, one obtains $M_{-\xi} = \infty$ since $\lambda(-\xi) = \infty$.]

Table 2 summarizes the simulation results for the tail probabilities of the ARCH(1) process based on 10^6 simulations. We notice a substantial agreement between the crude Monte Carlo estimates and those produced by our algorithm for small values of u. More importantly, we observe that the relative error remains bounded in all of the cases considered, while the simulation results using the state-dependent algorithm in Blanchet, Lam and Zwart (2012) *show that the relative error based on their algorithm increases as the parameter $u \to \infty$*. When compared with the state-independent algorithm of Blanchet, Lam and Zwart (2012), our simulations give comparable numerical results to those they report, although direct comparison is difficult due to the unquantified role of bias in their formulas. (In contrast, from a numerical perspective, the bias is negligible in our formulas,

TABLE 2
Importance sampling estimation for the tail probability of ARCH(1) financial process with $a = 1$, 1.9×10^{-5}

u	**P{$V > u$}**	**LCL**	**UCL**	**C**	**RE**	**Crude est.**
			$a = 1$			
1.0e+01	7.73e−02	7.64e−02	7.83e−02	1.71e+00	6.21e+00	7.75e−02
1.0e+02	4.34e−03	4.23e−03	4.45e−03	2.11e+00	1.29e+01	4.28e−03
1.0e+03	2.04e−04	1.99e−04	2.09e−04	2.20e+00	1.28e+01	2.07e−04
1.0e+04	9.00e−06	8.88e−06	9.12e−06	2.14e+00	6.83e+00	9.00e−06
1.0e+05	4.11e−07	4.04e−07	4.18e−07	2.15e+00	8.51e+00	NA
			$a = 1.9 \times 10^{-5}$			
1.0e+01	4.45e−08	4.38e−08	4.52e−08	9.82e−07	8.38e+00	NA
1.0e+02	2.02e−09	1.98e−09	2.05e−09	9.82e−07	9.29e+00	NA
1.0e+03	9.59e−11	8.77e−11	1.04e−10	1.03e−06	4.38e+01	NA
1.0e+04	4.15e−12	4.05e−12	4.26e−12	9.85e−07	1.32e+01	NA
1.0e+05	1.91e−13	1.83e−13	1.99e−13	1.00e−06	2.19e+01	NA

as it involves the convergence of a Markov chain near the *center* of its distribution, which is known to occur at a geometric rate.) We emphasize that our method also applies to a wider class of problems, as illustrated by the previous example. Finally, we remark that a variant of the ARCH(1) process is the GARCH(1, 1) financial process, which can be implemented by similar methods. Numerical results for this model are roughly analogous, but further complications arise which can be addressed as in our preprint under the same title in Math arXiv. For a further discussion of examples governed by Letac's model E and its generalizations, see Collamore and Vidyashankar (2013b), Section 3.

5. Proofs of results concerning running time of the algorithm. The proof of the first estimate will rely on the following.

LEMMA 5.1. *Under the conditions of Theorem 2.3, there exist positive constants β and $\rho \in (0, 1)$ such that*

$$\mathbf{E}_\xi\big[h(V_n)|V_{n-1}\big] \le \rho h(V_{n-1}) \qquad \text{on } \{V_{n-1} \ge \overline{M}\} \tag{5.1}$$

for some $\overline{M} < \infty$, where $h(x) := x^{-\beta}\mathbf{1}_{\{x>1\}} + \mathbf{1}_{\{x\le 1\}}$.

PROOF. Assume without loss of generality (w.l.o.g.) that $V_{n-1} = v > 1$. Then by the strong Markov property,

$$\mathbf{E}_\xi\big[h(V_n)|V_{n-1} = v\big] = \mathbf{E}_\xi\big[V_1^{-\beta}\mathbf{1}_{\{V_1>1\}}|V_0 = v\big] + \mathbf{P}_\xi\{V_1 \le 1|V_0 = v\}.$$

Using assumption (2.11), we obtain that the second term on the RHS is $o(v^{-\varepsilon})$, while the first term can be expressed as

$$v^\beta \mathbf{E}_\xi\big[\big(A_1 \max\{v^{-1}D_1, 1\} + v^{-1}B_1\big)^{-\beta}\mathbf{1}_{\{V_1>1\}}|V_0 = v\big] \sim v^\beta \mathbf{E}_\xi\big[A_1^{-\beta}\big]$$

as $v \to \infty$. Next observe that $\mathbf{E}_\xi[A_1^{-\beta}] = \lambda(\xi - \beta) < 1$ if $0 < \beta < \xi$. Thus, choosing $\beta = \varepsilon \in (0, \xi)$, where ε is given as in (2.11), we obtain that the lemma holds for any $\rho = (\mathbf{E}_\xi[A_1^{-\varepsilon}], 1)$ and $\overline{M} < \infty$ sufficiently large. □

PROOF OF THEOREM 2.3. We will prove (2.12)–(2.14) in three steps, each involving separate ideas and certain preparatory lemmas.

PROOF OF THEOREM 2.3, STEP 1. *Equation* (2.12) *holds.* Let $\overline{M}$ be given as in Lemma 5.1, and assume w.l.o.g. that $\overline{M} \ge \max\{M, 1\}$. Let $L \equiv \sup\{n \in \mathbb{Z}_+ : V_n \in (-\infty, \overline{M}]\}$ denote the last exit time of $\{V_n\}$ from $(-\infty, \overline{M}]$. Then it follows directly from the definitions that $K \le L$ on $\{K < \infty\}$, where we recall that K is the return time to the $\mathcal{C}$-set. Thus it is sufficient to verify that $\mathbf{E}_\xi[L] < \infty$.

To this end, we introduce two sequences of random times. Set $\mathcal{J}_0 = 0$ and $\mathcal{K}_0 = 0$ and, for each $i \in \mathbb{Z}_+$,

$$\mathcal{K}_i = \inf\{n > \mathcal{J}_{i-1} : V_n > \overline{M}\} \quad \text{and} \quad \mathcal{J}_i = \inf\{n > \mathcal{K}_i : V_n \in (-\infty, \overline{M}]\}.$$

Our main interest is in $\{\mathcal{K}_i\}$, the successive times that the process escapes from the interval $(-\infty, \overline{M}]$, and $\kappa_i := \mathcal{K}_i - \mathcal{K}_{i-1}$.

Let $\mathfrak{N}$ denote the total number of times that $\{V_n\}$ exits $(-\infty, \overline{M}]$ and subsequently returns to $(-\infty, \overline{M}]$. Then it follows that

$$L < \sum_{i=1}^{\mathfrak{N}+1} \kappa_i.$$

Then by the transience of $\{V_n\}$ in μ_ξ-measure, it follows that $\mathbf{E}_\xi[\mathfrak{N}] < \infty$.

It remains to show that $\mathbf{E}_\xi[\kappa_i] < \infty$, uniformly in the starting state $V_{\kappa_{i-1}} \in (\overline{M}, \infty]$. But note that the $\mathbf{E}_\xi[\kappa_i]$ can be divided into two parts; first, the sojourn time that the process $\{V_n\}$ spends in $(\overline{M}, \infty)$ prior to returning to $(-\infty, \overline{M}]$ and, second, the sojourn time in the interval $(-\infty, \overline{M}]$ prior to exiting again. Now if $\overline{K}$ denotes the first return time to $(-\infty, \overline{M}]$, then by Lemma 5.1,

$$\mathbf{P}_\xi\{\overline{K} = n | V_0 = v\} \le \rho^n \frac{h(v)}{h(\overline{M})} \le \rho^n.$$

Hence $\mathbf{E}_\xi[\overline{K}\mathbf{1}_{\{\overline{K}<\infty\}} | V_0 = v] \le \Theta < \infty$, uniformly in $v > \overline{M}$.

Thus, to establish the lemma, it is sufficient to show that $\mathbf{E}_\xi[\overline{N} | V_0 = v] < \infty$, uniformly in $v \in (-\infty, \overline{M}]$, where $\overline{N}$ denotes the total number of visits of $\{V_n\}$ to $(-\infty, \overline{M}]$. To this end, first note that $[-\overline{M}, \overline{M}]$ is petite. Moreover, it is easy to verify that $(-\infty, -\overline{M})$ is also petite for sufficiently large $\overline{M}$. Indeed, for large $\overline{M}$ and $V_0 < -\overline{M}$, (1.1) implies $V_1 = A_1 D_1 + B_1$ w.p. $p > 0$. Thus, $\{V_n\}$ satisfies a minorization with small set $(-\infty, -\overline{M})$. Consequently $(-\infty, \overline{M}]$ is petite and hence uniformly transient. We conclude $\mathbf{E}_\xi[\overline{N}] < \infty$, uniformly in $V_0 \in (-\infty, \overline{M}]$. □

Before proceeding to step 2, we need a slight variant of Lemma 4.1 in Collamore and Vidyashankar (2013b). In the following, let A^l be a typical ladder height of the process $S_n = \sum_{i=1}^n \log A_i$ in its ξ-shifted measure.

LEMMA 5.2. *Assume the conditions of Theorem 2.3. Then*

$$\lim_{u\to\infty} \mathbf{P}_\xi\left\{\frac{V_{T_u}}{u} > y \middle| T_u < K\right\} = \mathbf{P}_\xi\{\widehat{V} > y\} \tag{5.2}$$

for some r.v. $\widehat{V}$, where for all $y \ge 0$,

$$\mathbf{P}_\xi\{\log \widehat{V} > y\} = \frac{1}{\mathbf{E}_\xi[A^l]} \int_y^\infty \mathbf{P}_\xi\{A^l > z\}\, dz. \tag{5.3}$$

PROOF. It can be shown that

$$\frac{V_{T_u}}{u} \Rightarrow \widehat{V} \qquad \text{as } u \to \infty \tag{5.4}$$

in μ_ξ-measure, independent of $V_0 \in \mathcal{C}$ [see Collamore and Vidyashankar (2013b), Lemma 4.1].

Set $y > 1$. Then by (5.4), $\mathbf{P}_\xi\{V_{T_u}/u > y\} \to \mathbf{P}_\xi\{\widehat{V} > y\}$ as $u \to \infty$; and using the independence of this result on its initial state, we likewise have that $\mathbf{P}_\xi\{V_{T_u}/u > y | T_u \geq K\} \to \mathbf{P}_\xi\{\widehat{V} > y\}$ as $u \to \infty$. Hence we conclude (5.2), provided that $\liminf_{u\to\infty} \mathbf{P}_\xi\{T_u < K\} > 0\}$. But by the transience of $\{V_n\}$, $\mathbf{P}_\xi\{T_u < K\} \to \mathbf{P}_\xi\{K = \infty\} > 0$ as $u \to \infty$. $\square$

PROOF OF THEOREM 2.3, STEP 2. *Equation* (2.13) *holds*. With respect to the measure μ_ξ, it follows by Lemma 9.13 of Siegmund (1985) that

$$\frac{T_u}{\log u} \to \frac{1}{\Lambda'(\xi)} \qquad \text{in probability} \tag{5.5}$$

(since $\Lambda'(\xi) = \mathbf{E}_\xi[\log A]$). Hence, conditional on $\{T_u < K\}$, $(T_u/\log u) \to (\Lambda'(\xi))^{-1}$ in probability.

To show that convergence in probability implies convergence in expectation, it suffices to show that the sequence $\{T_u/\log u\}$ is uniformly integrable. Let $\overline{M}$ be given as in Lemma 5.1, and first suppose that $\overline{M} \leq M$ and $\operatorname{supp}(V_n) \subset [-M, \infty)$ for all n. Then, conditional on $\{T_u < K\}$,

$$T_u > n \quad \Longrightarrow \quad V_i \in (\overline{M}, u), \qquad i = 1, \dots, n.$$

Now apply Lemma 5.1. Iterating (5.1), we obtain $\mathbf{E}[h(V_n)\prod_{i=1}^n \mathbf{1}_{V_i \notin \mathcal{C}} | V_0] \leq \rho^n h(V_0)$, $n = 1, 2, \dots$. Then, using the explicit form of the function h in Lemma 5.1, we conclude that with β given as in Lemma 5.1,

$$\mathbf{P}_\xi\{T_u > n | T_u < K\} \leq \left(\frac{1}{\mathbf{P}_\xi\{T_u < K\}}\right)\rho^n u^\beta \qquad \text{for all } n. \tag{5.6}$$

Now $\mathbf{P}_\xi\{T_u < K\} \downarrow \Theta > 0$ as $u \to \infty$. Hence, letting $\mathbf{E}_\xi^{(u)}[\cdot]$ denote the expectation conditional on $\{T_u < K\}$, we obtain that for some $\overline{\Theta} < \infty$,

$$\mathbf{E}_\xi^{(u)}\left[\frac{T_u}{\log u}; \frac{T_u}{\log u} \geq \eta\right] \leq \overline{\Theta}\rho^{\eta \log u} u^\beta \tag{5.7}$$

and for sufficiently large η, the RHS converges to zero as $u \to \infty$. Hence $\{T_u/\log u\}$ is uniformly integrable.

If the assumptions at the beginning of the previous paragraph are not satisfied, then write $T_u = L + (T_u - L)$, where L is the last exit time from the interval $(-\infty, \overline{M}]$, as defined in the proof of Theorem 2.3, step 1. Then $(T_u - L)$ describes the length of the last excursion to level u after exiting $(-\infty, \overline{M}]$ forever. By a repetition of the argument just given, we obtain that (5.6) holds with $(T_u - L)$ in place of T_u; hence $\{(T_u - L)/\log u\}$ is uniformly integrable. Next observe by the proof of Theorem 2.3, step 1, that $\mathbf{E}_\xi[L/\log u] \downarrow 0$ as $u \to \infty$. The result follows. $\square$

Turning now to the proof of the last equation in Theorem 2.3, assume for the moment that $(V_0/u) = v > 1$ (we will later remove this assumption); thus, the process starts above level u and so its dual measure agrees with its initial measure. Also define

$$L(z) = \inf\{n : |V_n| \le z\} \qquad \text{for any } z \ge 0.$$

LEMMA 5.3. *Let $(V_0/u) = v > 1$ and $t \in (0,1)$. Then under the conditions of Theorem 2.3,*

$$\lim_{u\to\infty} \frac{1}{\log u} \mathbf{E}\left[L(u^t) \middle| \frac{V_0}{u} = v \right] = \frac{1-t}{|\Lambda'(0)|}. \tag{5.8}$$

PROOF. For notational simplicity, we will suppress the conditioning on $(V_0/u) = v$ in the proof. We begin by establishing an upper bound. Define

$$S_n^{(u)} := \sum_{i=1}^{n} X_i^{(u)} \qquad \text{where } X_i^{(u)} := \log\bigl(A_i + u^{-t}\bigl(A_i|D_i| + |B_i|\bigr)\bigr).$$

Then it can be easily seen that

$$\log|V_n| - \log(vu) \le S_n^{(u)} \qquad \text{for all } n < L\bigl(u^t\bigr). \tag{5.9}$$

Now let $\widetilde{L}_u(u^t) = \inf\{n : S_n^{(u)} \le -(1-t)\log u - \log v\}$. Then $L(u^t) \le \widetilde{L}_u(u^t)$ for all u.

By Wald's identity, $\mathbf{E}[S_{\widetilde{L}_u(u^t)}] = \mathbf{E}[X_1^{(u)}]\mathbf{E}[\widetilde{L}_u(u^t)]$. Thus, letting

$$O_u := \bigl|S_{\widetilde{L}_u(u^t)} - (1-t)\log u - \log v\bigr|$$

denote the overjump of $\{S_n^{(u)}\}$ over a boundary at level $(1-t)\log u + \log v$, we obtain

$$L\bigl(u^t\bigr) \le \frac{(1-t)\log u + \log v + \mathbf{E}[O_u]}{|\mathbf{E}[X_1^{(u)}]|}. \tag{5.10}$$

Since $\mathbf{E}[X_1^{(u)}] \to \Lambda'(0)$ as $u \to \infty$, the required upper bound will be established once we show that

$$\lim_{u\to\infty} \frac{1}{\log u} \mathbf{E}[O_u] = 0. \tag{5.11}$$

To establish (5.11), note as in the proof of Lorden's inequality [Asmussen (2003), Proposition V.6.1] that $\mathbf{E}[O_u] \le \mathbf{E}[Y_u^2]/\mathbf{E}[Y_u]$, where Y_u has the negative ladder height distribution of the process $\{S_n^{(u)}\}$. Next observe by Corollary VIII.4.4 of Asmussen (2003) that

$$\mathbf{E}[Y_u] = m_u^{(1)} e^{\mathfrak{S}_u} \to \mathbf{E}[Y] \qquad \text{as } u \to \infty, \tag{5.12}$$

where Y has the negative ladder height distribution of $\{S_n\}$, and $m_u^{(j)} := |\mathbf{E}[X^{(u)}]|$, $j = 1, 2, \ldots$ and $\mathfrak{S}_u := \sum_{n=1}^{\infty} n^{-1}\mathbf{P}\{S_n^{(u)} > 0\}$. We observe that $\mathfrak{S}_u$ is the so-called Spitzer series. Similarly, an easy calculation [cf. Siegmund (1985), page 176] yields

$$\mathbf{E}[Y_u^2] = m_u^{(2)} e^{\mathfrak{S}_u} - 2m_u^{(1)} e^{\mathfrak{S}_u} \sum_{n=1}^{\infty} \frac{1}{n}\mathbf{E}[(S_n^{(u)})^+] \to \mathbf{E}[Y^2], \qquad u \to \infty. \tag{5.13}$$

Since $\mathbf{E}[(\log A)^3] < \infty \Longrightarrow \mathbf{E}[Y^j] < \infty$ for $j = 1, 2$, it follows that $\mathbf{E}[O_u] \to \mathbf{E}[Y^2]/\mathbf{E}[Y] < \infty$, implying (5.11). Thus (5.8) holds as an upper bound.

To establish a corresponding lower bound, fix $s \in (t, 1)$ and define

$$\widetilde{L}(u^s) = \inf\{n : S_n \le -(1-s)\log u - \log v\}.$$

Observe that $V_n \ge A_n V_{n-1} - |B_n|$ for $V_{n-1} \ge 0$, and iterating yields

$$V_n \ge (A_1 \cdots A_n)V_0 - W \qquad \text{where } W := \lim_{n\to\infty} \sum_{i=1}^{n} \prod_{j=i+1}^{n} A_j |B_i|. \tag{5.14}$$

Since $(V_0/u) = v$, it follows from the definition of $\widetilde{L}$ that

$$\widetilde{L}(u^s) \ge n \quad \Longleftrightarrow \quad (A_1 \cdots A_k)V_0 > u^s \qquad \text{for all } k < n.$$

But by (5.14), $(A_1 \cdots A_k)V_0 > u^s \Longrightarrow V_k > u^t$ on $\{W \le (u^s - u^t)\}$. Thus for all n, $\widetilde{L}(u^s) \ge n \Longrightarrow L(u^t) \ge n$ on $\{W \le (u^s - u^t)\}$, and consequently

$$\mathbf{E}[L(u^t)] \ge \mathbf{E}[\widetilde{L}(u^s); W \le (u^s - u^t)]. \tag{5.15}$$

Next recall that for some $\overline{C} > 0$,

$$\mathbf{P}\{W > u^s - u^t\} \sim \overline{C} u^{-s\xi} \qquad \text{as } u \to \infty. \tag{5.16}$$

As $\widetilde{L}(u^t)$ is the time required for the negative-drift random walk $\{S_n + \log v\}$ to reach the level $-(1-s)\log u$, Heyde's (1966) a.s. convergence theorem for renewal processes gives that

$$\frac{\widetilde{L}(u^s)}{\log u} \to \frac{(1-s)}{|\Lambda'(0)|} \qquad \text{a.s. as } u \to \infty \tag{5.17}$$

(since $\mathbf{E}[\log A] = \Lambda'(0) < 0$). Hence for any $\varepsilon > 0$,

$$\lim_{u\to\infty} \mathbf{P}\left\{\frac{\widetilde{L}(u^t)}{\log u} \notin (r - \varepsilon, r + \varepsilon)\right\} = 0 \qquad \text{where } r := \frac{1-s}{|\Lambda'(0)|}. \tag{5.18}$$

Substituting (5.16) and (5.18) into (5.15) and letting $\varepsilon \to 0$, we obtain

$$\liminf_{u\to\infty} \frac{1}{\log u}\mathbf{E}[L(u^t)] \ge \frac{1-s}{|\Lambda'(0)|}. \tag{5.19}$$

The required lower bound follows by letting $s \downarrow t$. $\square$

LEMMA 5.4. *Assume the conditions of the previous lemma. Then*

$$\lim_{t\downarrow 0}\left\{\limsup_{u\to\infty}\frac{1}{\log u}\mathbf{E}\big[L(M)-L\big(u^t\big)\big]\right\}=0. \tag{5.20}$$

PROOF. Apply Lemma 3.1 with $\beta=0$ to obtain that, for some $\alpha>0$,

$$\mathbf{E}\big[|V_n|^\alpha|V_{n-1}=w\big]\le\rho|w|^\alpha \qquad \text{for all } w\notin\mathcal{C},$$

where $\rho\in(0,1)$ and $\mathcal{C}=[-M,M]$, for some positive constant M. Since this equation holds for all $n<L(M)$ (the first entrance time into the set $\mathcal{C}$), iterating this equation yields

$$\mathbf{E}[\mathbf{1}_{\{L(M)>n\}}|V_0=w]\le\rho^n\left(\frac{|w|}{M}\right)^\alpha \qquad \text{for all } n. \tag{5.21}$$

Now apply this equation to obtain an estimate for $L(M)-L(u^t)$. Since $|V_{L(u^t)}|\le u^t$, the previous equation [with $V_{L(u^t)}$ in place of V_0] gives

$$\mathbf{P}\big\{L(M)-L\big(u^t\big)>n\big\}\le\rho^n\left(\frac{u^t}{M}\right)^\alpha \qquad \text{for all } n. \tag{5.22}$$

Set $J_t(u)=L(M)-L(u^t)$ and $t'=t\alpha/(-\log\rho)$. Summing (5.22) over all $n\ge t'\log u$ yields that

$$\mathbf{E}\big[J_t(u)\mathbf{1}_{\{J_t(u)\ge t'\log u\}}\big]\le\frac{\rho^{t'\log u}}{1-\rho}\left(\frac{u^t}{M}\right)^\alpha=\frac{1}{(1-\rho)M^\alpha}. \tag{5.23}$$

Hence

$$\limsup_{u\to\infty}\frac{1}{\log u}\mathbf{E}\big[L(M)-L\big(u^t\big)\big]\le t'. \tag{5.24}$$

Since $t'\downarrow 0$ as $t\downarrow 0$, we conclude (5.20). □

PROOF OF THEOREM 2.3, STEP 3. *Equation* (2.14) *holds.* By Lemmas 5.3 and 5.4,

$$H_u(v):=\frac{1}{\log u}\mathbf{E}\left[L(M)\middle|\frac{V_0}{u}=v\right]\to\frac{1}{|\Lambda'(0)|} \qquad \text{as } u\to\infty. \tag{5.25}$$

Let $\hat\mu_u$, $\hat\mu$ denote the probability laws of the r.v.'s V_{T_u}/u, $\widehat{V}$ appearing in the statement of Lemma 5.2. Then, using the strong Markov property, it follows that $L(M)$, conditional on $V_0/u\sim\hat\mu_u$, is equal in distribution to $K-T_u$, conditional on $\{T_u<K\}$. Thus it is sufficient to verify that

$$\lim_{u\to\infty}\frac{1}{\log u}\mathbf{E}\left[L(M)\middle|\frac{V_0}{u}\sim\hat\mu_u\right]:=\lim_{u\to\infty}\int_{v\ge 0}H_u(v)\,d\hat\mu_u(v)=\frac{1}{|\Lambda'(0)|}. \tag{5.26}$$

This result will follow from (5.25), provided that we can show that the limit can be taken inside the integral in the above equation.

To do so, express the inner quantity in (5.26) as

$$\int_{v\geq 0} H_u(v)\,d(\hat{\mu}_u-\hat{\mu})(v)+\int_{v\geq 0} H_u(v)\,d\hat{\mu}(v). \tag{5.27}$$

To deal with the first term, begin by obtaining an upper bound for $H_u(v)$. First note by a slight modification of (5.23) [with $t=1$, $\theta=-\log\rho$ and $J_1(u)$, t' replaced with $L(M)$, r, resp.] that

$$\mathbf{E}\big[L(M)\mathbf{1}_{\{L(M)>r\log u\}}\big||V_0|\leq u\big]\leq \frac{u^{-r\theta}}{1-\rho}\left(\frac{u}{M}\right)^{\alpha} \tag{5.28}$$

for all $r>0$ and some $\alpha>0$. Now choose $r>\alpha/\theta$. Then the RHS is bounded above by $\Theta_1<\infty$, independent of u. Consequently,

$$\frac{1}{\log u}\mathbf{E}\big[L(M)\big||V_0|\leq u\big]\leq r+\frac{\Theta_1}{\log u}. \tag{5.29}$$

Next, we extend this estimate to the case where $(V_0/u)=v>1$. To this end, viewing an excursion time as the sum of the time to first reach $[-u,u]$ and then reach $\mathcal{C}$, we obtain

$$\mathbf{E}\left[L(M)\middle|\frac{V_0}{u}=v\right]\leq \sup_{w\in(M,u]}\mathbf{E}\big[L(M)\big||V_0|=w\big]+\mathbf{E}\left[L(u)\middle|\frac{V_0}{u}=v\right]. \tag{5.30}$$

For the second term, observe

$$|V_{n-1}|>u \quad\Longrightarrow\quad |V_n|\leq |V_{n-1}|\left(A_n+\frac{\widetilde{B}_n}{u}\right);$$

thus, $\mathbf{E}[L(u)|(V_0/u)=v]$ is bounded above by the length of time for the classical random walk

$$S_n^{(u)}:=S_{n-1}^{(u)}+\log\left(A_n+\frac{\widetilde{B}_n}{u}\right),\qquad n=1,2,\ldots,$$

starting from $S_0^{(u)}=\log(vu)$, to reach the level $\log u$. Denote this sojourn time by $L^*(u)$. Applying Lorden's inequality [Asmussen (2003), Proposition V.6.1] to $\{S_n^{(u)}\}$, we obtain [with $\Lambda'''(0)<\infty$] that

$$\mathbf{E}\big[L^*(u)\big]\leq \Theta_2(u)\log v+\Theta_3(u)\to\frac{\log v}{m_1}+\frac{m_2}{m_1^2},\qquad u\to\infty,$$

where m_i denotes the ith moment of the ladder height distribution for the sequence $\{\log A_i\}$; cf. the discussion following (5.12) above. Substituting this last bound and (5.29) into (5.30), we deduce that for some constant $\overline{\Theta}$, uniformly in $u\geq u_0$ for some finite constant u_0,

$$H_u(v):=\frac{1}{\log u}\mathbf{E}\left[L(M)\middle|\frac{V_0}{u}=v\right]\leq\overline{\Theta}+\frac{2\log v}{m_1}. \tag{5.31}$$

Returning to (5.27) and using the above upper bound, we now show that

$$\left|\int_{v\geq 0}\left(\overline{\Theta}+\frac{2\log v}{m_1}\right)d(\hat{\mu}_u-\hat{\mu})(v)\right|\to 0 \qquad \text{as } u\to\infty. \tag{5.32}$$

Since $\hat{\mu}_u \Rightarrow \hat{\mu}$, by Lemma 5.2, it is sufficient to show that $\int_{v\geq 0}\log v\, d\hat{\mu}_u(v)$ is uniformly bounded in u, which would follow from the uniform integrability of $\{|\log V_{T_u}-\log u|\}$. To this end, we apply the corollary to Theorem 2 of Lai and Siegmund (1979). Note that $V_{T_u}=\widetilde{V}_{T_u}$, where $\widetilde{V}_n=S_n+\delta_n$ for a sequence $\{\delta_n\}$ which is slowly changing [cf. Collamore and Vidyashankar (2013b), Lemma 4.1]. Also, using Collamore and Vidyashankar (2013b), Lemma 5.5, it is easy to verify that

$$\frac{\xi}{2}\mathbf{E}_\xi\big[|\delta_{T_u}-\delta_{T_u-1}|\mathbf{1}_{\{T_u<K\}}\big]\leq \mathbf{E}_\xi\big[\log(\overline{Z}^{(p)})^\xi\big]<\infty. \tag{5.33}$$

Note that conditions (6)–(8) of Lai and Siegmund (1979) are also satisfied with $\alpha=1$. In this regard, notice that Theorem 2 of their article is actually valid if their equation (8) is replaced by uniform continuity in probability of $\{\delta_n\}$, as given in equation (4.2) of Woodroofe (1982), and the latter condition holds since δ_n converges w.p.1 to a proper r.v. We conclude $\{|\log V_{T_u}-\log u|\}$ is uniformly integrable. Then (5.32) follows since $\hat{\mu}_u\Rightarrow\hat{\mu}$.

Finally, applying the dominated convergence theorem to the second term in (5.27) and invoking (5.25), we conclude $(\log u)^{-1}\mathbf{E}[L(M)|(V_0/u)\sim\hat{\mu}]\to 1/|\Lambda'(0)|$, as required. □

6. Proof of optimality. The idea of the proof is similar to Collamore (2002), Theorem 3.4, but new technical issues arise since we deal with a process generated by (1.1) rather than a random walk process.

PROOF OF THEOREM 2.4. Let $\boldsymbol{\nu}\in\mathfrak{M}$. First we show that

$$\liminf_{u\to\infty}\frac{1}{\log u}\mathbf{E}_{\boldsymbol{\nu}}\big[(\mathcal{E}_u^{(\boldsymbol{\nu})})^2\big]\geq -2\xi. \tag{6.1}$$

To establish (6.1), set

$$\mu_{\mathfrak{D}}(E;w,q)=\begin{cases}\mu_\xi(E), & E\in\mathcal{B}(\mathbb{R}^3),\ w\in\mathbb{R}\ \text{ and }\ q=0;\\ \mu(E), & E\in\mathcal{B}(\mathbb{R}^3),\ w\in\mathbb{R}\ \text{ and }\ q=1.\end{cases}$$

(Intuitively, w corresponds to the level of the process $\{\log V_{n-1}/\log u\}$, while $q=1$ indicates that $\{V_n\}$ has exceeded level u by the previous time.)

If $\nu\ll\mu_{\mathfrak{D}}$, then by a standard argument [cf. Collamore (2002), equations (4.54), (4.55)], utilizing the Radon–Nikodym theorem,

$$\begin{aligned}\mathbf{E}_{\boldsymbol{\nu}}\big[(\mathcal{E}_u^{(\boldsymbol{\nu})})^2\big] &:= \mathbf{E}_{\boldsymbol{\nu}}\left[N_u^2\mathbf{1}_{\{T_u<K\}}\prod_{i=1}^K\left(\frac{d\mu}{d\nu}(Y_i;W_i,Q_i)\right)^2\right]\\ &= \mathbf{E}_{\mathfrak{D}}\left[N_u^2\mathbf{1}_{\{T_u<K\}}\prod_{i=1}^K\left(\frac{d\mu}{d\mu_{\mathfrak{D}}}(Y_i;W_i,Q_i)\right)^2\frac{d\mu_{\mathfrak{D}}}{d\nu}(Y_i;W_i,Q_i)\right].\end{aligned}$$

Note $\frac{d\mu}{d\mu_{\mathfrak{D}}} = \frac{d\mu}{d\mu_\xi}$ for $Q_i = 0$, while $\frac{d\mu}{d\mu_{\mathfrak{D}}} = 1$ for $Q_i = 1$. Hence

$$\mathbf{E}_{\boldsymbol{\nu}}\big[\big(\mathcal{E}_u^{(\boldsymbol{\nu})}\big)^2\big] = \mathbf{E}_{\mathfrak{D}}\left[N_u^2 \mathbf{1}_{\{T_u<K\}} \prod_{i=1}^{T_u}\left(\frac{d\mu}{d\mu_\xi}(Y_i)\right)^2 \prod_{j=1}^{K} \frac{d\mu_{\mathfrak{D}}}{d\nu}(Y_j; W_j, Q_j)\right]. \tag{6.2}$$

Thus setting

$$U_i = \log\left(\frac{d\nu}{d\mu_{\mathfrak{D}}}(Y_i; W_i, Q_i)\right) \quad \text{and} \quad R_n = \sum_{i=1}^{n} U_i,$$

we conclude by Jensen's inequality that

$$\begin{aligned} \mathbf{E}_{\boldsymbol{\nu}}\big[\big(\mathcal{E}_u^{(\boldsymbol{\nu})}\big)^2\big] &= \mathbf{E}_{\mathfrak{D}}\big[N_u^2 \mathbf{1}_{\{T_u<K\}} e^{-2\xi S_{T_u} - R_K}\big] \\ &\geq p_u \exp\big\{\mathbf{E}_{\mathfrak{D}}[-2\xi S_{T_u} - R_K | T_u < K]\big\}, \end{aligned} \tag{6.3}$$

where $p_u := \mathbf{P}_\xi\{T_u < K\} \to \Theta > 0$ as $u \to \infty$. It follows from (6.3) that

$$\begin{aligned} \liminf_{u\to\infty} \frac{1}{\log u} \log \mathbf{E}_{\boldsymbol{\nu}}\big[\big(\mathcal{E}_u^{(\boldsymbol{\nu})}\big)^2\big] &\geq -\limsup_{u\to\infty} \frac{1}{\log u} \mathbf{E}_\xi[2\xi S_{T_u} \mathbf{1}_{\{T_u<K\}}] \\ &\quad - \limsup_{u\to\infty} \frac{1}{\log u} \mathbf{E}_{\mathfrak{D}}[R_K \mathbf{1}_{\{T_u<K\}}]. \end{aligned} \tag{6.4}$$

To identify the first term on the RHS of (6.4), note by Wald's identity that

$$\mathbf{E}_\xi[\log A]\mathbf{E}_\xi[T_u \wedge K] = \mathbf{E}_\xi[S_{T_u} \mathbf{1}_{\{T_u<K\}}] + \mathbf{E}_\xi[S_K \mathbf{1}_{\{K \leq T_u\}}]. \tag{6.5}$$

Now $(\log u)^{-1}\mathbf{E}_\xi[T_u \wedge K] \to (\log u)^{-1}\mathbf{E}_\xi[T_u | T_u < K]$ as $u \to \infty$ (by Theorem 2.3). Also, $\mathbf{E}_\xi[S_K \mathbf{1}_{\{K\leq T_u\}}] \to \mathbf{E}[S_K e^{-\xi S_K} \mathbf{1}_{\{K<\infty\}}]$ as $u \to \infty$, which is obviously finite on $\{S_K > 0\}$, and which is finite on $\{S_K \leq 0\}$ since (as $e^{\xi x} \geq 1 + \xi x, x > 0$) it can be bounded by a constant multiple of $\mathbf{E}_\xi[e^{-\xi S_K} \mathbf{1}_{\{S_K\leq 0, K<\infty\}}] = \mathbf{E}[\mathbf{1}_{\{S_K \leq 0, K<\infty\}}] < \infty$. Thus, using that $\mathbf{E}_\xi[\log A] = \Lambda'(\xi)$, it follows from Theorem 2.3, equation (2.13), and the above discussion that the middle term of (6.5) must satisfy

$$\lim_{u\to\infty} \frac{1}{\log u} \mathbf{E}_\xi[S_{T_u} \mathbf{1}_{\{T_u<K\}}] = 1. \tag{6.6}$$

To handle the second limit on the RHS of (6.4), first assume, for the moment, that $\log(\frac{d\nu}{d\mu_{\mathfrak{D}}})$ is bounded from below by a finite constant. This assumption will later be removed. Recall that $U_i = \log(\frac{d\nu}{d\mu_{\mathfrak{D}}}(Y_i; W_i, Q_i))$ and $R_n = \sum_{i=1}^n U_i$. Now it follows by an application of Jensen's inequality that

$$\begin{aligned} \mathbf{E}_{\mathfrak{D}}\big[U_n | (W_n, Q_n) = (w, q)\big] &= \int_{\mathbb{R}^3} \log\left(\frac{d\nu}{d\mu_{\mathfrak{D}}}(y; w, q)\right) d\mu_{\mathfrak{D}}(y; w, q) \\ &\leq \log \int_{\mathbb{R}^3} d\nu(y; w, q) = 0 \end{aligned} \tag{6.7}$$

[where we have suppressed the dependence on (w, q) in the above integrals], and consequently, after a short argument, we conclude that

$$\mathcal{M}_n := R_n \mathbf{E}_{\mathfrak{D}}[\mathbf{1}_{\{T_u < K\}} | \mathfrak{F}_n]$$

is a supermartingale. Hence by the optional sampling theorem,

$$\limsup_{u \to \infty} \frac{1}{\log u} \mathbf{E}_{\mathfrak{D}}[R_K \mathbf{1}_{\{T_u < K\}}] \le 0. \tag{6.8}$$

Then (6.1) follows from (6.6) and (6.8). If $\log(\frac{d\nu}{d\mu_{\mathfrak{D}}})$ is not bounded from below by a constant, then we can replace ν with a larger measure, $\nu^{(\varepsilon)} := \nu + \varepsilon \mu_{\mathfrak{D}}$, where $\varepsilon > 0$. Then the entire proof can be repeated without significant change, and we again conclude (6.1) upon letting $\varepsilon \downarrow 0$. We omit the details, which are straightforward.

Next, we show that strict inequality holds in (6.1) when $\boldsymbol{\nu} \in \mathfrak{M}$ differs from the dual measure. Now if $\nu \neq \mu_{\mathfrak{D}}$, then, in view of (6.7), there exists a point (w, q) where

$$\mathbf{E}_{\mathfrak{D}}[U_n | W_n = w, Q_n = q] = -2\Delta \qquad \text{for some } \Delta > 0. \tag{6.9}$$

Then, from the definition of U and an application of the Radon–Nikodym theorem, it follows from the continuity assumption (C_0) that for some neighborhood G of w,

$$\mathbf{E}_{\mathfrak{D}}[U_n | W_n = w, Q_n = q] \le -\Delta, \qquad w \in G. \tag{6.10}$$

We now show that by sharpening the estimate in Jensen's inequality on the set $G \times \{q\}$, we obtain a *strict* inequality in (6.1). As before, we begin by assuming that $\log(\frac{d\nu}{d\mu_{\mathfrak{D}}})$ is bounded from below by a constant. Then by repeating our previous argument, but using the sharper estimate (6.10) when $w \in G$ and q given as in (6.9), together with Jensen's inequality for the remaining values of (w, q), we obtain that

$$\mathcal{M}_n^* := (U_1^* + \cdots + U_n^*) \mathbf{E}_{\mathfrak{D}}[\mathbf{1}_{\{T_u < K\}} | \mathfrak{F}_n],$$
$$U_i^* := U_i + \Delta \mathbf{1}_{\{W_n \in G\}} \mathbf{1}_{\{Q_n = q\}},$$

is a supermartingale. Applying the optional sampling theorem, we deduce that

$$\mathbf{E}_{\mathfrak{D}}[R_K \mathbf{1}_{\{T_u < K\}}] \le -\Delta \{ \mathbf{1}_{\{q=0\}} \mathbf{E}_{\mathfrak{D}}[\mathcal{O}_u^{(0)}] + \mathbf{1}_{\{q=1\}} \mathbf{E}_{\mathfrak{D}}[\mathcal{O}_u^{(1)}] \}, \tag{6.11}$$

where

$$\mathcal{O}_u^{(0)} := \sum_{n=0}^{T_u} \mathbf{1}_{\{W_n \in G\}}$$

and

$$\mathcal{O}_u^{(1)} := \sum_{n=T_u+1}^{K} \mathbf{1}_{\{W_n \in G\}}.$$

Note that $\mathcal{O}_u^{(0)}$ denotes the occupation time which the scaled process $\{\log V_n/\log u\}$ spends in the interval G during a trajectory starting at time 0 and ending at time T_u, while $\mathcal{O}_u^{(1)}$ denotes the occupation time that $\{\log V_n/\log u\}$ spends in the interval G during a trajectory starting at time T_u and ending at time K. Note that for all $n \in \mathbb{Z}_+$,

$$V_0 - W \le \frac{V_n}{A_1 \cdots A_n} \le V_0 + W \qquad \text{where } W := \sum_{i=1}^{\infty} \frac{|B_i| + A_i|D_i|}{A_1 \cdots A_i}.$$

Now suppose that $G' := [u^{s'}, u^{t'}] \subset [u^s, u^t] \subset G$, where $s < s' < t' < t$. Then in the ξ-shifted measure, the transient process $\{V_n\}$ enters G' w.p. $p_u \to \Theta > 0$. Now, in the previous equation, take V_0 to be the position of this process at its first passage time into G', so that $V_0 \ge u^{s'}$. Since W is a proper r.v. w.p.1 in the ξ-shifted measure, it follows that for some $\varepsilon > 0$, $\mathbf{P}\{(V_0 + W)/(V_0 - W) - 1 > u^{-\varepsilon}\} \to 0$ as $u \to \infty$ (and an analogous estimate holds when $V_0 < W$). Thus we see that $\{\log V_n\}$ is well-approximated by $\{S_n\}$. Since, as a multiplicative random walk, the occupation time of $\{e^{S_n}\}$ in G' is at least $c \log u$ for some $c > 0$, we conclude (after a short argument) that

$$\liminf_{u\to\infty} \frac{1}{\log u} \mathbf{E}\big[\mathcal{O}_u^{(0)}\big] \ge \eta > 0. \tag{6.12}$$

Substituting this estimate into (6.11) yields, for the case $q = 0$ in (6.9), that

$$\limsup_{u\to\infty} \frac{1}{\log u} \mathbf{E}_{\mathfrak{D}}[R_K \mathbf{1}_{\{T_u<K\}}] \le -\Delta\eta < 0. \tag{6.13}$$

Now substituting (6.13) and (6.6) into (6.4), we obtain that the LHS of (6.4) is $\ge -2\xi + \Delta\eta$, as required.

If $q = 1$ in (6.9), the argument is similar. Here we study a trajectory in the original measure, beginning at the level V_{T_u} and returning to the set K. Setting $V_0 \overset{\mathcal{D}}{=} V_{T_u}$, then we may again observe that $\{\log V_n\}$ behaves similarly to a random walk or, more precisely,

$$\sup_n \left| V_n - V_0 \prod_{i=1}^{n} A_i \right| \le W' \qquad \text{where } W' := \sum_{i=1}^{\infty} \widetilde{B}_i \prod_{j=i+1}^{\infty} A_j \tag{6.14}$$

as long as $\{V_0, \dots, V_{n-1}\}$ is nonnegative. Then by a straightforward argument based on the law of large numbers,

$$\liminf_{u\to\infty} \frac{1}{\log u} \mathbf{E}\big[\mathcal{O}_u^{(1)}\big] \ge \tilde{\eta} > 0 \tag{6.15}$$

and so we obtain that the LHS of (6.4) is $\ge -2\xi + \Delta\tilde{\eta}$. (For more details, see our preprint under the same title in Math arXiv.)

If $\log(\frac{d\nu}{d\mu_{\mathfrak{D}}})$ is not bounded from below by a constant, then replace ν with $\nu^{(\varepsilon)} := \nu + \varepsilon\mu_{\mathfrak{D}}$, where $\varepsilon > 0$, and the proof carries through with little modification. Finally, to complete the proof of theorem, note that if we do not have $\nu \ll \mu_{\mathfrak{D}}$, as we have assumed throughout this proof, then by an application of the Radon–Nikodym theorem, $\nu = \nu_a + \nu_s$, where $\nu_a \ll \mu_{\mathfrak{D}}$ and $\nu_s \perp \mu_{\mathfrak{D}}$. The proof can now be repeated, replacing everywhere ν with ν_a; cf. Collamore (2002), proof of Theorem 3.4. We omit the details. $\square$

REFERENCES

ALSMEYER, G., IKSANOV, A. and RÖSLER, U. (2009). On distributional properties of perpetuities. *J. Theoret. Probab.* **22** 666–682. MR2530108

ASMUSSEN, S. (2003). *Applied Probability and Queues*, 2nd ed. Springer, New York. MR1978607

ASMUSSEN, S. and GLYNN, P. W. (2007). *Stochastic Simulation*: *Algorithms and Analysis*. Springer, New York. MR2331321

BLANCHET, J., LAM, H. and ZWART, B. (2012). Efficient rare-event simulation for perpetuities. *Stochastic Process. Appl.* **122** 3361–3392. MR2956109

COLLAMORE, J. F. (2002). Importance sampling techniques for the multidimensional ruin problem for general Markov additive sequences of random vectors. *Ann. Appl. Probab.* **12** 382–421. MR1890070

COLLAMORE, J. F. (2009). Random recurrence equations and ruin in a Markov-dependent stochastic economic environment. *Ann. Appl. Probab.* **19** 1404–1458. MR2538076

COLLAMORE, J. F. and VIDYASHANKAR, A. N. (2013a). Large deviation tail estimates and related limit laws for stochastic fixed point equations. In *Random Matrices and Iterated Random Functions* (G. Alsmeyer and M. Löwe, eds.) 91–117. Springer, Heidelberg.

COLLAMORE, J. F. and VIDYASHANKAR, A. N. (2013b). Tail estimates for stochastic fixed point equations via nonlinear renewal theory. *Stochastic Process. Appl.* **123** 3378–3429. MR3071384

COLLAMORE, J. F., VIDYASHANKAR, A. N. and XU, J. (2013). Rare event simulation for stochastic fixed point equations related to the smoothing transform. In *Proceedings of the Winter Simulation Conference* 555–563.

DUPUIS, P. and WANG, H. (2005). Dynamic importance sampling for uniformly recurrent Markov chains. *Ann. Appl. Probab.* **15** 1–38. MR2115034

GOLDIE, C. M. (1991). Implicit renewal theory and tails of solutions of random equations. *Ann. Appl. Probab.* **1** 126–166. MR1097468

HEYDE, C. C. (1966). Some renewal theorems with application to a first passage problem. *Ann. Math. Statist.* **37** 699–710. MR0193692

ISCOE, I., NEY, P. and NUMMELIN, E. (1985). Large deviations of uniformly recurrent Markov additive processes. *Adv. in Appl. Math.* **6** 373–412. MR0826590

KESTEN, H. (1973). Random difference equations and renewal theory for products of random matrices. *Acta Math.* **131** 207–248. MR0440724

LAI, T. L. and SIEGMUND, D. (1979). A nonlinear renewal theory with applications to sequential analysis. II. *Ann. Statist.* **7** 60–76. MR0515684

LETAC, G. (1986). A contraction principle for certain Markov chains and its applications. In *Random Matrices and Their Applications* (*Brunswick*, *Maine*, 1984). *Contemp. Math.* **50** 263–273. Amer. Math. Soc., Providence, RI. MR0841098

NUMMELIN, E. (1984). *General Irreducible Markov Chains and Nonnegative Operators*. Cambridge Univ. Press, Cambridge. MR0776608

SIEGMUND, D. (1976). Importance sampling in the Monte Carlo study of sequential tests. *Ann. Statist.* **4** 673–684. MR0418369

SIEGMUND, D. (1985). *Sequential Analysis*: *Tests and Confidence Intervals*. Springer, New York. MR0799155

VERVAAT, W. (1979). On a stochastic difference equation and a representation of nonnegative infinitely divisible random variables. *Adv. in Appl*. *Probab.* **11** 750–783. MR0544194

WOODROOFE, M. (1982). *Nonlinear Renewal Theory in Sequential Analysis*. SIAM, Philadelphia, PA. MR0660065

J. F. COLLAMORE
DEPARTMENT OF MATHEMATICAL SCIENCES
UNIVERSITY OF COPENHAGEN
UNIVERSITETSPARKEN 5
DK-2100 COPENHAGEN Ø
DENMARK
E-MAIL: collamore@math.ku.dk

G. DIAO
A. N. VIDYASHANKAR
DEPARTMENT OF STATISTICS
GEORGE MASON UNIVERSITY
4400 UNIVERSITY DRIVE, MS 4A7
FAIRFAX, VIRGINIA 22030
USA
E-MAIL: gdiao@gmu.edu
avidyash@gmu.edu