Submitted
arXiv: math.PR/0000000

RARE EVENT SIMULATION FOR PROCESSES GENERATED VIA STOCHASTIC FIXED POINT EQUATIONS

By Jeffrey F. Collamore*, Guoqing Diao, and Anand N. Vidyashankar†

University of Copenhagen and George Mason University*

In a variety 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 $V \stackrel{\mathcal{D}}{=} f(V)$, for some random function f. This paper is concerned with computational methods for evaluating these tail probabilities. We introduce a novel dynamic importance sampling algorithm, involving an exponential shift over a *random* time interval, for estimating such 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 our 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 that demonstrate the ease and scope of the implementation.

1. Introduction. Recently there has been much attention focused on the study of stochastic fixed point equations. This interest has been largely fueled by contemporary applications, including aspects of financial time series modeling, actuarial mathematics, the study of perpetuities, and the analysis of algorithms. An original motivation for this article came from one such application, namely the desire to provide a useful and implementable algorithm—that is theoretically justified—for calculating the probability of

*Corresponding author. Research supported in part by Danish Research Council (SNF) Grant "Point Process Modelling and Statistical Inference," No. 09-092331.

†Research Supported by grants from NDC Health Corporation and NSF DMS 000-03-07057.

AMS 2000 subject classifications: Primary 65C05, 91G60, 68W40, 60H25; secondary 60F10, 60G40, 60J05, 60J10, 60J22, 60K15, 60K20, 60G70, 68U20, 91B30, 91B70, 91G70.

Keywords 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.

ruin for an insurance company under the Cramér-Lundberg model, but in the presence of stochastic investments. A closely related problem is that of obtaining precise estimates for the tail probabilities in the GARCH(1,1) and ARCH(1) financial time series models, which are needed for Value-at-Risk or expected shortfall calculations in risk management. In spite of the relevance of these and a number of related problems, effective computational methods for evaluating these rare event probabilities have, to date, been investigated rather poorly.

This paper is concerned with importance sampling methods for calculating the stationary tail probabilities of a stochastic fixed point equation (SFPE). In general, an SFPE assumes the form

$$V \stackrel{\mathcal{D}}{=} f(V), \tag{1.1}$$

where f is a random function satisfying certain regularity conditions and is independent of V. When $f(v) = Av+B$, where $\mathbf{E}[\log A] < 0$, this problem has a long history beginning with the works of Solomon (1972), Kesten (1973), and Vervaat (1979). In this work, we consider a generalization of their model, namely Letac's (1986) Model E, whose SFPE is given by

$$V \stackrel{\mathcal{D}}{=} A\max(D, V) + B. \tag{1.2}$$

This recurrence equation is known to be quite general in the sense that it characterizes a wide variety of mathematical problems of applied importance. These applications include the ruin problem with investments, the GARCH(1,1) and ARCH(1) financial time series models, the AR(1) time series model with random coefficients, and the classical GI/G/1 queue, among others.

In a recent work, Collamore and Vidyashankar (2011) extended earlier results of Kesten (1973) and Goldie (1991), establishing

$$\lim_{u\to\infty} u^{\xi}\mathbf{P}(V > u) = C \tag{1.3}$$

for finite positive constants C and ξ, where C is identified as the difference of the ξth moments of a perpetuity sequence and another "backward sequence," and ξ is identified as the nonzero solution to the equation $\mathbf{E}\left[A^{\xi}\right] = 1$. To obtain an *exact* estimate for $\mathbf{P}(V > u)$, however, it is natural to turn to computational methods and particularly to importance sampling.

In large deviation problems for sums of i.i.d. or Markov random walks, importance sampling methods have been developed by a number of authors. Following the lines of Hammersley and Handscomb (1964), an early work is Siegmund (1976), who introduced an algorithm for computing the level crossing probabilities in the sequential probability ratio test. This method was later generalized to other large deviation problems in, *e.g.*, Asmussen (1985), Lehtonen and Nyrhinen (1992), Bucklew et al. (1990), Collamore (2002), Chan and Lai (2007), and Blanchet and Liu (2010).

The problem we consider here is different, since we study processes with *both* multiplicative and additive components. Indeed, if one directly adopts importance sampling methods for either light- or heavy-tailed sums, then the resulting estimator will not be efficient. Instead, we propose a different approach, involving a *dual* change of measure over a random time interval, simulated over the finite-time excursions of $\{V_n\}$ emanating from a certain set $\mathcal{C} \subset \mathbb{R}$ and returning to this set. This dual change of measure can be viewed as a *dynamic importance sampling* algorithm, as introduced formally in Dupuis and Wang (2005); *i.e.*, the change of measure depends on the outcome of the simulation experiment.

The motivation for our algorithm is the observation that the SFPE (1.3) induces a forward recursive sequence, namely,

$$V_n = A_n \max(D_n, V_{n-1}) + B_n, \quad n = 1, 2, \dots, \quad V_0 = v, \tag{1.4}$$

where $\{(A_n, B_n, D_n) : n \in \mathbb{Z}_+\}$ is an i.i.d. sequence having the same law as (A, B, D). Then $\{V_n\}$ is a Harris recurrent Markov chain, and hence it returns with probability one to any set $\mathcal{C}$ intersecting the support of its stationary measure. Thus we may study $\{V_n\}$ over the excursions from a set $\mathcal{C}$. It is important to observe that in many applications, such as with perpetuities, the mathematical process under study is usually obtained through the *backward* iterates of the given SFPE (as described Letac (1986) or Collamore and Vidyashankar (2011), Section 2.1). In particular, the linear recursion $f(v) = Av+B$ induces the backward recursive sequence or *perpetuity sequence*

$$Z_n = V_0 + \frac{B_1}{A_1} + \frac{B_2}{A_1 A_2} + \cdots + \frac{B_n}{A_1 \dots A_n}, \quad n = 1, 2, \dots. \tag{1.5}$$

However, $\{Z_n\}$ is not Markovian, and it is much less natural to simulate the sequence $\{Z_n\}$ as compared with the 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. We note that the ruin problem with investments is similar, where again the underlying process is typically described via a backward sequence.

We remark that another importance sampling algorithm, specifically for perpetuities with positive B_i's, was recently proposed in Blanchet et al. (2011). Their approach is considerably different from ours, as they analyze (1.5) directly, and thus their methods do not generalize to the other processes studied in this paper, such as the ruin problem with investments or further processes governed by Letac's Model E. Moreover, in the context of perpetuity sequences, their methods yield quite different results compared with ours; see the discussion in Section 4 below. In another direction, although distinct from the main focus of this paper, there is also an interesting connection of our work to the problem of simulating the stationary distribution of perpetuities, which has received some recent attention in the literature (*cf.* Fill and Huber (2010) or Devroye and Fawzi (2010)).

We conclude this introduction with a brief discussion of our results and some of the main contributions of this article. Section 2 is devoted to a description of our main results. Utilizing a dual change of measure over a random time interval and the forward recursive sequence $\{V_n\}$ generated by the given SFPE, we obtain an algorithm which, based on our main theorems, is: (i) consistent; (ii) asymptotically efficient in the sense that it has bounded relative error; and (iii) optimal among a wide class of possible importance sampling algorithms.

In Section 3, we provide a proof of consistency and asymptotic efficiency for our algorithm. The proof of consistency requires that we relate $\mathbf{P}(V > u)$ to the number of exceedances of $\{V_n\}$ over a cycle which emanates from a given set $\mathcal{C} \subset \mathbb{R}$ and terminates upon the return to $\mathcal{C}$. The representation formula we obtain involves an embedding of the given Markov chain into a Markov additive process and uses aspects of Markov renewal theory.

In the proof of efficiency, we encounter perpetuity sequences similar to (1.5), albeit more general. In the context of (1.5), it is known that while the series converges to a finite limit under minimal conditions, the necessary and sufficient condition for the L_β convergence of $\{Z_n\}$ is that $\mathbf{E}\left[A^{-\beta}\right] < 1$; *cf.* Alsmeyer et al. (2009). However, our analysis will involve moments of quantities similar to $\{Z_n\}$, but where $\mathbf{E}\left[A^{-\beta}\right]$ is *greater* than one, and hence our perpetuity sequences will necessarily diverge in L_β. To circumvent this problem, we will study these perpetuity sequences over randomly stopped intervals in the sense described in the previous paragraph, that is, over cycles emanating from, and returning to, a given subset $\mathcal{C}$ of $\mathbb{R}$. Thus, instead we study Z_K, where K denotes the first return time to $\mathcal{C}$, and we establish the L_β convergence of this quantity. We establish this convergence not only under the assumption that $\mathbf{E}\left[A^{-\beta}\right] \in (1, \infty)$ but, more surprisingly, also when $\mathbf{E}\left[A^{-\beta}\right] = \infty$. This result, obtained in the proof of Theorem 2.3 below, will rely on estimates for the return times of $\{V_n\}$ to $\mathcal{C}$. It is worth noting that if 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 *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 which we consider here 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.

In Section 4, which contains examples, numerical results and a discussion of the implementation of the algorithm, we describe various numerical strategies for dealing with models of practical interest, such as ruin models with stochastic investments and the ARCH(1) and GARCH(1,1) financial time series models.

Sections 5 and 6 are devoted to sharp asymptotic estimates for the running time of the algorithm and to the optimality of the algorithm, respectively. In particular, motivated by the Wentzell-Freidlin theory of large deviations, we consider other possible measures for generating the process $\{V_n\}$, which are allowed to depend on the level of the scaled process $\{\log V_{n-1}/\log u\}$ and on whether or not $\{V_n\}$ has exceeded u prior to the present time. We then show that our algorithm is, in fact, the *only one* which attains bounded relative error, establishing rather definitively the validity of the algorithm, at least in an asymptotic sense as the tail parameter $u \to \infty$. In proving these results, particularly relating to the running time of the algorithm, we encounter a variety of nonstandard issues which we resolve using techniques involving first entrance and last exit times of the Markov chain $\{V_n\}$ generated under various measures, and nonlinear renewal theory. Finally, some concluding remarks are given in Section 7.

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

2.1. *Background, hypotheses, and the algorithm.* In this section we introduce our importance sampling algorithm and discuss its theoretical properties. We begin by describing the so-called forward and backward iterates of an SFPE. In general, an SFPE can be written as a function of the unknown random variable V and an environmental random vector Y, specifically,

$$V \stackrel{\mathcal{D}}{=} F(V,Y), \tag{2.1}$$

where $F : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ is a deterministic measurable function which is continuous in its first component. For convenience, we also write $F_y(v) = F(v,y)$, where $y \in \mathbb{R}^d$. In particular, Letac's (1986) Model E—which will be the main focus of this article—is given by $V \stackrel{\mathcal{D}}{=} F_Y(V)$, where

$$F_Y(v,Y) = A \max(D,v) + B \quad \text{and} \quad Y = (A,B,D),$$

where typically (A, B, D) are correlated.

Let v be an element of the range of F, and let $\{Y_n\}$ be an independent and identically distributed (i.i.d.) sequence of random variables such that $Y_n \stackrel{\mathcal{D}}{=} Y$ for all n. Then the forward equation generated by the SFPE (2.1) is defined by

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

and the backward equation generated by this SFPE is defined by

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

While the forward equation is always Markovian, the backward equation need not be Markovian; however, for every v and every 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—the natural way to analyze the process is through its forward iterates.

Returning to Letac's Model E, the forward equation reduces to

$$V_n = A_n \max\left(D_n, V_{n-1}\right) + B_n, \quad n = 1, 2, \ldots, \quad V_0 = v, \tag{2.4}$$

which upon iteration yields (with $B_0 \equiv V_0$) that

$$V_n = \max\left(\sum_{i=0}^{n} B_i \prod_{j=i+1}^{n} A_j, \bigvee_{k=1}^{n} \left[\sum_{i=k}^{n} B_i \prod_{j=i+1}^{n} A_j + D_k \prod_{j=k}^{n} A_j \right] \right). \tag{2.5}$$

The regularity properties of $\{V_n\}$ have been described in Collamore and Vidyashankar (2011). To state the properties relevant for this article, we first need to introduce some notations and hypotheses, as follows. Let

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

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

$$\mu_\alpha(E) := \frac{1}{\lambda(\alpha)} \int_E e^{\alpha x} d\mu(x, y, z), \quad E \in \mathcal{B}(\mathbb{R}^3), \quad \alpha \in \mathbb{R}, \tag{2.6}$$

where here and in the following, $\mathcal{B}(\mathbb{R}^d)$ denotes the Borel sets of $\mathbb{R}^d$, for any $d \in \mathbb{Z}_+$. Let $\mathbf{E}_\alpha\left[\cdot\right]$ denote expectation with respect to this α-shifted measure.

For any random variable 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 $\{X_i\}$, 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' denote its first derivative, $f^{(2)}$ its second derivative, and so on.

Hypotheses:

(H_1): The random variable A has an absolutely continuous component with respect to Lebesgue measure and satisfies $\Lambda(\xi) = 0$ for some $\xi \in (0, \infty)$. Moreover, $\Lambda^{(3)}$ is finite on $\{0, \xi\}$.

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

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

The parameter ξ appearing in (H_1) will play an important role in the sequel, since we will use the ξ-shifted measure defined in (2.6) to develop our algorithm. Under the above hypotheses, it is shown in Collamore and Vidyashankar (2011), Section 5, that the following path properties hold:

LEMMA 2.1. *Assume Letac's Model E, and let $\{V_n\}$ denote the forward recursive sequence corresponding to this SFPE. Assume that (H_1), (H_2), and (H_3) are satisfied. Then:*

(i) *$\{V_n\}$ is φ-irreducible and geometrically ergodic. Moreover,*

$$\lim_{n\to\infty} \mathbf{P}\,(V_n > u) = \mathbf{P}\,(V > u)\,.$$

(ii) *Under the measure μ_ξ,*

$$V_n \nearrow +\infty \quad \text{w.p. } 1 \quad \text{as} \quad n \to \infty.$$

Thus, the Markov chain $\{V_n\}$ is transient in the ξ-shifted measure.

We note that the proof of part (i), providing the uniqueness of the solution of the SFPE, is obtained from Letac's principle. The following tail estimate is fundamental for the development of the results of this paper.

THEOREM 2.1. *Assume Letac's Model E, and suppose that (H_1), (H_2), and (H_3) are satisfied. Then*

$$\lim_{u\to\infty} u^{\xi}\mathbf{P}\,(V > u) = C \tag{2.7}$$

for a finite positive constant C.

The limit result (2.7) appears in Goldie (1991), while the explicit identification of the constant C is given in Collamore and Vidyashankar (2011).

The above theorem shows that the tail of V decays at a polynomial rate and hence for large u, $\mathbf{P}(V > u)$ constitutes a rare event probability. Thus standard Monte Carlo methods break down for large u, in the sense that the relative error of the estimator will tend to infinity as the probability in question tends to zero (*cf.* Asmussen and Glynn (2007), Chapter 6). The essential reason is that the Markov chain will frequently fail to reach the high level u. Large deviation theory then suggests that we consider shifted distributions, and based on known techniques used in the classical ruin problem, it is natural to shift using the parameter $\xi > 0$ satisfying $\mathbf{E}\left[A^{\xi}\right] = 1$. As a starting point, we observe that under the measure μ_ξ, the process $\{V_n\}$ will be transient by Lemma 2.1 (ii). Thus under the shifted measure μ_ξ, $\{V_n\}$ will ultimately attain the high level u.

Hence a reasonable first step is to simulate the process from the distribution μ_ξ. To relate $\mathbf{P}(V > u)$ 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 $\{V_n\}$, and define a probability measure γ on $\mathcal{C}$ by setting

$$\gamma(E) = \frac{\pi(E)}{\pi(\mathcal{C})}, \quad \text{for all Borel subsets } E \text{ of } \mathcal{C}. \tag{2.8}$$

Let $K := \inf\{n \geq 1 : V_n \in \mathcal{C}\}$ denote the first return time to $\mathcal{C}$. Then we will establish below that

$$\mathbf{P}(V > u) = \mathbf{E}_\gamma\left[N_u\right], \quad \text{where} \quad N_u := \sum_{i=0}^{K-1} \mathbf{1}_{\{V_n > n\}}. \tag{2.9}$$

This representation formula suggests that we simulate $\{V_n\}$ *over a cycle* emanating from the set $\mathcal{C}$ and then returning to $\mathcal{C}$.

For (2.9) to be useful, it is necessary that we simulate the process over a cycle that terminates. Since $\{V_n\}$ is transient in the ξ-shifted measure, it is natural to simulate in the *original* measure after the rare event $\{V_n > u\}$ has occurred. Thus, we propose to simulate the process under a *dual* change of measure, namely

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

where μ_ξ is defined as in (2.6) and ξ is given as in (H_1). Roughly speaking, the dual measure shifts the distribution of $\log A_n$ on a path of random duration terminating at time T_u, reverting to the original measure thereafter. Let $\mathbf{E}_{\mathfrak{D}}[\cdot]$ denote expectation with respect to the dual measure $(\mathfrak{D})$.

Finally, to relate the simulated quantity in the ξ-shifted measure to the required probability in the original measure, introduce an appropriate weighting factor. In the proof of Theorem 2.2 below, we will show that $\mathbf{E}_{\mathfrak{D}}[\mathcal{E}_u] =$

$\pi(\mathcal{C})\mathbf{E}_{\mathfrak{D}}\left[N_u e^{-\xi S_{T_u}}\mathbf{1}_{\{T_u<K\}}|V_0 \sim \gamma\right]$, where γ is given as in (2.8). Using this identity, it is natural to introduce the importance sampling estimator

$$\mathcal{E}_u = N_u e^{-\xi S_{T_u}}\mathbf{1}_{\{T_u<K\}}. \tag{2.10}$$

Then it will be shown in the proof of Theorem 2.2 that $\pi(\mathcal{C})\mathcal{E}_u$ is an unbiased estimator for $\mathbf{P}(V > u)$. However, since in practice the stationary distribution π and hence γ is seldom known—even if the underlying distribution of (A, B, D) is known—we first run multiple realizations of $\{V_n\}$ according to the known μ and thereby estimate $\pi(C)$ and γ. Let $\hat{\pi}_k(C)$ denote the estimate obtained for $\pi(C)$, and let $\hat{\mathcal{E}}_{u,n}$ denote the estimate obtained upon averaging the realizations of $\mathcal{E}_u$. This yields the estimator $\hat{\pi}_k(\mathcal{C})\hat{\mathcal{E}}_{u,n}$. The above discussion can be formalized as an algorithm as follows:

Algorithm. *Step 1.* Estimate the stationary measure π of the process $\{V_n\}$. Namely, generate i.i.d. realizations $\{V_{n,j} : j = 1, \ldots, k\}$ of V_n and estimate

$$\hat{\pi}_k(\mathcal{C}) = \frac{1}{k}\sum_{j=1}^{k}\mathbf{1}_{\{V_{n,j}\in\mathcal{C}\}}.$$

Set $\hat{\gamma}_k(E) = \hat{\pi}_k(E)/\hat{\pi}_k(\mathcal{C})$, for any Borel set $E \subset \mathcal{C}$, and set $\hat{\gamma}_k(\mathcal{C}^c) = 0$.

Step 2. Generate a new process having the initial state $V_0 \sim \hat{\gamma}_k$. Now, given V_{n-1}, generate V_n using the forward recursion of the SFPE and the shifted measure μ_ξ. If $V_n \in \mathcal{C}$, then stop and set $\mathcal{E}_u = 0$. If $V_n > u$, then go to Step 3. If not, then repeat Step 2.

Step 3. Generate the process $\{V_n\}$ according to its original distribution until the first time the process enters the set $\mathcal{C}$.

Step 4. Compute the importance sampling estimator, namely

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

where N_u denotes the number of exceedances of the process $\{V_n\}$ above level u prior to time K. We notice here that $\mathcal{E}_u = \mathcal{E}_u(k)$, since the initial distribution of the Markov chain depends on k. (We will frequently suppress the dependence on k when there is no cause for confusion.)

Step 5. Repeat Steps 2-4 n times, each time calculating the importance sampling estimator $\mathcal{E}_u$, yielding the jth realization $\mathcal{E}_{u,j}$, $j = 1, \ldots, n$. Then compute the average,

$$\hat{\mathcal{E}}_{u,n} = \hat{\mathcal{E}}_{u,n}(k) = \frac{1}{n}\sum_{j=1}^{n}\mathcal{E}_{u,j}(k).$$

Finally, estimate $\mathbf{P}(V > u)$ by setting this quantity equal to $\hat{\pi}_k(\mathcal{C})\hat{\mathcal{E}}_{u,n}(k)$.

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

2.2. *Consistency and efficiency of the algorithm.* Our first result is concerned with consistency; that is, for any fixed u, the estimator given in (2.10) converges to $\mathbf{P}(V > u)$ with probability one (w.p. 1) as $u \to \infty$.

THEOREM 2.2. *Assume that hypotheses* (H_1), (H_2), *and* (H_3) *hold. Then for any* $\mathcal{C}$ *intersecting* $\operatorname{supp}(\pi)$ *and for any* u *sufficiently large such that* $u \notin \mathcal{C}$, *the algorithm is strongly consistent; that is,*

$$\lim_{k\to\infty} \lim_{n\to\infty} \hat{\pi}_k(\mathcal{C})\hat{\mathcal{E}}_{u,n}(k) = \mathbf{P}(V > u) \quad \text{a.s.} \tag{2.11}$$

REMARK 2.1. If the stationary distribution π of $\{V_n\}$ is known on $\mathcal{C}$ (as would be the case if $\mathcal{C} = \{v\}$ for some point $v \in \mathbb{R}$), then it will follow from the proof that $\pi(\mathcal{C})\hat{\mathcal{E}}_{u,n}$ is an unbiased estimator for $\mathbf{P}(V > u)$.

We emphasize that the choice of the set $\mathcal{C} := [-M, M]$ is not critical for obtaining consistency, and $\mathcal{C}$ can be taken to be any set having a non-empty intersection with the support of the stationary measure π. However, under an appropriate choice of M, we will establish that the estimator is strongly efficient in the sense that it has bounded relative error. Before stating this next result, we introduce a slight modification of hypothesis (H_2).

Hypothesis:
(H_2'): $\mathbf{E}\big[\big(A^{-1}|B|^2\big)^{\alpha}\big] < \infty$ and $\mathbf{E}\big[\big(A|D|^2\big)^{\alpha}\big] < \infty$, for some $\alpha > \xi$.

THEOREM 2.3. *Assume that hypotheses* (H_1), (H_2'), *and* (H_3) *hold and that* $\lambda(\alpha) < \infty$ *for some* $\alpha < -\xi$. *Then there exists an* $M > 0$ *such that*

$$\sup_{u \geq 0} \sup_{k \in \mathbb{Z}_+} u^{2\xi} \mathbf{E}_{\mathfrak{D}} \left[\mathcal{E}_u^2 \middle| V_0 \sim \gamma_k \right] < \infty. \tag{2.12}$$

Moreover, even if $\lambda(\alpha) = \infty$ *for all* $\alpha < -\xi$, *then* (2.12) *still holds provided that* $\mathbf{E}\left[\left(|D| + A^{-1}|B|\right)^{\alpha}\right] < \infty$ *for all* $\alpha > 0$.

Eq. (2.12) implies that our estimator exhibits bounded relative error, that is, the ratio $\mathbf{E}_{\mathfrak{D},\gamma}\left[\mathcal{E}_u^2\right] / \left(\mathbf{E}_{\mathfrak{D},\gamma}\left[\mathcal{E}_u\right]\right)^2$ is uniformly bounded in u and, in particular, remains bounded in the limit as $u \to \infty$. It is worth noticing that *even if* $\lambda(-\xi) < \infty$, the relative error could not be bounded without the presence of the term $\mathbf{1}_{\{T_u<K\}}$ in the definition of $\mathcal{E}_u$. For further discussion of this point and its relationship to perpetuity estimates, see the discussion in Section 3 following the proof of this theorem.

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 $\{V_n\}$, generated in its α-shifted measure, with $\alpha = 0$ and $\alpha = -\xi$. Indeed, we will use this approach in Lemma 3.3 to provide a formula for M. Admittedly, this method may not always yield a good choice for M. For instance, a relatively large value for M could lead to an increase in the initial rejections, thus leading to an increase in the number of Monte Carlo simulations needed for accurate results. Indeed, the choice of M is problem-dependent and we outline an alternative numerical method in Section 4 to address this issue.

2.3. *Running time of the algorithm.* Next we focus on the running time of the proposed algorithm, which will clearly depend on the following three quantities under the dual change of measure: (i) the return time K on $\{K < \infty\}$; (ii) the time T_u needed to reach the level u; and (iii) the time $K - T_u$ to return to the set $\mathcal{C}$, where the latter quantity is evaluated on the set $\{K > T_u\}$. The expected behavior of these three quantities is the subject of our next main theorem. First we need to impose a mild regularity condition.

Hypothesis:
(H_4): $\mathbf{P}_\xi\big(V_1 \le 1 | V_0 = v\big) = \mathbf{o}(v^{-\epsilon})$ as $v \to \infty$, for some $\epsilon > 0$.

THEOREM 2.4. *Assume that hypotheses (H_1)-(H_4) hold. Then*:

(i) *Under the measure $\mu_{\mathfrak{D}}$, the return time K has finite expectation on $\{K < \infty\}$; that is, $\mathbf{E}_{\mathfrak{D}}\left[K \mathbf{1}_{\{K<\infty\}}\right] < \infty$.*

(ii) *Conditional on the event $\{T_u < K\}$, we have under the measure $\mu_{\mathfrak{D}}$ that*

$$\mathbf{E}_{\mathfrak{D}}\left[\frac{T_u}{\log u}\,\middle|\, T_u < K\right] \to \frac{1}{\Lambda'(\xi)} \quad \text{as } u \to \infty.$$

(iii) *Furthermore, conditional on the event $\{T_u < K\}$, we also have that*

$$\mathbf{E}_{\mathfrak{D}}\left[\frac{K - T_u}{\log u}\,\middle|\, T_u < K\right] \to \frac{1}{|\Lambda'(0)|} \quad \text{as } u \to \infty.$$

Note that the ultimate objective of the algorithm is to minimize cost, *i.e.*, the total number of Monte Carlo simulations needed to attain a given accuracy. Thus, it is upon combining Theorem 2.3 with Theorem 2.4 that we see that our algorithm is actually efficient. For further details, see Remark 2.2 below.

Several sufficient conditions for (H_3') can be provided. For the case $B > 0$ and $D = 0$ (which corresponds to the special case of a perpetuity sequence), it is easy to see that (H_3') always holds, since by Chebyshev's inequality,

$$\mathbf{P}_\xi(V_1 \le 1 | V_0 = v) \le \mathbf{P}\left(A \le v^{-1}\right) \le \mathbf{E}_\xi\left[A^{-\epsilon}\right] v^{-\epsilon}, \tag{2.13}$$

and for $\epsilon > 0$ sufficiently small, $\mathbf{E}_\xi\left[A^{-\epsilon}\right] = \lambda(\xi - \epsilon) \in (0, 1)$. When B and D assume values in $\mathbb{R}$, a slight modification of (2.13) yields that (H_4) holds under appropriate moment conditions on B/A and D. In general, (H_4) is a very weak condition which is easily verified in all of the examples considered in this article.

2.4. *Optimality of the algorithm.* We conclude with a comparison of our algorithm to other algorithms obtained using alternative measure transformations. A natural alternative would be to simulate with a 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 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 $q \in \{0, 1\}$. Intuitively, (w, q) denotes a realization of (W_n', Q_n), where

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

Now suppose that we shift the driving sequence $\{Y_n\}$ of the process $\{V_n\}$, *i.e.*, we change the distribution of $Y_n := (A_n, B_n, D_n) \sim \mu$. In a general sense, we would like to allow the simulation distribution of Y_n to *depend* on (W_n, Q_n), *i.e.* on the level of the scaled process $\{\log V_n / \log u\}$ at the prior time $n - 1$, and on whether or not the process $\{V_n\}$ has exceeded the level u by that time. To be more precise, let W_n' be defined as above and set

$$W_n = \begin{cases} W_n', & \text{if } W_n' \in [0, 1], \\ (W_n' \wedge 1) \vee 0 & \text{otherwise.} \end{cases}$$

Next, suppose that $\{\nu_n\}$ is a family of random measures satisfying

$$\nu_n(E) = \nu(E; W_n, Q_n), \quad \text{for all } E \in \mathcal{B}(\mathbb{R}^3), \tag{2.14}$$

for some given measure ν.

Condition (C_0): For any measure ν, where $\mu \ll \nu$,

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

is piecewise continuous as a function of w, for any fixed $Q_n \in \{0,1\}$.

Class $\mathfrak{M}$: Define

$$\mathfrak{M} = \Big\{\{\nu_n\} : \nu_n(\cdot) = \nu(\cdot; W_n, Q_n) \text{ and } \nu \text{ satisfies } (C_0), \text{ for all } n\Big\}. \tag{2.15}$$

Thus, $\mathfrak{M}$ denotes the class of all sequences of measures $\{\nu_n\}$ obtained from the construction (2.14), where (C_0) is satisfied and we also assume that μ is absolutely continuous with respect to $\nu(\cdot; w, q)$ for any $w \in [0,1]$ and $q \in \{0,1\}$. Set $\boldsymbol{\nu} = \{\nu_1, \nu_2, \ldots\}$, and let $\mathbf{E}_{\boldsymbol{\nu}}[\cdot]$ denote expectation with respect to this state-dependent collection of simulation distributions.

¿From a practical perspective, (C_0) is no real restriction on the class of possible simulation distributions. In practice, it is natural to simulate using a class of exponential transforms determined by a multidimensional parameter $\boldsymbol{\beta} = \boldsymbol{\beta}(w,q)$, where for some normalizing factor c,

$$\mu_{\boldsymbol{\beta}}(E) = \frac{1}{c}\int_E e^{\beta_1 x + \beta_2 y + \beta_3 z} d\mu(x,y,z), \quad E \in \mathcal{B}(\mathbb{R}^3),$$

and $\mathfrak{L}(\log A_n, B_n, D_n) \sim \mu_{\boldsymbol{\beta}}$. In other words, 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), where w corresponds to the realization of $\log V_{n-1}/\log u$ while q corresponds to the realization of $\mathbf{1}_{\{T_u<n\}}$. It is reasonable to expect that the optimal $\boldsymbol{\beta}(w,q)$ depends on w, or that it changes in the event that a threshold is achieved, *e.g.*, when the process V_n first exceeds u. Such scenarios will always lead to members $\boldsymbol{\nu} = (\nu_1, \nu_2, \ldots)$ belonging to the class $\mathfrak{M}$ and, in particular, satisfying (C_0).

Now suppose that simulation is performed using a modification of our previous algorithm, where $Y_n \sim \nu_n$ for some $\boldsymbol{\nu} = \{\nu_1, \nu_2, \ldots\} \in \mathfrak{M}$. For the importance sampling estimator, set

$$\mathcal{E}_u^{(\boldsymbol{\nu})} = N_u \left\{\prod_{i=1}^{T_u} \frac{d\mu}{d\nu_i}(Y_i; W_i, Q_i)\right\} \mathbf{1}_{\{T_u<K\}}. \tag{2.16}$$

Let $\hat{\pi}_k$ denote an empirical estimate for π, as obtained in Step 1 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 easily shown by a slight modification of Theorem 2.2 that

$$\lim_{n\to\infty} \hat{\pi}_{k(n)} \hat{\mathcal{E}}_{u,n}^{(\boldsymbol{\nu})} = \mathbf{P}\left(V > u\right), \tag{2.17}$$

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

THEOREM 2.5. *Assume that the conditions of Theorems 2.3 and 2.4 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 for any initial state $v \in \mathcal{C}$,*

$$\liminf_{u\to\infty} \frac{1}{\log u} \log\left(u^{2\xi} \mathbf{E}_\nu \left[\left(\mathcal{E}_u^{(\boldsymbol{\nu})}\right)^2 \middle| V_0 = v \right]\right) \geq 0. \tag{2.18}$$

Moreover, equality holds in this inequality 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 in the class $\mathfrak{M}$.*

REMARK 2.2. In practice, the objective should be to minimize the total number of *random variables* generated in order to obtain a given accuracy. This grows according to

$$\mathrm{Var}\left(\mathcal{E}_u^{(\boldsymbol{\nu})}\right) \left\{ c_1 \mathbf{E}_{\boldsymbol{\nu}} \left[K \,|T_u < K \right] + c_2 \mathbf{E}_{\boldsymbol{\nu}} \left[K \mathbf{1}_{\{T_u \geq K\}} \right] \right\} \quad \text{as } u \to \infty \tag{2.19}$$

for appropriate constants c_1 and c_2; *cf.* Hammersley and Handscomb (1964), Siegmund (1976). However, as a consequence of Theorem 2.4 (ii) and (iii), we have under the dual measure $(\mathfrak{D})$ that

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

for some positive constant Θ, while the second term in (2.19) converges to a finite constant, by Theorem 2.4 (i). Thus, under the dual measure, it is sufficient to minimize the second moment in the asymptotic limit as $u \to \infty$. Consequently we conclude that simulation under the dual measure $(\mathfrak{D})$ is indeed asymptotically efficient and optimal.

3. Proofs of consistency and efficiency.

Set $A_0 \equiv 1$, and define

$$S_n = \sum_{i=1}^n \log A_i, \quad i = 1, 2, \ldots; \quad S_0 = 0;$$

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

where

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

In the following lemma, we summarize some regularity properties satisfied by $\{Z_n\}$ and the corresponding perpetuity sequence $\bar{Z}^{(p)}$.

LEMMA 3.1. *Assume Letac's Model E, and suppose that (H_1), (H_2), and (H_3) are satisfied. Then:*

(i) *$Z_n \to Z$ with respect to the measure μ_ξ, where Z is a proper random variable supported on $(0, \infty)$.*

(ii) *$\mathbf{E}_\xi\big[(\bar{Z}^{(p)})^\xi\big] < \infty$, and for all n and all u,*

$$\big|Z_n \mathbf{1}_{\{n<K\}}\big| \le \bar{Z}^{(p)} \quad \text{and} \quad \big|Z_{T_u}\mathbf{1}_{\{T_u<K\}}\big| \le \bar{Z}^{(p)}. \tag{3.1}$$

Moreover,

$$Z_{T_u}\mathbf{1}_{\{T_u<K\}} \le \sum_{n=0}^{\infty} \frac{\tilde{B}_n}{A_0 \cdots A_n}\mathbf{1}_{\{n \le T_u < K\}}. \tag{3.2}$$

PROOF. These properties through (3.1) are established in Collamore and Vidyashankar (2011), Lemma 5.5. (Although we have replaced the stopping time τ of that article with K in (3.1), the proofs remain unchanged.)

For (3.2), apply Eq. (5.32) in Lemma 5.5 of Collamore and Vidyashankar (2011) to obtain that

$$Z_n \le \sum_{i=0}^{n} \frac{\tilde{B}_i}{A_0 \cdots A_i}, \quad n = 0, 1, \ldots. \tag{3.3}$$

Then (3.2) follows from (3.3). □

A second basic result which will be used throughout this article is a change of measure formula, which allows one to compare an expectation under a particular measure to an expectation under another measure. Frequently we will take this other measure to be the dual measure. In the following lemma assume, quite generally, that ν_n denotes the distribution of $\{(A_n, B_n, D_n)\}$, possibly depending on the past history of $\{V_n\}$, that is on the vector $(V_0, \ldots, V_{n-1})$. To be more precise, let $\mathbf{v}_n = (v_0, \ldots, v_{n-1})$ denote a vector in $\mathbb{R}^n$, and introduce the family of measures

$$\{\nu_n(E; \mathbf{v}_n) : E \in \mathcal{B}(\mathbb{R}^3), \mathbf{v}_n \in \mathbb{R}^n, n \in \mathbb{Z}_+\}.$$

Then our objective is to compare expectations under the original measure to expectations under the driving sequence

$$Y_n := (A_n, B_n, D_n) \sim \nu_n(\cdot; V_0, \ldots, V_{n-1}).$$

Let $\mathbf{E}_{\boldsymbol{\nu}(n)}[\cdot]$ denote expectation with respect to the family $\{\nu_1, \ldots, \nu_n\}$ just described. With a slight abuse of notation, we will surpress the dependence on n in this last expectation. Then we have the following:

LEMMA 3.2. *Assume Letac's Model E, and suppose that $\mu \ll \nu_n$ a.s. for all n. Let $g : \mathbb{R}^\infty \to [0,\infty]$ be a deterministic function, and let g_n denote its projection onto the first $n+1$ coordinates; that is, $g_n(x_0,\ldots,x_n) = g(x_0,\ldots,x_n,0,0,\ldots)$. Then for any $n \in \mathbb{N}$,*

$$\mathbf{E}\left[g_n(V_0,\ldots,V_n)\right] = \mathbf{E}_{\boldsymbol{\nu}}\left[\left(\prod_{i=1}^{n} \frac{d\mu}{d\nu_i}(Y_i)\right) g_n(V_0,\ldots,V_n)\right]. \tag{3.4}$$

PROOF. This result follows by induction. Assume that the result holds for some positive integer n, for any given function g, and consider

$$\begin{aligned} &\mathbf{E}_{\boldsymbol{\nu}}\left[\left(\prod_{i=1}^{n+1} \frac{d\mu}{d\nu_i}(Y_i)\right) g_{n+1}(V_0,\ldots,V_{n+1})\right] \\ &\qquad = \mathbf{E}_{\boldsymbol{\nu}}\left[\left(\prod_{i=1}^{n} \frac{d\mu}{d\nu_i}(Y_i)\right) h_n(V_0,\ldots,V_n)\right], \end{aligned} \tag{3.5}$$

where

$$h_n(V_0,\ldots,V_n) := \mathbf{E}_{\boldsymbol{\nu}}\left[\frac{d\mu}{d\nu_{n+1}}(Y_{n+1}) g_{n+1}(V_0,\ldots,V_{n+1}) \middle| V_0,\ldots,V_n\right].$$

By direct calculation,

$$h_n(V_0,\ldots,V_n) = \mathbf{E}\left[g_{n+1}(V_0,\ldots,V_{n+1})\middle| V_0,\ldots,V_n\right].$$

Now apply the inductive hypothesis to obtain

$$\mathbf{E}_{\boldsymbol{\nu}}\left[\left(\prod_{i=1}^{n} \frac{d\mu}{d\nu_i}(Y_i)\right) h_n(V_0,\ldots,V_n)\right] = \mathbf{E}\Big[h_n(V_0,\ldots,V_n)\Big].$$

Substituting these last two equations into the right-hand side of (3.5) yields (3.4), as required. □

We will be particularly interested in applying the previous lemma to dual measures of the form

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

where $\alpha \in \operatorname{dom}(\Lambda)$. By conditioning on $\{T_u = m, K = n\}$ and summing over all possible values of m and n, we obtain the following:

Corollary 3.1. *Assume the conditions of the previous lemma. Then for any* $\alpha \in \mathrm{dom}\,(\Lambda)$,

$$\mathbf{E}\left[g_K(V_0,\ldots,V_K)\right] = \mathbf{E}_{\mathfrak{D}_\alpha}\left[g_K(V_0,\ldots,V_K)e^{-\alpha S_{T_u}}\left(\lambda(\alpha)\right)^{T_u}\mathbf{1}_{\{T_u<K\}}\right] + \mathbf{E}_{\mathfrak{D}_\alpha}\left[g_K(V_0,\ldots,V_K)e^{-\alpha S_K}\left(\lambda(\alpha)\right)^{K}\mathbf{1}_{\{T_u\geq K\}}\right], \tag{3.6}$$

where $\mathbf{E}_{\mathfrak{D}_\alpha}[\cdot]$ *denotes expectation with respect to the dual measure in* ($\mathfrak{D}_\alpha$) *above.*

For notational convenience, here and in the following, we write $\mathbf{E}_{\mathfrak{D}_\xi}[\cdot]$ as $\mathbf{E}_{\mathfrak{D}}[\cdot]$, *i.e.*, we suppress ξ in the special case that $\alpha = \xi$. Moreover, we use the notation $\mathbf{P}_\alpha(\cdot)$ and $\mathbf{E}_\alpha[\cdot]$ when the random variables under consideration have the distribution μ_α for all n.

We now turn to the proof of Theorem 2.2. The main complications in the proof are that: (i) we work with the return times rather than the regeneration times of the Markov chain; and (ii) we utilize an empirical approximation to the stationary distribution. The consequence of (i) is that the cycles induced by the return times are not i.i.d. We begin with a summary of the main steps of the proof.

Step 1. Letting $K_1, K_2, \ldots$ denote the return times of the process $\{V_n\}$ to $\mathcal{C}$, we show that the stopped process $X_n = V_{K_n}$, $n = 1, 2, \ldots$, is a stationary Markov chain whose stationary distribution is given by $\gamma(E) = \pi(E)/\pi(\mathcal{C})$, for any Borel set $E \subset \mathcal{C}$, where π is the stationary distribution of $\{V_n\}$.

Step 2. We derive an expression for the total number of exceedances N_u above the level u which occur over a cycle emanating from $\mathcal{C}$ and terminating upon its return to $\mathcal{C}$, assuming that the initial distribution satisfies $V_0 \sim \gamma$.

Step 3. We relate the expression obtained in Step 2 to the object of interest, namely $\mathbf{P}(V > u)$.

Step 4. We verify that, if the true stationary distribution γ is replaced with the empirical distribution γ_k, then the resulting estimator is consistent in the asymptotic limit where $\gamma_k \Rightarrow \gamma$.

Proof of Theorem 2.2. *Step 1.* Let $K_1, K_2, \ldots$ denote the successive return times of $\{V_n\}$ to $\mathcal{C}$; that is, $K_0 = 0$ and $K_n = \inf\{i > K_{n-1} : V_i \in \mathcal{C}\}$ for $n = 1, 2, \ldots$. Set

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

Note that $\{X_n\}$ is a Markov chain with state space $\mathcal{C}$. We will now show that $\{X_n\}$ has a stationary distribution given by $\gamma(E) = \pi(E)/\pi(\mathcal{C})$, for all Borel sets $E \subset \mathcal{C}$. First observe that $\{V_n\}$ is φ-irreducible and geometrically ergodic, by Lemma 2.1 (i). It follows by the definitions of irreducibility and ergodicity that $\{X_n\}$ is also φ-irreducible and geometrically ergodic. Thus $\{V_n\}$ and $\{X_n\}$ are both positive chains.

Hence applying the law of large numbers (Meyn and Tweedie (1993), Theorem 17.1.7 with $f(x) = \mathbf{1}_E(x)$, where $\mathbf{1}_E$ is the indicator function on the set E), we obtain that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{V_i \in E\}} = \pi(E) \text{ a.s.,} \quad \text{for all} \quad E \subset \mathcal{C}. \tag{3.7}$$

Next, let $\mathfrak{N}_n := \sum_{i=1}^{n} \mathbf{1}_{\{X_i \in \mathcal{C}\}}$ denote the number of visits of $\{X_n\}$ to $\mathcal{C}$ occurring by time n. Since $\{V_n\}$ is a recurrent Markov chain, it follows that $\mathfrak{N}_n \uparrow \infty$ w.p. 1. Applying (3.7) with $E = \mathcal{C}$, we obtain that $\pi(\mathcal{C}) = \lim_{n\to\infty} \mathfrak{N}_n/n$ a.s. Similarly, applying (3.7) with $E \subset \mathcal{C}$ and utilizing the definition of $\{X_n\}$, we also obtain that

$$\pi(E) = \lim_{n\to\infty} \frac{\mathfrak{N}_n}{n} \left(\frac{1}{\mathfrak{N}_n} \sum_{i=1}^{\mathfrak{N}_n} \mathbf{1}_{\{X_i \in E\}} \right) \quad \text{a.s.}$$

Hence

$$\lim_{n\to\infty} \frac{1}{\mathfrak{N}_n} \sum_{i=1}^{\mathfrak{N}_n} \mathbf{1}_{\{X_i \in E\}} = \frac{\pi(E)}{\pi(\mathcal{C})} \quad \text{a.s.} \tag{3.8}$$

Finally, by yet another application of the law of large numbers for Markov chains, now applied to $\{X_n\}$, we obtain that the left-hand side of (3.8) converges to the stationary distribution of $\{X_n\}$. Thus, it follows that $\gamma(E) := \pi(E)/\pi(\mathcal{C})$ is the stationary distribution of $\{X_n\}$.

Step 2. Let $N_{u,1}, N_{u,2}, \ldots$ denote the number of exceedances above level u which occur over the successive cycles starting from $\mathcal{C}$; that is, $N_{u,n} := \sum_{i=K_{n-1}}^{K_n - 1} \mathbf{1}_{\{V_i > u\}}$. Set $S_0^N = 0$ and $S_n^N = N_{u,1} + \cdots + N_{u,n}$, $n = 1, 2, \ldots$. Then $\{(X_n, S_n^N) : n = 0, 1, \ldots\}$ is a Markov random walk. To view $\{S_n^N\}$ as a sum of the functionals of a Markov chain, first introduce the adjoined Markov chain $\big\{\big(X_n, N_{u,n}\big)\big\}$. Since the distribution of $N_{u,n}$ is determined entirely by X_n, where $\{X_n\}$ has stationary distribution γ, it follows that the stationary distribution of this adjoined chain is given by

$$\tilde{\gamma}(E \times F) = \int_E \mathbf{P}\left(N_{u,1} \in F \middle| X_1 = x\right) d\gamma(x). \tag{3.9}$$

Moreover, as $\{X_n\}$ is a positive Harris chain, it follows that $\big\{\big(X_n, N_{u,n}\big)\big\}$ is also a positive Harris chain. Hence the law of large numbers for Markov chains applies, yielding (after observing $\mathbf{E}_{\tilde{\gamma}}[N_u] = \mathbf{E}_\gamma[N_u]$) that

$$\mathbf{E}_\gamma\big[N_u\big] = \lim_{n\to\infty} \frac{S_n^N}{n} := \lim_{n\to\infty} \frac{1}{n} \sum_{n=0}^{K_n - 1} \mathbf{1}_{\{V_i > u\}} \quad \text{a.s.,} \tag{3.10}$$

provided that the expectation on the left-hand side is finite.

To see that this quantity is finite, recall that $\{V_n\}$ is geometrically recurrent. Then since $\{K_n\}$ denotes the return times of this process to its $\mathcal{C}$-set, we have that $\sup_{v \in \mathcal{C}} \mathbf{E}_v [K_n - K_{n-1}] < \infty$. Since $N_{u,n}$ denotes the number of exceedances above level u occurring over the nth such cycle, we obviously have $N_{u,n} \le (K_n - K_{n-1})$, and thus $\mathbf{E}_\gamma [N_u] < \infty$.

Step 3. Next, we identify the limit on the right-hand side of (3.10). First recall that $\mathfrak{N}_n$ is the number of returns to $\mathcal{C}$ occurring by time n; thus $K_{\mathfrak{N}_n}$ is the last return time occurring in $[0, n]$. Now apply the law of large numbers for Markov chains to obtain that

$$\mathbf{P}(V > u) := \pi(u, \infty) = \lim_{n \to \infty} \frac{1}{n} \left\{ \sum_{i=0}^{K_{\mathfrak{N}_n} - 1} \mathbf{1}_{\{V_i > u\}} + \sum_{i=K_{\mathfrak{N}_n}}^{n} \mathbf{1}_{\{V_i > u\}} \right\}. \tag{3.11}$$

We claim that the last term on the right-hand side is asymptotically negligible. To justify this fact, we apply a Markovian renewal theorem given in Iscoe et al. (1985). To this end, first introduce the augmented chain $\{(V_n, \eta_n)\}$, where $\{\eta_n\}$ is an i.i.d. sequence of Bernoulli random variables, independent of $\{V_n\}$, with $\mathbf{P}(\eta_n = 1) = \delta$. Here δ is the constant appearing in the minorization condition of the Markov chain $\{V_n\}$. Then the event $\{V_n \in \mathcal{C}, \eta_n = 1\}$ is a *regeneration time* (Nummelin (1984), Chapter 4). Let $I(n)$ denote the last regeneration time occurring in the interval $[0, n]$, and let $J(n)$ denote the first regeneration time occurring after this time. Since regeneration only occurs at the return times to $\mathcal{C}$ for which $\eta_i = 1$ (so regeneration is "less frequent" than returns to $\mathcal{C}$), we have that $I(n) \le K_{\mathfrak{N}_n} \le n \le J(n) - 1$. Hence

$$\sum_{i=K_{\mathfrak{N}_n}}^{n} \mathbf{1}_{\{V_i > u\}} \le \sum_{i=I(n)}^{J(n)-1} \mathbf{1}_{\{V_i > u\}}. \tag{3.12}$$

Now apply Lemma 6.2 of Iscoe et al. (1985) to the right-hand side of this equation. Letting τ denote a typical interregeneration time and letting ν denote the measure appearing in the minorization for $\{V_n\}$, then Lemma 6.2 of Iscoe et al. (1985) states that

$$\lim_{n \to \infty} \sum_{i=I(n)}^{J(n)-1} \mathbf{1}_{\{V_i > u\}} = \frac{1}{\mathbf{E}[\tau]} \mathbf{E}\left[\tau \Big(\sum_{i=0}^{n} \mathbf{1}_{\{V_i > u\}} \Big) \Big| V_0 \sim \nu \right]. \tag{3.13}$$

Since $\{V_n\}$ is geometrically ergodic, $\mathbf{E}[\tau] < \infty$. Substituting (3.12) and (3.13) into (3.11), we conclude

$$\mathbf{P}(V > u) = \lim_{n \to \infty} \frac{\mathfrak{N}_n}{n} \left(\frac{1}{\mathfrak{N}_n} \sum_{i=0}^{K_{\mathfrak{N}_n} - 1} \mathbf{1}_{\{V_i > u\}} \right). \tag{3.14}$$

Since $\mathfrak{N}_n/n \to \pi(\mathcal{C})$ as $n \to \infty$, it follows from (3.10) and (3.14) that

$$\mathbf{P}\left(V > u\right) = \pi(\mathcal{C})\mathbf{E}_{\gamma}\big[N_u\big]. \tag{3.15}$$

Next apply Corollary 3.1 to the second term on the right-hand side of (3.15) to obtain that

$$\mathbf{P}\left(V > u\right) = \pi(\mathcal{C})\mathbf{E}_{\mathfrak{D},\gamma}\left[N_u e^{-\xi S_{T_u}} \mathbf{1}_{\{T_u < K\}}\right] := \pi(\mathcal{C})\mathbf{E}_{\mathfrak{D},\gamma}\left[\mathcal{E}_u\right]. \tag{3.16}$$

This shows that *if we could choose* $V_0 \sim \gamma$, that is, if we knew the stationary distribution of $\{V_n\}$ exactly, then $\pi(\mathcal{C})\mathcal{E}_u$ would be an unbiased estimator for $\mathbf{P}\left(V > u\right)$.

Step 4. Now suppose that we approximate γ with the empirical measure γ_k described in Step 1 of the algorithm. Then it remains to show that

$$\begin{aligned} \lim_{k\to\infty} \mathbf{E}_{\mathfrak{D}}\left[N_u e^{-\xi S_{T_u}} \mathbf{1}_{\{T_u < K\}}\middle| V_0 \sim \gamma_k\right] \\ = \mathbf{E}_{\mathfrak{D}}\left[N_u e^{-\xi S_{T_u}} \mathbf{1}_{\{T_u < K\}}\middle| V_0 \sim \gamma\right]. \end{aligned} \tag{3.17}$$

Set

$$H(v) = \mathbf{E}_{\mathfrak{D}}\left[\mathbf{E}_{\mathfrak{D}}\left[N_u \left| \mathfrak{F}_{T_u}\right.\right] e^{-\xi S_{T_u}} \mathbf{1}_{\{T_u < K\}}\middle| V_0 = v\right]. \tag{3.18}$$

We now claim that $H(v)$ is uniformly bounded in $v \in \mathcal{C}$. To establish this claim, first apply Theorem 4.2 of Collamore and Vidyashankar (2011) to obtain that

$$\mathbf{E}_{\mathfrak{D}}\left[N_u \left| \mathfrak{F}_{T_u}\right.\right] \mathbf{1}_{\{T_u < K\}} \le \left(C_1(u)\log\left(\frac{V_{T_u}}{u}\right) + C_2(u)\right)\mathbf{1}_{\{T_u < \tau\}}, \tag{3.19}$$

where $\tau \geq K$ is the first regeneration time, and $C_i(u) \to C_i < \infty$ for $i = 1, 2$. Moreover, by the definition of $\{Z_n\}$,

$$e^{-\xi S_{T_u}} = u^{-\xi}\left(\frac{V_{T_u}}{u}\right)^{-\xi} Z_{T_u}^{\xi}. \tag{3.20}$$

Substituting the last two equations into (3.18) yields

$$|H(v)| \le \Theta \mathbf{E}_{\mathfrak{D}}\left[\left|Z_{T_u}^{\xi} \mathbf{1}_{\{T_u < \tau\}}\right| \middle| V_0 = v\right] \tag{3.21}$$

for some finite constant Θ, independent of u. Next apply Lemma 5.5 (ii) of Collamore and Vidyashankar (2011) and the definition of $\bar{Z}^{(p)}$ to obtain that

$$\mathbf{E}_{\mathfrak{D}}\left[\left|Z_{T_u}^{\xi} \mathbf{1}_{\{T_u < \tau\}}\right| \middle| V_0 = v\right] \le \mathbf{E}_{\mathfrak{D}}\left[\left(\bar{Z}^{(p)}\right)^{\xi}\middle| V_0 = v\right]. \tag{3.22}$$

By Lemma 3.1 (ii), the right-hand side of (3.22) is bounded, uniformly in $v \in \mathcal{C}$. Consequently $H(v)$ is bounded in $v \in \mathcal{C}$. Moreover, the function H is also continuous, as can be seen by observing that for any fixed l, $\{V_1(w), \ldots, V_l(w)\}$ converges pointwise to $\{V_1(v), \ldots, V_l(v)\}$ as the initial state $w \to v$. Summing over all realizations $\{K = l\}$ and using the definition $H(v) = \mathbf{E}_{\mathfrak{D}}\left[N_u e^{-\xi S_{T_u}} \mathbf{1}_{\{T_u < K\}} | V_0 = v\right]$ obtains continuity. Finally, since γ_k and γ are both supported on $\mathcal{C}$, it then follows from the weak convergence $\gamma_k \Rightarrow \gamma$ that

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

which is (3.17). □

Next we turn to the proof of efficiency. Suppose now that $Y_n \sim \mu_\beta$ for all n, where $Y = (A_n, B_n, D_n)$ and $\beta \in \operatorname{dom}(\Lambda)$, and where μ_β is defined according to (2.6). Let

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

and note by the definition of μ_β that

$$\Lambda_\beta(\alpha) = \Lambda(\alpha + \beta) - \Lambda(\beta). \tag{3.23}$$

In the following lemma, we summarize some standard results concerning the return times to the set $\mathcal{C}$ under this β-shifted measure. Letting P denote 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

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

where $\rho \in (0,1)$ and $\mathcal{C}$ is some Borel subset of $\mathbb{R}$. We note that this definition differs slightly from the more standard definition given in Meyn and Tweedie (1993), but will be more convenient for our purposes here.

LEMMA 3.3. *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, in this case, we may take* $\rho = \rho_\beta$ *and* $\mathcal{C} = [-M_\beta, M_\beta]$, *where*

$$\rho_\beta := t \lambda_\beta(\alpha), \quad \text{for some } t \in \left(1, \frac{1}{\lambda_\beta(\alpha)}\right), \tag{3.24}$$

and

$$M_\beta := \begin{cases} \left(\mathbf{E}_\beta\left[\tilde{B}^\alpha\right]\right)^{1/\alpha} \left(\lambda_\beta(\alpha)\,(t-1)\right)^{-1/\alpha}, & \text{if } \alpha \in (0,1), \\ \left(\mathbf{E}_\beta\left[\tilde{B}^\alpha\right]\right)^{1/\alpha} \left((\lambda_\beta(\alpha))^{1/\alpha}\,(t^{1/\alpha}-1)\right)^{-1}, & \text{if } \alpha \geq 1. \end{cases} \tag{3.25}$$

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

$$\sup_{v \in \mathcal{C}} \mathbf{P}_\beta\left(K > n \middle|\, V_0 = v\right) \leq \rho_\beta^n, \quad \text{for all } n \in \mathbb{Z}_+. \tag{3.26}$$

Proof. To verify $(\mathcal{D})$ and identify ρ_β and M_β explicitly, first observe that

$$|V_n| \leq A_n|V_{n-1}| + \tilde{B}_n, \quad \text{for all } n, \tag{3.27}$$

where $\tilde{B}_n := A_n|D_n| + |B_n|$. If $\alpha \geq 1$, then it follows by Minkowskii's inequality that

$$\begin{aligned} \mathbf{E}_\beta\left[|V_1|^\alpha \middle| V_0 = v\right] &\leq \left(\left(\mathbf{E}_\beta\left[A^\alpha\right]\right)^{1/\alpha} v + \left(\mathbf{E}_\beta\left[\tilde{B}^\alpha\right]\right)^{1/\alpha}\right)^\alpha \\ &= \rho_\beta v^\alpha \left(\frac{1}{t^{1/\alpha}} + \frac{\left(\mathbf{E}_\beta\left[\tilde{B}^\alpha\right]\right)^{1/\alpha}}{\rho_\beta^{1/\alpha} v}\right)^\alpha, \quad \text{where} \quad \rho_\beta := t\lambda_\beta(\alpha). \end{aligned} \tag{3.28}$$

Thus $(\mathcal{D})$ holds with $h(v) = |v|^\alpha$ and $\rho_\beta = t\lambda_\beta(\alpha) < 1$ for $t \in \left(1, (\lambda_\beta(\alpha))^{-1}\right)$, provided that the quantity inside the parentheses on the right of (3.28) is less than one for $v \notin \mathcal{C}$. Setting $t^{-1/\alpha} + \left(\mathbf{E}_\beta\left[\tilde{B}^\alpha\right]\right)^{1/\alpha}/(\rho_\beta^{1/\alpha} v) = 1$ and solving for v, we obtain the expression given on the right-hand side of (3.25) for the case $\alpha \geq 1$.

If $\alpha < 1$, then an analogous condition is obtained by employing the deterministic inequality $|x+y|^\alpha \leq |x|^\alpha + |y|^\alpha$ in place of Hölder's inequality.

Once $(\mathcal{D})$ is established, the proof of (3.26) is obtained as in Nummelin and Tuominen (1982) or Nummelin (1984). Namely, observe by $(\mathcal{D})$ and an inductive argument that

$$\mathbf{E}\left[h(V_n)\prod_{i=1}^{n} \mathbf{1}_{\{V_i \notin \mathcal{C}\}} \,\middle|\, V_0\right] \leq \rho_\beta^n h(V_0), \quad n = 1, 2, \ldots. \tag{3.29}$$

Since $h(V_n) \geq h(V_0)$ on $\{V_n \notin \mathcal{C},\ V_0 \in \mathcal{C}\}$, it follows that $\mathbf{P}\left(K > n\right) \leq \rho_\beta^n$, uniformly in $V_0 \in \mathcal{C}$. □

Proof of Theorem 2.3. In the proof, set

$$P_n = A_0 \cdots A_n, \quad n = 0, 1, \ldots,$$

and 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

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

note by definition that $V_n e^{-S_n} := V_n/P_n := Z_n$. Since $V_{T_u} > u$, it follows that $0 \le u e^{-S_{T_u}} \le Z_{T_u}$. Consequently by Lemma 3.1 (ii),

$$u^{2\xi}\mathbf{E}_{\mathfrak{D}}\left[\mathcal{E}_u^2\right] \le \mathbf{E}_{\mathfrak{D}}\left[N_u^2\left(\sum_{n=0}^{\infty}\frac{\tilde{B}_n}{P_n}\right)^{2\xi}\mathbf{1}_{\{n\le T_u<K\}}\right], \tag{3.30}$$

where $\{\tilde{B}_n\}$ is defined as in the discussion prior to Lemma 3.1. If $2\xi \ge 1$, then using Minkowskii's inequality on the right-hand side, we obtain

$$\begin{aligned}\left(u^{2\xi}\mathbf{E}_{\mathfrak{D}}\left[\mathcal{E}_u^2\right]\right)^{1/2\xi} &\le \sum_{n=0}^{\infty}\left(\mathbf{E}_{\mathfrak{D}}\left[N_u^2\left(\frac{\tilde{B}_n}{P_n}\right)^{2\xi}\mathbf{1}_{\{n\le T_u<K\}}\right]\right)^{1/2\xi} \\ &= \sum_{n=0}^{\infty}\left(\mathbf{E}\left[N_u^2 P_n^{-\xi}\tilde{B}_n^{2\xi}\mathbf{1}_{\{n\le T_u<K\}}\right]\right)^{1/2\xi},\end{aligned} \tag{3.31}$$

where in the last step we have used the change-of-measure formula in Corollary 3.1. Now apply Hölder's inequality to the right-hand side. Using the independence of $(A_n, \tilde{B}_n)$ from $\mathbf{1}_{\{n-1<T_u\wedge K\}}$, we obtain that the left-hand side of (3.31) is bounded above by

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

where $r^{-1} + s^{-1} = 1$. Now set $\zeta = s\xi$ for the remainder of the proof. Next, by Lemma 3.2 (*cf.* Corollary 3.1), the last term on the right-hand side of the previous equation may be expressed as

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

On the right-hand side, we have used the fact that for all $n < T_u$, the dual measure $(\mathfrak{D}_{-\zeta})$ agrees with the $(-\zeta)$-shifted measure. Substituting the last equation into the upper bound for (3.31), we conclude that

$$\left(u^{2\xi}\mathbf{E}_{\mathfrak{D}}\left[\mathcal{E}_u^2\right]\right)^{1/2\xi} \le \sum_{n=0}^{\infty}\mathcal{J}_n\left\{\left(\lambda(-\zeta)\right)^{n-1}\mathbf{P}_{-\zeta}\left(n-1<T_u\wedge K\right)\right\}^{1/2\zeta}, \tag{3.32}$$

where

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

Since $N_u \le K$, Lemma 3.3 (with $\beta = 0$) yields

$$\sup_{v\in\mathcal{C}} \mathbf{E}\left[N_u^{2r}\,|V_0 = v\right] < \infty, \text{ for any finite constant } r. \tag{3.33}$$

Moreover, for sufficiently small $s > 1$ and $\zeta = s\xi$, it follows by (H_2') that $\mathbf{E}\left[\left(A^{-1}\tilde{B}^2\right)^{\zeta}\right] < \infty$. Thus, to show that the quantity on the left-hand side of (3.32) is finite, it is sufficient to show that for some $\zeta > \xi$ and some $t > 1$,

$$\mathbf{P}_{-\zeta}\left(n - 1 < T_u \wedge K\right) \le \left(t\lambda(-\zeta)\right)^{-n+1}, \text{ for all } n \ge \text{some } N_0, \tag{3.34}$$

and that this last equation holds uniformly in u and uniformly in $v \in \mathcal{C}$. Note that $\{T_u \wedge K > n-1\} \subset \{K > n-1\}$, and by Lemma 3.3,

$$\sup_{v\in\mathcal{C}} \mathbf{P}_{-\zeta}\left(K > n-1|\, V_0 = v\right) \le \left(t\lambda(-\zeta)\right)^{-n+1}, \tag{3.35}$$

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 $M > M_0$, so that (3.33) will hold.) This proves (3.34), thus establishing the theorem for the case $2\xi \ge 1$.

If $2\xi < 1$ then, in place of Minkowskii's inequality, we use the deterministic inequality $|x+y|^\alpha \le |x|^\alpha + |y|^\alpha$, for $\alpha \in (0, 1]$. This yields (3.31), but without the powers "$1/2\xi$" on the left- and right-hand sides. Once this modification has been made, the previous argument can be repeated to obtain that $u^{2\xi}\mathbf{E}_{\mathfrak{D}}\left[\mathcal{E}_u^2\right]$ is bounded uniformly in u and $v \in \mathcal{C}$, completing the proof.

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

We modify the previous argument, now employing a truncation argument as follows. First assume $2\xi \ge 1$. Then as before, we obtain that $\left(u^{2\xi}\mathbf{E}_{\mathfrak{D}}\left[\mathcal{E}_u^2\right]\right)^{1/2\xi}$ is bounded above by the right-hand side of (3.31), and it is sufficient to show that $\mathbf{E}\left[P_{n-1}^{-\zeta}\mathbf{1}_{\{n-1<T_u\wedge K\}}\right] < \infty$ for some $\zeta > \xi$. Set $W_n = P_{n-1}^{-\zeta}\mathbf{1}_{\{n-1<T_u\wedge K\}}$, and first observe that $\mathbf{E}\left[W_n\right] < \infty$. Indeed, using the inequality (3.27), namely

$$|V_n| \le A_n|V_{n-1}|\left(1 + \frac{\tilde{B}_n}{A_n|V_{n-1}|}\right), \quad n = 1, 2, \ldots,$$

and observing that $n - 1 < T_u \wedge K \Longrightarrow |V_i| \in (M, u)$ for $i = 1, \ldots, n-1$, we obtain as a rough upper bound that

$$A_i^{-\zeta} \le \left(\frac{u}{M}\right)^{\zeta}\left(1 + \frac{\tilde{B}_i}{MA_i}\right)^{\zeta}, \; i = 1, \ldots, n-1 \text{ on } \{n-1 < T_u \wedge K\}. \tag{3.36}$$

Then this equation yields an upper bound for P_{n-1}. Using the assumption that $\mathbf{E}\left[(A^{-1}\tilde{B})^{\alpha}\right] < \infty$ for all $\alpha > 0$, we conclude by (3.36) that $\mathbf{E}\left[W_n\right] < \infty$.

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

$$\mathbf{E}\left[W_k \mathbf{1}_{F_k^c}\right] \leq \frac{1}{k^2}, \quad k = 1, 2, \ldots. \tag{3.37}$$

Then it remains to show that

$$\sum_{k=0}^{\infty} \mathbf{E}\left[W_k \mathbf{1}_{F_k}\right] < \infty. \tag{3.38}$$

To verify that (3.38) holds, set $\bar{A}_{0,k} = 1$ and

$$\bar{A}_{n,k} = A_n \mathbf{1}_{\{A_n \geq L_k\}} + L_k \mathbf{1}_{\{A_n < L_k\}}, \quad n = 1, 2, \ldots.$$

Let $\lambda_k(\alpha) = \mathbf{E}\big[\bar{A}_{1,k}^{\alpha}\big]$, and define $\bar{W}_k = (\bar{A}_0 \cdots \bar{A}_{k-1})^{-\zeta}\mathbf{1}_{\{k-1<T_u \wedge K\}}$. Since $\{k-1 < T_u \wedge K\} \subset \{k-1 < K\}$, it follows by Lemma 3.2 (similarly to Corollary 3.1) that

$$\mathbf{E}\left[\bar{W}_k\right] \leq (\lambda_k(-\zeta))^{k-1}\mathbf{E}_{-\zeta}\left[\mathbf{1}_{\{K>k-1\}}\mathbf{1}_{F_k}\right]. \tag{3.39}$$

To evaluate the expectation on the right-hand side, we apply a direct argument to obtain a drift condition (rather than Lemma 3.3, as we used before). In particular, start with the forward recursion

$$|V_{n,k}| \leq \bar{A}_{n,k}|V_{n-1,k}|\left(1 + \frac{\tilde{B}_n}{\bar{A}_{n,k}|V_{n-1,k}|}\right), \quad n = 1, 2, \ldots. \tag{3.40}$$

Write $\mathbf{E}_{-\zeta,w}\left[\cdot\right] = \mathbf{E}_{-\zeta}\left[\cdot|V_{0,k} = w\right]$. Now let $\beta > 0$ and take the βth moment in (3.40) for the one-step transition which starts at $V_{0,k} = w$. To simplify the expression on the right-hand side, use the definition (2.6) of $\mu_{-\zeta}$ (but applied to the truncated random variable $\bar{A}_{1,k}$) to express this expectation in terms of the original measure. This yields

$$\mathbf{E}_{-\zeta,w}\left[|V_{1,k}|^{\beta}\right] \leq \frac{w^{\beta}}{\lambda_k(-\zeta)}\mathbf{E}\left[(\bar{A}_{1,k})^{\beta-\zeta}\left(1 + \frac{\tilde{B}_1}{w\bar{A}_{1,k}}\right)^{\beta}\right]. \tag{3.41}$$

Hence by Hölder's inequality,

$$\mathbf{E}_{-\zeta,w}\left[|V_{1,k}|^{\beta}\right] \leq \rho_k w^{\beta}\left(t^{-q}\mathbf{E}\left[\left(1 + \frac{\tilde{B}_1}{w\bar{A}_{1,k}}\right)^{q\beta}\right]\right)^{1/q}, \tag{3.42}$$

where $\rho_k := \left(\mathbf{E}\left[\left(\bar{A}_{1,k}\right)^{p(\beta-\zeta)}\right]\right)^{1/p}\left(t/\lambda_k(-\zeta)\right)$ 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 has been chosen sufficiently close to one so that the expectation in the definition of ρ_k is finite. Since $L_k \downarrow 0$ as $k \to \infty$,

$$\lim_{k\to\infty} \lambda_k(\alpha) = \lambda(\alpha), \quad \alpha \geq 0, \tag{3.43}$$

by a monotone convergence argument. Moreover, (H_1) implies that λ achieves its minimum on the positive axis, at which point $\lambda(\hat{\beta}) < 1$. Thus we conclude that for $t \in \left(1, \left(\lambda(\hat{\beta})\right)^{-1/p}\right)$ and for some constant $\rho \in (0,1)$,

$$\lim_{k\to\infty} \lambda_k(-\zeta)\rho_k = t\left(\lambda(\hat{\beta})\right)^{1/p} < \rho. \tag{3.44}$$

Then $\lambda_k(-\zeta)\rho_k \leq \rho$ for all $k \geq k_0$, and with this value of ρ, we obtain by (3.42) that

$$\mathbf{E}_{-\zeta,w}\left[|V_{1,k}|^\beta\right] \leq \frac{\rho w^\beta}{\lambda_k(-\zeta)}, \quad \text{for all } k \geq k_0, \tag{3.45}$$

provided that

$$t^{-q}\mathbf{E}\left[\left(1+\frac{\tilde{B}_1}{w\bar{A}_{1,k}}\right)^{q\beta}\right] \leq 1. \tag{3.46}$$

Our next objective is to find a set $\mathcal{C} = [-M, M]$ such that for all $w \notin \mathcal{C}$, (3.46) holds. First assume $q\beta \geq 1$ and apply Minkowskii's inequality to the left-hand side of (3.46). Then set this quantity equal to one, solve for w, and set $w = M_k$. After some algebra, this yields

$$M_k = \frac{1}{t^{1/\beta}-1}\left(\mathbf{E}\left[\left(\frac{\tilde{B}_1}{\bar{A}_{1,k}}\right)^{q\beta}\right]\right)^{1/q\beta}. \tag{3.47}$$

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

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

To complete the proof observe that, as in the proof of Lemma 3.3 (*cf.* (3.29)), we obtain upon iterating (3.45) with $\mathcal{C} = [-M, M]$ that

$$\mathbf{E}_{-\zeta}\left[\mathbf{1}_{\{K>k-1\}}\mathbf{1}_{F_k}\right] \leq \left(\frac{\rho}{\lambda(-\zeta)}\right)^{-k+1}, \quad \text{for all } k \geq k_0. \tag{3.48}$$

Note that on the set F_k, the sequences $\{V_{n,k} : 1 \leq n \leq k\}$ and $\{V_n : 1 \leq n \leq k\}$ agree, and thus the event $\{K > k-1\}$ coincides for these two sequences. Finally, substituting (3.48) into (3.39) yields (3.38), as required. If $2\xi < 1$, then the above argument may be modified in the same manner as for Case 1, and the required estimate still holds. $\square$

Returning to the remark concerning the assumption that $\lambda(\alpha) < \infty$ for some $\alpha < -\xi$, we notice that $\mathbf{E}_{\mathfrak{D}}\Big[\Big(P_n^{-1}\tilde{B}_n\Big)^{2\xi} \Big] = \infty$ (see Alsmeyer et al. (2009)). Thus, even though the expectation of the perpetuity sequence is infinite, it converges when studied on the set $\{T_u < K\}$. We emphasize here that some of the above calculations could be made much simpler if we were to work with assumptions such as $B > 0$ and $D > 0$. However, these are quite strong assumptions from an applied perspective.

4. Examples and simulations. In this section we provide several examples which illustrate how to implement the algorithm. As seen in Sections 2 and 3, while the use of a drift condition provides a formula for M, it may not be optimal in a practical sense. This is due to the fact that the estimate for V_n^α typically uses Minkowskii- or Hölder-type inequalities, which are usually not very sharp. We begin by outlining an alternative method for obtaining M and use it to verify that it yields meaningful answers from a practical perspective.

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

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

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

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

In this expression, α is necessarily 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 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 $\{(A_i, B_i, D_i) : 1 \leq i \leq N\}$ denote a collection of i.i.d random variables having the same distribution as that of (A, B, D). The numerical method for finding an optimal choice of M proceeds as follows:

Step 1. Using a root finding algorithm, obtain ξ. That is, solve for ξ in the equation $\mathbf{E}[A^{\xi}] = 1$. If an analytic expression for $\mathbf{E}[A^{\xi}]$ is not available, numerical approximations such as Gauss-Hermite quadrature methods can be applied.

Step 2. For $\mathbf{E}_{\beta}\left[A^{\alpha}\right] < 1$, use a Monte Carlo procedure to compute $\mathbf{E}_{\beta}\left[|V_1|^{\alpha} | V_0 = v\right]$ 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 $\beta = 0$ and $\beta = -\xi$ and where $\rho_{\beta} < 1$. Then select α so that it provides the smallest possible value of v. Set $M_{\beta} > v$ for $\beta = 0$ and $\beta = -\xi$.

Step 3. Set $M = \max(M_0, M_{-\xi})$.

We now turn to some specific examples.

4.2. *The ruin problem with stochastic investments.* Assume that the fluctuations in the insurance business are 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 premiums income rate, $\{\zeta_n\}$ its claims losses, and N_t the number of Poisson claims occurring in the interval $[0, t]$. It is assumed that $\{\zeta_n\}$ is i.i.d. and independent of $\{N_t\}$. We now depart from this classical model by assuming that at the discrete times $n = 1, 2, \ldots$, the surplus capital is invested, earning stochastic returns $\{R_n\}$ which are assumed to be i.i.d. Let $L_n := -\left(X_n - X_{n-1}\right)$ 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

$$Y_n = R_n Y_{n-1} - L_n, \quad n = 1, 2, \ldots, \quad Y_0 = u. \tag{4.2}$$

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

Our objective is to determine the probability of ruin using importance sampling, namely to estimate

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

Iterating (4.2) yields $Y_n = \left(R_1 R_2 \cdots R_n\right)\left(Y_0 - \mathcal{L}_n\right)$, where

$$\mathcal{L}_n := \frac{L_1}{R_1} + \cdots + \frac{L_n}{R_1 R_2 \cdots R_n}, \tag{4.4}$$

and hence $\Psi(u) = \mathbf{P}\left(\mathcal{L}_n > u, \text{ some } n\right)$. Setting $\mathcal{L} = \left(\sup_{n\in\mathbb{Z}_+} \mathcal{L}_n\right) \vee 0$, then from (4.4) and an elementary argument we obtain that $\mathcal{L}$ satisfies the SFPE

$$\mathcal{L} \stackrel{\mathcal{D}}{=} (A\mathcal{L} + B)^+ , \quad \text{where} \quad A := \frac{1}{R_1} \text{ and } B = \frac{L_1}{R_1}. \tag{4.5}$$

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

4.2.1. *Implementing the algorithm.* To implement the algorithm in this example, we generated investment returns according to the Black-Scholes model. Specifically, we chose

$$A_n = \exp\left\{-\left(\mu - \frac{\sigma^2}{2}\right) - \sigma Z_n\right\}, \quad \text{for all } n, \tag{4.6}$$

where $\{Z_n\}$ is an i.i.d. sequence of standard Gaussian random variables. Then

$$\Lambda(\alpha) = -\alpha\left(\mu - \frac{\sigma^2}{2}\right) + \frac{\alpha^2\sigma^2}{2}. \tag{4.7}$$

Thus $\xi = 2\mu/\sigma^2 - 1$ and $\mu_\xi \sim \text{Normal}(\mu - \sigma^2/2, \sigma^2)$.

We set $\mu = 0.2$ and $\sigma^2 = 0.25$. Regarding the insurance model, we set the premiums rate $c = 1$, $\{\zeta_n\}$ to be exponential with parameter 1, and $\{N_t\}$ to be a Poisson process with parameter 1/2. Applying the procedure described at the beginning of this section to this model, we obtained that $M_0 = 0 = M_{-\xi}$. Thus $M = 0$.

In this example, we can actually deduce that $M = 0$ by a more elementary argument. Arguing as in the proof of Lemma 3.3, we obtain that $M_\beta = \min_{i=1,2} M_\beta^{(i)}$, where

$$M_\beta^{(1)} = \inf_{\alpha\in(0,1)\cap\Phi} \frac{||B_1^+||_{\beta,\alpha}}{(1 - ||A_1||_{\beta,\alpha}^\alpha)^{1/\alpha}}, \quad M_\beta^{(2)} = \inf_{\alpha\in[1,\infty)\cap\Phi} \frac{||B_1^+||_{\beta,\alpha}}{1 - ||A_1||_{\beta,\alpha}}, \tag{4.8}$$

and $\Phi = \{\alpha \in \mathbb{R} : \mathbf{E}_\beta\left[A^\alpha\right] < 1\}$. Here, $||\cdot||_{\beta,\alpha}$ denotes the L_α norm under the β-shifted measure μ_β, where we again consider the two cases $\beta = 0$ and $\beta = -\xi$. For each of these cases, this infimum may be computed numerically, yielding $M_0 = 0 = M_{-\xi}$ and thus $M = 0$, just as before.

In choosing $M = 0$, the algorithm in Section 2 simplifies considerably, since in this case we can take the measure γ to be a point mass at the origin. Moreover, it can be observed that $\{0\}$ is an atom of the the Markov chain $\{V_n\}$. Consequently, it follows that a cycle originating at $\{0\}$ and then stopped upon its return to $\{0\}$ forms a *regeneration cycle* of the Markov chain (*cf.* Nummelin (1984), Chapter 4).

We implemented the algorithm in Section 2 to estimate the probabilities of ruin for $u = 10,\ 20,\ 100,\ 500,\ 10^3,\ 5 \times 10^3,\ 10^4,\ 5 \times 10^4$, and 10^5. In all our simulations, the distribution in Step 1 was based on $k = 10^4$, and V_{1000} was taken as an approximation to the limit random variable V. We arrived at this choice using extensive exploratory analysis comparing V_{1000} and V_n. The comparisons involved studying the empirical cumulative distribution functions and two-sample comparisons using Kolmogrov-Smirnov tests between V_{1000} and other values of V_n. Specifically, based on 10^5 samples and two-sample Kolmogrov-Smirnov tests, there were no statistically significant differences between V_{1000} and V_{2000}, V_{5000}, and $V_{10,000}$ (with p-values ≥ 0.185).

Table 4.1 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 example 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 of tail decay in Theorem 2.1 and the relative error (RE). Even in the extreme tail—far below the probabilities of practical interest in this problem—the algorithm works effectively and is clearly seen to have bounded relative error. For comparison, we also present here 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 estimate and the crude Monte Carlo estimates are close, *which provides an empirical validation of the algorithm for small values of u.* Admittedly, the value of the crude estimate for $u = 10^5$ is questionable. Finally, since all of the conditions of Theorems 2.3 and 2.5 are satisfied, our algorithm is optimal and efficient, as can also be seen from the values of the relative error.

Table 4.1. Importance sampling estimation for the ruin probability with investments obtained using $M = 0$.

u	$\mathbf{P}(V > u)$	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
2.0e+01	3.66e-02	3.52e-02	3.81e-02	2.21e-01	1.98e+01	3.54e-02
1.0e+02	1.33e-02	1.28e-02	1.39e-02	2.11e-01	2.12e+01	1.29e-02
5.0e+02	4.95e-03	4.74e-03	5.15e-03	2.06e-01	2.09e+01	4.85e-03
1.0e+03	3.27e-03	3.14e-03	3.41e-03	2.07e-01	2.12e+01	3.21e-03
5.0e+03	1.25e-03	1.19e-03	1.30e-03	2.06e-01	2.18e+01	1.23e-03
1.0e+04	8.13e-04	7.78e-04	8.49e-04	2.04e-01	2.24e+01	8.01e-04
5.0e+04	3.06e-04	2.93e-04	3.20e-04	2.02e-01	2.22e+01	3.27e-04
1.0e+05	1.98e-04	1.90e-04	2.07e-04	1.98e-01	2.16e+01	2.10e-04

4.3. *Perpetuity sequences.* A similar mathematical problem arises when one estimates the tail of a perpetuity sequence $\{\mathcal{L}_n\}$ as defined in (4.4), namely

$$\mathcal{L}^* := B_0 + A_0 B_1 + A_0 A_1 B_2 + \cdots . \tag{4.9}$$

Note that by choosing $A_i = R_i^{-1}$ and $L_i = B_i$ (and $B_0 \equiv 0$), we are back in the setting of the previous example (except that we now consider $\lim_{n\to\infty} \mathcal{L}_n$ rather than $\mathcal{L} := \left(\sup_{n\in\mathbb{Z}_+} \mathcal{L}_n\right) \vee 0$, which we considered before). This leads to the SFPE

$$\mathcal{L}^* = A\mathcal{L}^* + B. \tag{4.10}$$

If $B \geq 0$, then this is a special case of (4.5).

The sequence (4.9) is of importance in life insurance mathematics. In that context, $\{A_n\}$ denotes the discounted financial returns due, *e.g.*, to inflation, while $\{B_n\}$ denotes the future financial obligations of the company, generally taken to be positive. Then it is of interest to study $\mathbf{P}\left(\mathcal{L}^* > u\right)$ as $u \to \infty$, *i.e.*, the *stationary* tail on this sequence.

Note that (4.9) can be viewed a *backward* sequence generated by the SFPE (4.10), while we simulate the *forward* sequence generated by the same SFPE. These two limiting distributions must be the same due to Letac's principle (*cf.* Letac (1986), Collamore and Vidyashankar (2011), Lemma 2.1).

4.4. *The ARCH(1) process.* We now estimate the tail probabilities of the ARCH(1) financial process, which were also studied in Blanchet et al. (2011). Originally introduced by Engle (1982), this process models the squared returns on an asset using the recurrence equation

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

where $A_n = b\zeta_n^2$ and $B_n = a\zeta_n^2$ and $\{\zeta_n\}$ is an i.i.d. Gaussian sequence. These assumptions imply $\mathbf{E}\left[\log A\right] < \infty$, and the additional assumptions of our theorems are easily seen to be satisfied. Setting $V_n = R_n^2$, we obtain that $V := \lim_{n\to\infty} V_n$ satisfies the SFPE $V \stackrel{\mathcal{D}}{=} AV + B$, which is the same as (4.10). Again, it is of interest to determine the tail behavior of $\{V_n\}$ under stationarity, that is, $\mathbf{P}\left(V > u\right)$ for large values of u. Next we describe how our algorithm can be implemented to estimate these probabilities.

4.4.1. *Implementation.* We set $b = 4/5$ and considered the following values for a: $1.9 \times 10^{-5}, 1$, and 2. It can be shown that

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

We solved the equation $\mathbf{E}\big[A_n^{\xi}\big] = 1$ 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 formulae in (4.8) for M, we obtained (upon taking the limit as $\beta \to 0$ and using the Taylor approximation $\Gamma(\beta + 1/2) = \Gamma(1/2) + \beta\Gamma'(1/2) + \mathbf{O}(\beta^2)$) that $M_0 = 0.362, 0.724, 6.879 \times 10^{-6}$ when $a = 1, 2,$ and 1.9×10^{-5}, respectively. Next we observed that $M_{-\xi}$ does not exist, since $\lambda(-\xi) = \infty$ in this example. However, using the numerical procedure at the beginning of this section, we deduced that, in fact, $M_{-\xi} = 0$. Thus we set $M = M_0$. We also computed M_0 using the algorithm described at the beginning of this section and obtained similar estimates.

Table 4.2 summarizes the simulation results for 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 in the paper of Blanchet, Lam, and Zwart (2011) *show that the relative error based on their algorithm increases as the parameter* $u \to \infty$.

4.5. *The GARCH(1,1) process.* A variant of the last example is the so-called GARCH(1,1) financial process introduced by Bollerslev (1986). Here the logarithmic returns on an asset are modeled as $R_n = \sigma_n \zeta_n$, where σ_n denotes the stochastic volatility at time n and $\{\zeta_n\}$ is i.i.d. Gaussian. It is assumed that the squared volatility satisfies the recurrence equation

$$\sigma_n^2 = a_0 + b_1 \sigma_{n-1}^2 + a_1 R_{n-1}^2, \quad n = 1, 2, \ldots. \tag{4.11}$$

Then $V := \lim_{n\to\infty} \sigma_n^2$ satisfies the SFPE

$$V \stackrel{\mathcal{D}}{=} AV + B, \quad \text{where } A = \big(b_1 + a_1 \zeta_1^2\big) \text{ and } B = a_0. \tag{4.12}$$

To study the large exceedances of $\{R_n\}$ under stationarity, it is of primary interest to determine $\mathbf{P}(V > u)$ as $u \to \infty$. Let $\hat{\mu}$ denote the distribution of A. Then it is easy to see that

$$\hat{\mu}((-\infty, x]) = \begin{cases} 0 & \text{for } x \le b_1, \\ 1 - 2\Phi\left(\sqrt{\frac{x - b_1}{a_1}}\right) & \text{for } x > b_1, \end{cases}$$

where $\Phi(\cdot)$ is the distribution function of a standard Gaussian random variable. Hence, under the ξ-shifted measure, the random variable A has the probability law

$$\hat{\mu}_\xi(E) := \int_E y^\xi d\mu(y), \quad E \in \mathcal{B}(\mathbb{R}).$$

Table 4.2. Importance sampling estimation for the tail probability of ARCH(1) financial process with $a = 1, 2, 1.9 \times 10^{-5}$.

u	$\mathbf{P}(V > u)$	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
2.0e+01	3.43e-02	3.35e-02	3.51e-02	1.92e+00	1.18e+01	3.43e-02
1.0e+02	4.34e-03	4.23e-03	4.45e-03	2.11e+00	1.29e+01	4.28e-03
5.0e+02	5.07e-04	4.96e-04	5.18e-04	2.15e+00	1.13e+01	5.21e-04
1.0e+03	2.04e-04	1.99e-04	2.09e-04	2.20e+00	1.28e+01	2.07e-04
5.0e+03	2.32e-05	2.28e-05	2.36e-05	2.17e+00	8.08e+00	2.00e-05
1.0e+04	9.00e-06	8.88e-06	9.12e-06	2.14e+00	6.83e+00	9.00e-06
5.0e+04	1.07e-06	1.05e-06	1.10e-06	2.21e+00	1.27e+01	2.00e-06
1.0e+05	4.11e-07	4.04e-07	4.18e-07	2.15e+00	8.51e+00	NA
			$a = 2$			
1.0e+01	1.62e-01	1.60e-01	1.64e-01	3.57e+00	5.99e+00	1.62e-01
2.0e+01	7.73e-02	7.64e-02	7.83e-02	4.33e+00	6.21e+00	7.78e-02
1.0e+02	1.08e-02	1.05e-02	1.11e-02	5.25e+00	1.34e+01	1.06e-02
5.0e+02	1.28e-03	1.25e-03	1.31e-03	5.43e+00	1.14e+01	1.33e-03
1.0e+03	5.07e-04	4.96e-04	5.18e-04	5.45e+00	1.13e+01	5.44e-04
5.0e+03	5.96e-05	5.81e-05	6.11e-05	5.57e+00	1.27e+01	7.70e-05
1.0e+04	2.32e-05	2.28e-05	2.36e-05	5.51e+00	8.08e+00	3.50e-05
5.0e+04	2.64e-06	2.60e-06	2.68e-06	5.44e+00	7.60e+00	3.00e-06
1.0e+05	1.07e-06	1.05e-06	1.10e-06	5.61e+00	1.27e+01	1.00e-06
			$a = 1.9 \times 10^{-5}$			
1.0e+01	4.45e-08	4.38e-08	4.52e-08	9.82e-07	8.38e+00	NA
2.0e+01	1.75e-08	1.72e-08	1.78e-08	9.80e-07	1.00e+01	NA
1.0e+02	2.02e-09	1.98e-09	2.05e-09	9.82e-07	9.29e+00	NA
5.0e+02	2.66e-10	1.99e-10	3.32e-10	1.13e-06	1.27e+02	NA
1.0e+03	9.59e-11	8.77e-11	1.04e-10	1.03e-06	4.38e+01	NA
5.0e+03	1.04e-11	1.02e-11	1.06e-11	9.75e-07	1.01e+01	NA
1.0e+04	4.15e-12	4.05e-12	4.26e-12	9.85e-07	1.32e+01	NA
5.0e+04	4.78e-13	4.66e-13	4.91e-13	9.86e-07	1.34e+01	NA
1.0e+05	1.91e-13	1.83e-13	1.99e-13	1.00e-06	2.19e+01	NA

Unlike in the ARCH(1) example, there is no closed form expression for $\hat{\mu}_\xi$, so generating data from this distribution is more difficult. However, generating data from the shifted empirical distribution is easy, since it amounts to generating data from a multinomial distribution. Let $\hat{\mu}_{\beta,k}$ denote the empirical distribution based on a sample of k observations from μ_β for $\beta = 0$ and $\beta = \xi$; thus,

$$\hat{\mu}_{\xi,k}(E) := \frac{\int_E y^\xi d\mu_{0,k}(y)}{\int_{\mathbb{R}} y^\xi d\mu_{0,k}(y)}, \quad E \in \mathcal{B}(\mathbb{R}).$$

In our simulations, we chose $k = 10^6$ and $a_0 = 10^{-7}$, $a_1 = 0.11$, and $b_1 = 0.88$. Also, solving the equation $\mathbf{E}[A^\xi] = 1$ numerically, we obtained $\xi =$

1.838. Furthermore, the minimum of $||A||_\alpha$ is 0.98. Hence $M_0 = 50a_0$. Under the $(-\xi)$-shifted measure, the minimum of $||A||_\alpha$ is 0 and thus $M_{-\xi} = a_0$. Therefore $M = 50a_0$. The same value of M was obtained using the procedure described at the beginning of this section.

Results for the estimation of the tail probabilities are summarized in Table 4.3. As in the previous examples, the algorithm works effectively and is clearly seen to have bounded relative error. In addition, we considered the value $a_0 = 0.01$. In all of these simulation results, we see once again that the importance sampling estimate agrees very well with the crude Monte Carlo estimate for small values of u (where it is sensible to make this comparison).

Table 4.3. Importance sampling Estimation for the tail probability of GARCH(1,1) financial process.

u	$\mathbf{P}(V > u)$	LCL	UCL	C	R.E.
1.0e+01	3.61e-12	3.42e-12	3.80e-12	2.49e-10	2.67e+01
2.0e+01	1.03e-12	9.89e-13	1.07e-12	2.54e-10	1.95e+01
1.0e+02	5.24e-14	5.05e-14	5.43e-14	2.49e-10	1.88e+01
5.0e+02	2.58e-15	2.52e-15	2.64e-15	2.36e-10	1.17e+01
1.0e+03	7.75e-16	7.38e-16	8.11e-16	2.53e-10	2.40e+01
5.0e+03	3.96e-17	3.85e-17	4.06e-17	2.49e-10	1.37e+01
1.0e+04	1.09e-17	1.07e-17	1.12e-17	2.46e-10	1.28e+01
5.0e+04	5.78e-19	5.59e-19	5.97e-19	2.51e-10	1.67e+01
1.0e+05	1.56e-19	1.53e-19	1.60e-19	2.43e-10	1.28e+01

We end this section with a brief comparison of our work with that of Blanchet, Lam, and Zwart (2011). As shown above, our method works effectively and is comparatively easy to implement. The complex nature of their algorithm can be attributed to their alternative approach, which relies on using complex asymptotic estimates. Furthermore, a critical issue with their algorithm is that it depends on several parameters, and effectively tuning them to produce optimal results seems to be challenging even when specialized to the simpler case of perpetuities without invoking Markov state-dependence. Additionally, their results on crude Monte Carlo—even for small values of u corresponding to $\Delta = 0.1$ and 0.5 in their paper—are quite different from ours. We emphasize that our crude Monte Carlo estimates are based on five million simulations and on the approximation of the limit variable V by V_{2000}. The latter approximation was based on extensive exploratory analysis. We are not able to explain this discrepancy with their results (as the implementation details are not available in their manuscript). However, we emphasize that our crude Monte Carlo estimates and importance sampling estimates are quite close to one another for moderate values of u, as one should expect. Finally, it is unclear from their numerical results whether their state-dependent sampler yields unbiased point estimates or bounded relative error.

5. Proofs of results concerning running time of the algorithm. We now turn to the proof of Theorem 2.4, which will rely on ideas from nonlinear renewal theory and Markov chain theory. In the proof, we will utilize the following:

LEMMA 5.1. *There exist positive constants* $\bar{M}$ *and* β *and a constant* $\rho \in (0,1)$ *such that*

$$\mathbf{E}_\xi \left[h(V_n) | V_{n-1} \right] \leq \rho h(V_{n-1}) \quad \text{on} \quad \{V_{n-1} \geq \bar{M}\}, \tag{5.1}$$

where $h(x) := x^{-\beta} \mathbf{1}_{\{x>1\}} + \mathbf{1}_{\{x \leq 1\}}$.

PROOF. Without loss of generality, assume that $V_{n-1} = v > 1$. Then by strong Markov property,

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

By assumption (H_4), the second term on the right-hand side of the above expression is $\mathbf{o}(v^{-\epsilon})$. As for the first term, notice that it can be expressed as $v^\beta \mathbf{E}_\xi \left[(A_1 \max(v^{-1} D_1, 1) + v^{-1} B_1)^{-\beta} \mathbf{1}_{\{V_1>1\}} | V_0 = v \right]$. Using the boundedness of $\left\{ (A_1 \max(v^{-1} D_1, 1) + v^{-1} B_1)^{-\beta} \mathbf{1}_{\{V_1>1\}} | V_0 = v \right\}$ in v, it follows that $\mathbf{E}_\xi \left[(A_1 \max(v^{-1} D_1, 1) + v^{-1} B_1)^{-\beta} \mathbf{1}_{\{V_1>1\}} | V_0 = v \right]$ converges as $v \to \infty$ to $\mathbf{E}_\xi \left[A_1^{-\beta} \right] = \lambda(\xi - \beta) < 1$ if $0 < \beta < \xi$. Thus, choosing $\beta = \epsilon \in (0, \xi)$, where ϵ is given as in (H_4), we obtain that the lemma holds for any $\rho = \left(\mathbf{E}_\xi \left[A_1^{-\epsilon} \right], 1 \right)$ and $\bar{M} < \infty$ sufficiently large. $\square$

PROOF OF THEOREM 2.4 (i). Let $\bar{M}$ be given as in Lemma 5.1. Without loss of generality, we may assume that $\bar{M} \geq \max(M, 1)$. Let $L = \sup \left\{ n \in \mathbb{Z}_+ : V_n \in (-\infty, \bar{M}] \right\}$. Then it follows directly from the definitions that $K \leq L$ on $\{K < \infty\}$. Thus it is sufficient to verify that $\mathbf{E}_\xi [L] < \infty$.

To this end, introduce two sequences of random times, as follows. Begin by setting $\mathcal{J}_0 = 0$ and $\mathcal{K}_0 = 0$. Then for each $i \in \mathbb{Z}_+$, set

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

Intuitively, $\{\mathcal{K}_i\}$ denotes the successive times that the process escapes from the interval $(-\infty, \bar{M}]$, while $\{\mathcal{J}_i\}$ denotes the successive times that the process subsequently returns to $(-\infty, \bar{M}]$, where $0 < \mathcal{K}_1 < \mathcal{J}_1 < \dots$. For any integer i, let $\kappa_i := \mathcal{K}_i - \mathcal{K}_{i-1}$ denote the ith inter-escape time from the interval $(-\infty, \bar{M}]$. Finally, let $\mathfrak{N}$ denote the total number of cycles that exit $(-\infty, \bar{M}]$

and subsequently return to $(-\infty, \bar{M}]$. Then it follows from these definitions that

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

Observe that $\mathbf{E}_\xi [\mathfrak{N}] < \infty$. Indeed, when the process $\{V_n\}$ escapes from $(-\infty, \bar{M}]$, the probability that it ever returns to this interval is bounded above by this probability conditioned on the starting state $V_0 = \bar{M}$. But by the transience of $\{V_n\}$ in Lemma 2.1 (ii), that probability is less than one. Thus

$$\sup_{v > \bar{M}} \mathbf{P}_\xi \left(V_n \in (-\infty, \bar{M}], \text{ some } n \geq 1 \middle| V_0 = v \right) \leq p < 1,$$

and consequently $\mathbf{E}_\xi [\mathfrak{N}] \leq \sum_{n=1}^{\infty} p^n < \infty$.

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

$$\mathbf{P}_\xi \left(\bar{K} = n \middle| V_0 = v \right) \leq \rho^n \frac{h(v)}{h(\bar{M})} \leq \rho^n.$$

Hence $\mathbf{E}_\xi \left[\bar{K} \mathbf{1}_{\{\bar{K} < \infty\}} \middle| V_0 = v \right] \leq \Theta$ for some finite constant Θ, uniformly in $v > \bar{M}$.

Thus, to establish the lemma, it is sufficient to show that $\mathbf{E}_\xi \left[\bar{N} \middle| V_0 = v \right] < \infty$, uniformly in $v \in (-\infty, \bar{M}]$, where $\bar{N}$ denotes the total number of visits of $\{V_n\}$ to $(-\infty, \bar{M}]$. To this end, observe that $[-\bar{M}, \bar{M}]$ is petite (Collamore and Vidyashankar (2011), Lemma 5.1). Moreover, it is easy to verify that $(-\infty, -\bar{M})$ is also petite for sufficiently large $\bar{M}$. Indeed, under Letac's Model E, we have that $\inf_{v \in (-\infty, -\bar{M})} \mathbf{P}_\xi \left(\max(D_1, V_0) = D_1 \middle| V_0 = v \right) > 0$, and in that case $V_1 = A_1 D_1 + B_1$. Thus, the transition kernel of $\{V_n\}$ satisfies a minorization with small set $(-\infty, -\bar{M})$. Consequently $(-\infty, \bar{M}]$ is petite and hence uniformly transient (Meyn and Tweedie (1993), Theorem 8.3.5, and the transience of $\{V_n\}$). We conclude that $\mathbf{E}_\xi \left[\bar{N} \right] < \infty$, uniformly in the initial state $V_0 \in (-\infty, \bar{M}]$, completing the proof. □

We now turn to the proof of Theorem 2.4 (ii). We begin with a slight variant of Lemma 4.1 in Collamore and Vidyashankar (2011).

LEMMA 5.2. *Assume Letac's Model E, and suppose that* (H_1), (H_2), *and*

(H_3) *are satisfied. Then*

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

for some random variable $\hat{V}$. The distribution of this random variable $\hat{V}$ is independent of the initial distribution of V_0 and is described as follows. If A^l is a typical ladder height of the process $S_n = \sum_{i=1}^n \log A_i$ in the ξ-shifted measure, then

$$\mathbf{P}_\xi\big(\log \hat{V} > y\big) = \frac{1}{\mathbf{E}_\xi\left[A^l\right]}\int_y^\infty \mathbf{P}_\xi\big(A^l > z\big)dz, \quad \text{for all } y \geq 0. \tag{5.3}$$

Proof. Note by definition of Z_n that

$$V_n = (A_0 \cdots A_n)\left(Z_n \mathbf{1}_{\{Z_n > 0\}} + Z_n \mathbf{1}_{\{Z_n \leq 0\}}\right).$$

Now it is shown in Lemma 5.2 of Collamore and Vidyashankar (2011) that $V_n \uparrow \infty$ w.p. 1 as $n \to \infty$ under the measure μ_ξ. Thus $T_u < \infty$ a.s., and at this exceedance time, we have that $\mathbf{1}_{\{T_u \leq 0\}} = 0$. Consequently, setting $X_n = Z_n$ on $\{Z_n > 0\}$ and $X_n = 1$ otherwise, we have that $V_{T_u} = V'_{T_u}$, where

$$V'_n = (A_0 \cdots A_n)\, X_n, \quad n = 0, 1, \ldots.$$

Since X_n is positive for all n, we can introduce the perturbed random walk

$$\log V'_n = S_n + \delta_n, \quad n = 0, 1, \ldots, \tag{5.4}$$

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

Note that S_n has a positive drift under μ_ξ-measure and that $\{(\log A_i, \delta_i) : i = 1, \ldots, n\}$ is independent of $\log A_j$ for all $j > n$. Thus classical nonlinear renewal theory can be applied to the sequence $\{\log V'_n\}$. To do so, we need to verify that the sequence $\{\delta_n\}$ is slowly changing. But by Lemma 3.1, we have that $\mathbf{1}_{\{Z_n \leq 0\}} \to 0$ a.s. and that $\delta_n := \log X_n$ converges to a proper random variable. Hence $\delta_n / n \to 0$ a.s., and consequently $\{\delta_n\}$ is slowly changing. Hence it follows from Theorem 4.2 of Woodroofe (1982) that

$$\frac{V'_{T_u}}{u} \Rightarrow \hat{V} \quad \text{as} \quad u \to \infty. \tag{5.5}$$

Note that this result holds *independent* of the initial distribution of $V_0 \in \mathcal{C}$.

Next observe that for any $y \geq 1$,

$$\mathbf{P}_\xi\left(\frac{V'_{T_u}}{u} > y\right) = \mathbf{P}_\xi\left(\frac{V'_{T_u}}{u} > y\,;\, T_u < K\right) + \mathbf{P}_\xi\left(\frac{V'_{T_u}}{u} > y\,;\, T_u \geq K\right). \tag{5.6}$$

By the strong Markov property,

$$\mathbf{P}_\xi\left(\frac{V'_{T_u}}{u} > y \Big| T_u \geq K\right) = \mathbf{P}_\xi\left(\frac{V'_{T_u}}{u} > y \Big| V_0 \sim \nu\right), \tag{5.7}$$

where ν is the distribution of V'_K conditional on $\{T_u \geq K\}$. Note by definition of K that $V'_K \in \mathcal{C}$. Hence it follows by (5.5) and (5.7) that

$$\lim_{u\to\infty} \mathbf{P}_\xi\left(\frac{V'_{T_u}}{u} > y \Big| T_u \geq K\right) = \mathbf{P}_\xi\left(\hat{V} > y\right), \quad y \geq 1. \tag{5.8}$$

Consequently by (5.5) and (5.6),

$$\lim_{u\to\infty} \mathbf{P}_\xi\left(\frac{V'_{T_u}}{u} > y \Big| T_u < K\right) = \mathbf{P}_\xi\left(\hat{V} > y\right), \quad y \geq 1, \tag{5.9}$$

provided that $\liminf_{u\to\infty} \mathbf{P}_\xi\left(T_u < K\right) > 0$. Finally, to verify that this last condition is fulfilled, observe that $\mathbf{P}_\xi\left(T_u < K\right)$ increases to $\mathbf{P}_\xi\left(K = \infty\right)$ as $u \to \infty$. Moreover, by Lemma 5.2 of Collamore and Vidyashankar (2011), we have that $V_n \uparrow \infty$ w.p. 1 as $n \to \infty$ in the ξ-shifted measure. Thus, starting from the stationary measure γ on $\mathcal{C}$, $\mathbf{P}_\xi\left(K = \infty\right)$ is strictly positive over any given cycle. Thus $\mathbf{P}_\xi\left(T_u < K\right) \to \Theta$ as $u \to \infty$, for some positive constant Θ. □

Proof of Theorem 2.4 (ii). Let V'_n be defined as in the proof of Lemma 5.2, and observe from the proof of this lemma that $\log V'_n = S_n + \delta_n$, where $\{\delta_n\}$ is slowly changing and $S_n = \sum_{i=1}^n \log A_i$.

Set $T'_u = \inf\{n : V'_n > u\}$. Then it follows by Lemma 9.13 of Siegmund (1985) that

$$\frac{T'_u}{\log u} \to \frac{1}{\Lambda'(\xi)} \quad \text{in probability} \tag{5.10}$$

with respect to the measure μ_ξ, since $\Lambda'(\xi) = \mathbf{E}_\xi[\log A]$. Now if $T_u < K$, then the process $\{V_n\}$ never returns to the set $(-\infty, 0]$, and hence we have in that case that $T_u = T'_u$. Moreover, since $V_n \uparrow \infty$ w.p. 1 as $n \to \infty$ in the ξ-shifted measure, $\mathbf{P}_\xi\left(T_u < K\right) \to \Theta$ as $u \to \infty$, for some positive constant Θ. Thus it follows from (5.10) that, 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 $\bar{M}$ be given as in Lemma 5.1, and first suppose that $\bar{M} \leq M$ and $\operatorname{supp}(V_n) \subset [-M, \infty)$ for all n. Then in this case, conditional on $\{T_u < K\}$,

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

Now apply Lemma 5.1. Iterating (5.1)—as in the proof of Lemma 3.3—we obtain that (3.29) holds with ρ in place of ρ_β. Consequently, using the explicit form of the function h in Lemma 5.1, we obtain that

$$\mathbf{P}_\xi \left(T_u > n \middle| T_u < K \right) \leq \left(\frac{1}{\mathbf{P}_\xi \left(T_u < K \right)} \right) \rho^n u^\beta, \quad \text{for all } n, \tag{5.11}$$

where β is given as in Lemma 5.1. Now $\mathbf{P}_\xi \left(T_u < K \right) \downarrow \Theta > 0$ as $u \to \infty$. Hence, letting $\mathbf{E}_\xi^{(u)} [\cdot]$ denote the expectation conditional on $\{T_u < K\}$, we conclude that

$$\mathbf{E}_\xi^{(u)} \left[\frac{T_u}{\log u}; \frac{T_u}{\log u} \geq \eta \right] \leq \bar{\Theta} \rho^{\eta \log u} u^\beta, \tag{5.12}$$

for some finite constant $\bar{\Theta}$, and for sufficiently large η, the right-hand side 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 $(-\infty, \bar{M}]$, as defined in the proof of Theorem 2.4 (i). Then $(T_u - L)$ describes the length of the last excursion to level u after exiting $(-\infty, \bar{M}]$ forever. By a repetition of the argument just given, we obtain that (5.11) holds with $(T_u - L)$ in place of T_u, and hence (5.12) also holds with $(T_u - L)$ in place of T_u. This implies $\{(T_u - L)/ \log u\}$ is uniformly integrable. Next observe by the proof of Theorem 2.4 (i) that $\mathbf{E}_\xi [L / \log u] \downarrow 0$ as $u \to \infty$. The result follows. □

In Theorem 2.4 (ii), we showed that

$$\mathbf{E}_\xi \left[T_u \middle| T_u < K \right] \sim \frac{\log u}{\Lambda'(\xi)} \quad \text{as} \quad u \to \infty, \tag{5.13}$$

which can be expected due to the fact that $\Lambda'(\xi) = \mathbf{E}_\xi [\log A]$. Specifically, if we were to replace $\{V_n\}$ with the multiplicative process $\{A_1 \cdots A_n\}$ and define $\tilde{T}_u = \inf\{n : (A_1 \cdots A_n) > u\}$, then by an elementary renewal argument applied to $S_n := \sum_{i=1}^n \log A_i$, we would obtain that (5.13) holds with the left-hand side replaced with $\mathbf{E}\big[\tilde{T}_u\big]$.

Next we turn to the proof of Theorem 2.4 (iii), which provides a similar asymptotic estimate for the return time to the set $\mathcal{C}$, assuming that $T_u < K$ and starting our analysis from the state $V_{T_u} > u$. Now if $T_u < K$, then starting at time T_u, we have $V_{T_u} \approx u$ and we would like to study its return time to $\mathcal{C} = [-M, M]$. Again, if we replace $\{V_n\}$ with the multiplicative process $\{(A_1 \cdots A_n)u\}$ and define $\tilde{K} = \inf\{n : (A_1 \cdots A_n)u \leq M\}$, then an elementary renewal argument yields

$$\mathbf{E}\big[\tilde{K}\big] \sim \frac{\log u}{|\Lambda'(0)|} \quad \text{as} \quad u \to \infty,$$

where we have used that $\mathbf{E}[\log A] = \Lambda'(0) < 0$. This heuristic argument serves as motivation for Theorem 2.4 (iii).

We will first establish Theorem 2.4 (iii) conditional on the event that $(V_0/u) = v > 1$, and later remove this assumption. Thus, we assume for the moment that the process *starts* at a level vu where $v > 1$, so the dual measure for the process agrees with its initial measure. For this process, define

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

LEMMA 5.3. *Let $(V_0/u) = v > 1$ and $t \in (0,1)$. Then under the conditions of Theorem 2.4,*

$$\lim_{u\to\infty} \frac{1}{\log u} \mathbf{E}\Big[L\left(u^t\right) \Big| \frac{V_0}{u} = v\Big] = \frac{1-t}{|\Lambda'(0)|}. \tag{5.14}$$

PROOF. For notational simplicity, we will suppress the conditioning on $(V_0/u) = v$ in the proof of the lemma. We begin by establishing an upper bound. First observe that, as in (3.27),

$$\frac{|V_n|}{|V_{n-1}|} \le A_n + \frac{(A_n|D_n| + |B_n|)}{|V_{n-1}|}. \tag{5.15}$$

Hence

$$\log\left(\frac{|V_n|}{|V_{n-1}|}\right) \le \log\Big(A_n + u^{-t}\left(A_n|D_n| + |B_n|\right)\Big), \quad n < L(u^t). \tag{5.16}$$

This shows that we may bound the process $\{|V_n| : n < L(u^t)\}$ by a classical random walk. More precisely, define

$$S_n^{(u)} := \sum_{i=1}^n X_i^{(u)} \quad \text{where} \quad X_i^{(u)} := \log\Big(A_i + u^{-t}\left(A_i|D_i| + |B_i|\right)\Big).$$

Then iterating (5.16) and using that $(V_0/u) = v$, we obtain

$$\log|V_n| - \log(vu) \le S_n^{(u)}, \quad \text{for all } n < L(u^t). \tag{5.17}$$

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

Since $\big\{S_n^{(u)}\big\}$ is a classical random walk and $\tilde{L}_u(u^t)$ is a stopping time, it follows by Wald's identity that $\mathbf{E}\big[S_{\tilde{L}_u(u^t)}\big] = \mathbf{E}\big[X_1^{(u)}\big]\mathbf{E}\big[\tilde{L}_u(u^t)\big]$. Thus letting

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

denote the overjump of the random walk $\{S_n^{(u)}\}$ over its boundary (which in this case is taken to be the level $(1-t)\log u + \log v$), we obtain

$$L(u^t) \le \frac{(1-t)\log u + \log v + \mathbf{E}[O_u]}{\left|\mathbf{E}\left[X_1^{(u)}\right]\right|}. \tag{5.18}$$

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

$$\lim_{u\to\infty} \frac{1}{\log u}\mathbf{E}[O_u] = 0. \tag{5.19}$$

To establish (5.19) note, as in the proof of Lorden's inequality (Asmussen (2003), Proposition V.6.1), that $\mathbf{E}[O_u] \le \mathbf{E}\left[Y_u^2\right]/\mathbf{E}[Y_u]$, 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

$$\mathbf{E}[Y_u] = m_u^{(1)} e^{\mathfrak{S}_u} \to \mathbf{E}[Y] \quad \text{as } u \to \infty, \tag{5.20}$$

where Y has the negative ladder height distribution of the process $\{S_n\}$,

$$m_u^{(j)} := \left|\mathbf{E}\left[X^{(u)}\right]\right|,\ j = 1, 2, \ldots, \quad \text{and} \quad \mathfrak{S}_u := \sum_{n=1}^{\infty} \frac{1}{n}\mathbf{P}\left(S_n^{(u)} > 0\right).$$

We observe that $\mathfrak{S}_u$ is the so-called Spitzer series. Similarly, an easy calculation (*cf.* Siegmund (1985), p. 176) yields

$$\mathbf{E}\left[Y_u^2\right] = m_u^{(2)} e^{\mathfrak{S}_u} - 2m_u^{(1)} e^{\mathfrak{S}_u} \sum_{n=1}^{\infty} \frac{1}{n}\mathbf{E}\left[\left(S_n^{(u)}\right)^+\right] \to \mathbf{E}\left[Y^2\right],\ u \to \infty. \tag{5.21}$$

Since $\mathbf{E}\left[(\log A)^3\right] < \infty \Longrightarrow \mathbf{E}\left[Y^j\right] < \infty$ for $j = 1, 2$, it follows that $\mathbf{E}[O_u] \to \mathbf{E}\left[Y^2\right]/\mathbf{E}[Y]$, *i.e.* to a finite constant, which implies (5.19). Thus (5.14) holds as an upper bound.

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

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

(Essentially, $\tilde{L}(z)$ is defined in the same way as $\tilde{L}_u(z)$, except that we now substitute the sequence $\{S_n\}$ in place of $\{S_n^{(u)}\}$.) Next observe that $V_n \ge A_n V_{n-1} - |B_n|$ provided that $V_{n-1} \ge 0$, and iterating this equation yields

$$V_n \ge (A_1 \cdots A_n) V_0 - W, \quad \text{where} \quad W := \lim_{n\to\infty} \sum_{i=1}^{n} \prod_{j=i+1}^{n} A_j |B_i|. \tag{5.22}$$

Now from the very definition of $\tilde{L}$ and the assumption that $(V_0/u) = v$, we obtain that

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

But by (5.22), $(A_1 \cdots A_k)V_0 > u^s \Longrightarrow V_k > u^t$ on $\{W \leq (u^s - u^t)\}$. Thus for all n, $\tilde{L}(u^s) \geq n \Longrightarrow L(u^t) \geq n$ on $\{W \leq (u^s - u^t)\}$, and consequently

$$\mathbf{E}\left[L(u^t)\right] \geq \mathbf{E}\left[\tilde{L}(u^s); W \leq (u^s - u^t)\right]. \tag{5.23}$$

To study the expectation on the right-hand side, recall that W satisfies the SFPE $W \stackrel{\mathcal{D}}{=} AW + |B|$, and hence by Theorem 2.1 of Collamore and Vidyashankar (2011),

$$\mathbf{P}\left(W > u^s - u^t\right) \sim \bar{C}u^{-s\xi} \quad \text{as } u \to \infty. \tag{5.24}$$

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

$$\frac{\tilde{L}(u^s)}{\log u} \to \frac{(1-s)}{|\Lambda'(0)|} \quad \text{a.s.} \quad \text{as } u \to \infty, \tag{5.25}$$

where in the denominator of this last expression, we have observed that $\mathbf{E}[\log A] = \Lambda'(0) < 0$. It follows from (5.25) that for any $\epsilon > 0$,

$$\lim_{u\to\infty} \mathbf{P}\left(\frac{\tilde{L}(u^t)}{\log u} \notin (r-\epsilon, r+\epsilon)\right) = 0, \quad \text{where} \quad r := \frac{1-s}{|\Lambda'(0)|}. \tag{5.26}$$

Substituting (5.24) and (5.26) into (5.24) and letting $\epsilon \to 0$, we obtain that

$$\liminf_{u\to\infty} \frac{1}{\log u}\mathbf{E}\left[L\left(u^t\right)\right] \geq \frac{1-s}{|\Lambda'(0)|}, \tag{5.27}$$

and the required lower bound is obtained by letting $s \downarrow t$. □

Lemma 5.4. *Under the conditions of Theorem 2.4,*

$$\lim_{t\downarrow 0}\left\{\limsup_{u\to\infty} \frac{1}{\log u}\mathbf{E}\Big[L(M) - L\left(u^t\right)\Big]\right\} = 0. \tag{5.28}$$

PROOF. Applying Lemma 3.3 with $\beta = 0$, we obtain that the drift condition $(\mathcal{D})$ holds, and hence for some $\alpha > 0$,

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

where $\rho \in (0,1)$, and $\mathcal{C} = [-M, M]$ (where $M \ge M_0$ and M_0 is given as in Lemma 3.3). Iterating this equation yields

$$\mathbf{E}\left[\mathbf{1}_{\{L(M)>n\}} \,\middle|\, V_0 = w\right] \le \rho^n \left(\frac{|w|}{M}\right)^\alpha, \quad \text{for all } n. \tag{5.29}$$

We now apply this result to obtain an estimate for $L(M) - L(u^t)$. Recall that $L(M) - L(u^t)$ measures the length of time required for the process, beginning at level $V_{L(u^t)}$, to enter the set $\mathcal{C} = [-M, M]$. Thus by definition, $\left|V_{L(u^t)}\right| \le u^t$. Hence we are in the setting of (5.29) with $w \le u^t$, and with $L(M)$ replaced by $L(M) - L(u^t)$. Using the strong Markov property and (5.29), we then obtain that

$$\mathbf{P}\Big(L(M) - L(u^t) > n\Big) \le \rho^n \left(\frac{u^t}{M}\right)^\alpha, \quad \text{for all } n. \tag{5.30}$$

Set $J_t(u) = L(M) - L(u^t)$ and $t' = t\alpha/(-\log\rho)$. It follows from the previous equation (upon summing over all $n \ge t' \log u$) that

$$\mathbf{E}\Big[J_t(u)\mathbf{1}_{\{J_t(u)\ge t'\log u\}}\Big] \le \frac{\rho^{t'\log u}}{1-\rho}\left(\frac{u^t}{M}\right)^\alpha = \frac{1}{(1-\rho)M^\alpha}. \tag{5.31}$$

Then

$$\limsup_{u\to\infty} \frac{1}{\log u}\mathbf{E}\Big[L(M) - L\left(u^t\right)\Big] \le t'. \tag{5.32}$$

Since $t' \downarrow 0$ as $t \downarrow 0$, we conclude (5.28). □

PROOF OF THEOREM 2.4 (iii). Set

$$H_u(v) = \frac{1}{\log u}\mathbf{E}\left[L(M) \,\middle|\, \frac{V_0}{u} = v\right].$$

Then it follows from Lemmas 5.3 and 5.4 that

$$\lim_{u\to\infty} H_u(v) = \frac{1}{|\Lambda'(0)|}. \tag{5.33}$$

Let $\hat\mu_u$, $\hat\mu$ denote the probability laws of the random variables V_{T_u}/u, $\hat V$, respectively, as given in 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

$$\lim_{u\to\infty} \frac{1}{\log u} \mathbf{E}\left[L(M) \,\middle|\, \frac{V_0}{u} \sim \hat{\mu}_u\right] := \lim_{u\to\infty} \int_{v\geq 0} H_u(v) d\hat{\mu}_u(v) = \frac{1}{|\Lambda'(0)|}. \tag{5.34}$$

This result will follow from (5.33), 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 (5.34) as

$$\int_{v\geq 0} H_u(v) d\left(\hat{\mu}_u - \hat{\mu}\right)(v) + \int_{v\geq 0} H_u(v) d\hat{\mu}(v). \tag{5.35}$$

To deal with the first term, we begin by obtaining an upper bound for $H_u(v)$. First note by a slight modification of (5.31) (with $t = 1$, $\theta = -\log \rho$, $J_1(u)$ replaced with $L(M)$, and t' replaced with an arbitrary constant r) that

$$\mathbf{E}\left[L(M)\mathbf{1}_{\{L(M)>r\log u\}} \,\middle|\, |V_0| \leq u\right] \leq \frac{u^{-r\theta}}{1-\rho}\left(\frac{u}{M}\right)^{\alpha}, \tag{5.36}$$

for all $r > 0$ and some $\alpha > 0$. Now choose $r > \alpha/\theta$. Then by the previous inequality,

$$\mathbf{E}\left[L(M)\mathbf{1}_{\{L(M)>r\log u\}} \,\middle|\, |V_0| \leq u\right] \leq \Theta_1, \tag{5.37}$$

for some finite constant Θ_1 which is independent of u. Consequently

$$\frac{1}{\log u}\mathbf{E}\left[L(M) \,\middle|\, |V_0| \leq u\right] \leq r + \frac{\Theta_1}{\log u}. \tag{5.38}$$

Next, we extend this estimate to the case where the starting state satisfies $(V_0/u) = v > 1$. To this end, we write $\mathbf{E}\left[L(M)|(V_0/u) = v\right]$ as a sum of two terms; first, the expected time for the process $\{V_n\}$ to reach $\mathcal{C} = [-M, M]$ starting from an initial state in $[-u, u]$; and second, the expected time to reach $[-u, u]$ starting from the initial state $(V_0/u) = v$. Using the strong Markov property, this yields

$$\mathbf{E}\left[L(M) \,\middle|\, \frac{V_0}{u} = v\right] \leq \sup_{w\in(M,u]} \mathbf{E}\left[L(M) \,\middle|\, |V_0| = w\right] + \mathbf{E}\left[L(u) \,\middle|\, \frac{V_0}{u} = v\right]. \tag{5.39}$$

For the second term, we bound $L(u)$ by the sojourn time for a classical random walk, as follows. Begin by observing that

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

Now taking logarithms, we see that $\mathbf{E}\left[L(u)\,|(V_0/u)=v\right]$ is bounded above by the length of time needed for the random walk

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

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

$$\mathbf{E}\left[L^*(u)\right] \leq \Theta_2(u)\log v + \Theta_3(u)$$

for constants $\Theta_2(u) \to m_1^{-1}$ and $\Theta_3(u) \to m_2/m_1^2$, where m_1 and m_2 are the first and second moments of the ladder height distribution for the sequence $\{\log A_i\}$; *cf.* the discussion following (5.19) above. (The required moment condition needed for this convergence to hold are satisfied due to the differentiability assumption in (H_1).) Finally, substituting this last bound and (5.38) into (5.39), we obtain that for some constant $\bar{\Theta}$, uniformly in $u \geq u_0$,

$$H_u(v) := \frac{1}{\log u}\mathbf{E}\left[L(M)\,\middle|\,\frac{V_0}{u} = v\right] \leq \bar{\Theta} + \frac{\log v}{m_1}. \tag{5.40}$$

Returning to (5.35), using the above upper bound, we now show that

$$\left|\int_{v\geq 0}\left(\bar{\Theta} + \frac{\log v}{m_1}\right) d\left(\hat{\mu}_u - \hat{\mu}\right)(v)\right| \to 0 \quad \text{as } u \to \infty. \tag{5.41}$$

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). First, write the process $\{V_n\}$ as a nonlinear renewal process as in (5.4). Then apply Lemma 3.1 to obtain (in the same notation as in (5.4)) that

$$\delta_n \leq \log\left(\sum_{i=0}^{\infty}\frac{\tilde{B}_i}{A_0\cdots A_i}\mathbf{1}_{\{K>i\}}\right), \quad n < K. \tag{5.42}$$

(This is the first equation in (3.1).) Next apply Jensen's inequality and Lemma 3.1 (ii) to obtain that

$$\frac{\xi}{2}\mathbf{E}_\xi\left[\left|\delta_{T_u} - \delta_{T_u-1}\right|\mathbf{1}_{\{T_u<K\}}\right] \leq \mathbf{E}_\xi\left[\log\left(\bar{Z}^{(p)}\right)^\xi\right] < \infty. \tag{5.43}$$

Notice also that conditions (6)-(8) of Lai and Siegmund (1979) are satisfied (with $\alpha = 1$). By (H_1), we have that $\mathbf{E}_\xi\left[(\log A)^2\right] < \infty$. By a slight modification of (5.43) followed by an application of Chebyshev's inequality, Eq. (7) of their article follows. Moreover, we observe that Theorem 2 of their article is actually valid if their Eq. (8) is replaced by uniform continuity in probability of $\{\delta_n\}$, as given in Eq. (4.2) of Woodroofe (1982). The latter property holds, since δ_n converges w.p. 1 to a proper random variable. We conclude that all the conditions in Lai and Siegmund (1979) are satisfied, and hence $\{|\log V_{T_u} - \log u|\}$ is uniformly integrable. Consequently (5.41) follows from the weak convergence $\hat{\mu}_u \Rightarrow \hat{\mu}$.

Finally, applying the dominated convergence theorem to the second term in (5.35) and using (5.33), we conclude that $(\log u)^{-1}\mathbf{E}\left[L(M)\,|(V_0/u) \sim \hat{\mu}\right] \to 1/|\Lambda'(0)|$, as required. □

6. Proof of optimality. Next we turn to the proof of the optimality theorem. We remark that a similar result was obtained in a different setting in Collamore (2002), although the current proof incorporates new aspects, most notably, the possibility that the alternative algorithm be state-dependent.

Proof of Theorem 2.4. We divide the proof into two parts, "asymptotic equivalence" and "uniqueness."

Part I: Asymptotic equivalence. We begin by showing that if $\left\{\nu(E; w, q) : E \in \mathcal{B}(\mathbb{R}^3), w \in \mathbb{R}, q \in \{0,1\}\right\}$ is any family of measures within the class $\mathfrak{M}$, where $\mathfrak{M}$ is given as in (2.15), then

$$\liminf_{u\to\infty} \frac{1}{\log u}\mathbf{E}_{\boldsymbol{\nu}}\left[\left(\mathcal{E}_u^{(\boldsymbol{\nu})}\right)^2\right] \geq \lim_{u\to\infty} \frac{1}{\log u}\mathbf{E}_{\mathfrak{D}}\left[\mathcal{E}_u^2\right] = -2\xi; \tag{6.1}$$

in other words, simulation under the dual measure used in our main algorithm is either asymptotically equivalent or preferable to any alternative in $\mathfrak{M}$.

To establish (6.1), set

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

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

First assume that $\nu \ll \mu_{\mathfrak{D}}$. Then by applying Lemma 4.2 twice, first to express the given expectation in terms of the original measure and then in terms of the dual measure, we obtain

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

where $Y_i := (\log A_i, B_i, D_i)$ is the driving sequence of the SFPE, $W_i := V_{i-1}/u$, and Q_i is the indicator function on $\{V_j > u, \text{for some } j < i\}$. Using the Radon-Nykodym theorem to simplify the last quantity in the previous display, we deduce that

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

Observe that by the definition of the dual measure, $\frac{d\mu}{d\mu_{\mathfrak{D}}} = \frac{d\mu}{d\mu_\xi}$ when $Q_i = 0$, because in this case, the process $\{V_n\}$ has not attained the level u by time i, and consequently the dual measure *agrees* with the ξ-shifted measure. Similarly, $\frac{d\mu}{d\mu_{\mathfrak{D}}} = 1$ when $Q_i = 1$, because in that case, the process $\{V_n\}$ has attained the level u by time i, in which case the dual measure agrees with the original measure. Thus, from the previous equation we obtain that

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

Now set

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

Substituting these definitions into (6.2) and noting that $\frac{d\mu}{d\mu_\xi}(Y_i) = e^{-\xi \log A_i}$, we obtain that

$$\mathbf{E}_{\boldsymbol{\nu}}\Big[\big(\mathcal{E}_u^{(\boldsymbol{\nu})}\big)^2\Big] = \mathbf{E}_{\mathfrak{D}}\left[N_u^2 \mathbf{1}_{\{T_u<K\}}\, e^{-2\xi S_{T_u} - R_K}\right]. \tag{6.3}$$

By Jensen's inequality and the observation that $N_u \geq 1$ on $\{T_u < K\}$, it follows that

$$\mathbf{E}_{\boldsymbol{\nu}}\Big[\big(\mathcal{E}_u^{(\boldsymbol{\nu})}\big)^2\Big] \geq p_u \exp\Big\{\mathbf{E}_{\mathfrak{D}}\left[-2\xi S_{T_u} - R_K \middle|\, T_u < K\right]\Big\}, \tag{6.4}$$

where

$$p_u := \mathbf{P}_\xi (T_u < K) \to \Theta \quad \text{as } u \to \infty.$$

where $\Theta := \mathbf{P}_\xi (K = \infty)$. Note that Θ is positive due to the transience of $\{V_n\}$ in its ξ-shifted measure, as stated in Lemma 2.1 (ii). Consequently (6.4) yields

$$\begin{aligned}(6.5)\qquad \liminf_{u\to\infty} \frac{1}{\log u} \log \mathbf{E}_{\boldsymbol{\nu}} \left[\left(\mathcal{E}_u^{(\boldsymbol{\nu})} \right)^2 \right] \geq &- \limsup_{u\to\infty} \frac{1}{\log u} \mathbf{E}_\xi \left[2\xi S_{T_u} \,|T_u < K \right] \\ &- \limsup_{u\to\infty} \frac{1}{\log u} \mathbf{E}_{\mathfrak{D}} \left[R_K \,|T_u < K \right].\end{aligned}$$

To identify the first term on the right-hand side of (6.5), note by Wald's identity that

$$\mathbf{E}_\xi [\log A]\, \mathbf{E}_\xi [T_u \wedge K] = \mathbf{E}_\xi \left[S_{T_u} \mathbf{1}_{\{T_u < K\}} \right] + \mathbf{E}_\xi \left[S_K \mathbf{1}_{\{T_u \geq K\}} \right],$$

which yields

$$(6.6)\qquad \mathbf{E}_\xi \left[S_{T_u} \,|T_u < K \right] = \mathbf{E}_\xi [\log A] \frac{\mathbf{E}_\xi [T_u \wedge K]}{\mathbf{P}_\xi (T_u < K)} - \frac{\mathbf{E}_\xi \left[S_K \mathbf{1}_{\{T_u \geq K\}} \right]}{\mathbf{P}_\xi (T_u < K)}.$$

Now consider the first term on the right-hand side. Since $\mathbf{E}_\xi [T_u \wedge K] = \mathbf{E}_\xi \left[T_u \mathbf{1}_{\{T_u < K\}} \right] + \mathbf{E}_\xi \left[K \mathbf{1}_{\{T_u \geq K\}} \right]$, we obtain that

$$(6.7)\qquad \frac{\mathbf{E}_\xi [T_u \wedge K]}{\mathbf{P}_\xi (T_u < K)} = \mathbf{E}_\xi \left[T_u \,|T_u < K \right] + \frac{\mathbf{E}_\xi \left[K \mathbf{1}_{\{T_u \geq K\}} \right]}{\mathbf{P}_\xi (T_u < K)}.$$

By Theorem 2.4 (ii),

$$(6.8)\qquad \lim_{u\to\infty} \frac{1}{\log u} \mathbf{E}_\xi \left[T_u \,|T_u < K \right] = \frac{1}{\Lambda'(\xi)}.$$

Moreover, as $u \to \infty$,

$$(6.9)\qquad \mathbf{E}_\xi \left[K \mathbf{1}_{\{T_u \geq K\}} \right] \to \mathbf{E}_\xi \left[K \mathbf{1}_{\{K < \infty\}} \right] < \infty,$$

where finiteness on the right-hand side is obtained by Theorem 2.4 (i). Since $\mathbf{P}_\xi (T_u < K) \to \mathbf{P}_\xi (K = \infty) = \Theta > 0$, by Lemma 2.1 (ii), it follows from (6.7) and the previous two equations that

$$(6.10)\qquad \lim_{u\to\infty} \frac{1}{\log u} \left(\frac{\mathbf{E}_\xi [T_u \wedge K]}{\mathbf{P}_\xi (T_u < K)} \right) = \frac{1}{\Lambda'(\xi)}.$$

Next consider the second term on the right-hand side of (6.6). By Corollary 3.1,

$$\mathbf{E}_\xi\left[e^{-\xi S_K}\mathbf{1}_{\{T_u\geq K\}}\right] = \mathbf{E}\big[\mathbf{1}_{\{T_u\geq K\}}\big] \to \mathbf{P}\left(K<\infty\right) \quad \text{as } u\to\infty.$$

Moreover, by the recurrence of the Markov chain $\{V_n\}$ in its original measure, $\mathbf{P}\left(K<\infty\right)=1$. Thus

$$1=\lim_{u\to\infty}\mathbf{E}_\xi\left[e^{-\xi S_K}\mathbf{1}_{\{T_u\geq K\}}\right]\geq \limsup_{u\to\infty}\mathbf{E}_\xi\left[e^{\xi|S_K|}\mathbf{1}_{\{S_K\leq 0\}}\mathbf{1}_{\{T_u\geq K\}}\right].$$

Applying the inequality $e^{\xi x}\geq 1+\xi x$, $x\geq 0$, to the expectation appearing on the right-hand side of the previous equation, we obtain

$$\limsup_{u\to\infty}\frac{1}{\log u}\Big(-\mathbf{E}_\xi\left[S_K\mathbf{1}_{\{T_u\geq K\}}\right]\Big)\leq 0. \tag{6.11}$$

Finally, substituting (6.10) and (6.11) into (6.6), using that $\mathbf{P}_\xi\left(T_u<K\right)\to\Theta>0$, and recalling that $\mathbf{E}_\xi\left[\log A\right]=\Lambda'(\xi)$, we conclude

$$\limsup_{u\to\infty}\frac{1}{\log u}\mathbf{E}_\xi\left[S_{T_u}\left|T_u<K\right.\right]\leq 1. \tag{6.12}$$

We now turn to the second limit on the right-hand side of (6.5). Our objective is to show that

$$\limsup_{u\to\infty}\frac{1}{\log u}\mathbf{E}_{\mathfrak{D}}\left[R_K\left|T_u<K\right.\right]\leq 0. \tag{6.13}$$

To establish (6.13), first introduce the assumption that

$$(\mathcal{A})\qquad \log\left(\frac{d\nu}{d\mu_{\mathfrak{D}}}\right) \text{ is bounded from below by a finite constant.}$$

This assumption will later be removed.

Claim: Under $(\mathcal{A})$, $\mathcal{M}_n:=R_n\mathbf{E}_{\mathfrak{D}}\left[\mathbf{1}_{\{T_u<K\}}\middle|\mathfrak{F}_n\right]$ is a supermartingale.

Proof of the Claim: Recall

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

Thus

$$\mathbf{E}_{\mathfrak{D}}\left[\mathcal{M}_n\middle|\mathfrak{F}_{n-1}\right]=\mathcal{M}_{n-1}+\mathbf{E}_{\mathfrak{D}}\left[U_n\mathbf{E}_{\mathfrak{D}}\left[\mathbf{1}_{\{T_u<K\}}\middle|\mathfrak{F}_n\right]\middle|\mathfrak{F}_{n-1}\right].$$

In the last term on the right-hand side, the inner conditional expectation is bounded above by one, and hence this entire term is bounded above by $\mathbf{E}_{\mathfrak{D}}\left[U_n \,|(W_n, Q_n)\right]$, where we have used the strong Markov property to replace $\mathfrak{F}_{n-1}$ with $(V_{n-1}, \mathbf{1}_{\{T_u \leq n-1\}})$, or equivalently (W_n, Q_n). Now condition on the event that $(W_n, Q_n) = (w, q)$ and apply Jensen's inequality to deduce that

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

where, for notational simplicity, we have suppressed the dependence on (w, q) in the above integrals. Since this inequality holds for any (w, q), we conclude that

$$\mathbf{E}_{\mathfrak{D}}\left[\mathcal{M}_n \middle| \mathfrak{F}_{n-1}\right] \leq \mathcal{M}_{n-1}. \tag{6.14}$$

Next, after using the inequality $\log x \leq x$, observe that

$$\mathbf{E}_{\mathfrak{D}}\left[U_n^+ \,|(W_n, Q_n) = (w, q)\right] \leq \int_{G^+(v,q)} d\nu(y; w, q) \leq 1,$$

where $G^+(v, q) := \left\{y : \log\left(\frac{d\nu}{d\mu_{\mathfrak{D}}}(y; w, q)\right) \geq 0\right\}$ and $U_n^+ = \max(U_n, 0)$. Moreover, since we are assuming that $\log\left(\frac{d\nu}{d\mu_{\mathfrak{D}}}\right)$ is bounded from below by a constant, it follows that U_n is also bounded from below by a constant, uniformly in n (where this constant is independent of $\mathfrak{F}_{n-1}$). Thus, setting $U_n^- = -\min(U_n, 0)$, we obtain that

$$\mathbf{E}_{\mathfrak{D}}\left[U_n^- \,|(W_n, Q_n) = (w, q)\right] \leq \Theta_0,$$

where Θ_0 is a finite constant independent of (w, q). Thus

$$\mathbf{E}\Big[|\mathcal{M}_n - \mathcal{M}_{n-1}|\Big|\mathfrak{F}_{n-1}\Big] \leq \Theta_0 + 1. \tag{6.15}$$

It follows that $\{\mathcal{M}_n\}_{n\in\mathbb{Z}_+}$ lies in L^1. By (6.14), we conclude that $\{\mathcal{M}_n\}_{n\in\mathbb{Z}_+}$ is a supermartingale. □

Returning to the proof of the theorem, observe that since $\{\mathcal{M}_n\}$ is a supermartingale, it follows from the optional sampling theorem that

$$\mathbf{E}_{\mathfrak{D}}\left[R_K \mathbf{1}_{\{T_u < K\}}\right] = \mathbf{E}_{\mathfrak{D}}\left[\mathcal{M}_K\right] \leq 0, \tag{6.16}$$

implying that

$$\limsup_{u\to\infty} \frac{1}{\log u} \mathbf{E}_{\mathfrak{D}}\left[R_K \mathbf{1}_{\{T_u < K\}}\right] \leq 0. \tag{6.17}$$

Substituting (6.12) and (6.17) into (6.5), we conclude that

$$\liminf_{u\to\infty} \frac{1}{\log u} \mathbf{E}_{\boldsymbol{\nu}} \left[\left(\mathcal{E}_u^{(\boldsymbol{\nu})} \right)^2 \right] \geq -2\xi. \tag{6.18}$$

This is the required result, since it is known by Theorem 2.3 that

$$\limsup_{u\to\infty} \frac{1}{\log u} \mathbf{E}_{\mathfrak{D}} \left[\mathcal{E}_u^2 \right] \leq -2\xi.$$

Thus we see that simulation under the dual measure is either asymptotically equivalent or optimal when compared with any other member in $\mathfrak{M}$.

If $\log\left(\frac{d\nu}{d\mu_{\mathfrak{D}}}\right)$ is not bounded from below by a constant, then in (6.2) we can replace ν with a larger measure, namely $\nu^{(\epsilon)} := \nu + \epsilon\mu_{\mathfrak{D}}$, where $\epsilon > 0$. Then (6.2) yields

$$\mathbf{E}_{\boldsymbol{\nu}}\Big[\left(\mathcal{E}_u^{(\boldsymbol{\nu})}\right)^2\Big] \geq \mathbf{E}_{\mathfrak{D}} \left[N_u^2 \mathbf{1}_{\{T_u<K\}} \prod_{i=1}^{T_u} \left(\frac{d\mu}{d\mu_\xi}(Y_i) \right)^2 \prod_{j=1}^{K} \frac{d\mu_{\mathfrak{D}}}{d\nu^{(\epsilon)}}(Y_j; W_j, Q_j) \right]. \tag{6.19}$$

Since $\log\left(\frac{d\nu^{(\epsilon)}}{d\mu_{\mathfrak{D}}}\right) \geq \log\epsilon$, the entire proof can be repeated without change, except that when applying Jensen's inequality we now obtain

$$\mathbf{E}_{\mathfrak{D}} \left[U_n \,|(W_n, Q_n) = (w,q) \right] \leq \log \int_{\mathbb{R}^3} d\nu_\epsilon(y; w, q) = \log(1+\epsilon) \leq \epsilon,$$

and hence in this case the optional sampling theorem yields

$$\mathbf{E}_{\mathfrak{D}} \left[R_K \mathbf{1}_{\{T_u<K\}} \right] \leq \epsilon \mathbf{E}_{\mathfrak{D}} \left[K \right]. \tag{6.20}$$

This last expectation can be decomposed into three terms,

$$\mathbf{E}_\xi \left[K \mathbf{1}_{\{K \leq T_u\}} \right] + \mathbf{E}_\xi \left[T_u \mathbf{1}_{\{T_u<K\}} \right] + \mathbf{E}_{\mathfrak{D}} \left[(K - T_u) \mathbf{1}_{\{T_u<K\}} \right].$$

Hence, by applying Theorem 2.4 (i)-(iii), we obtain

$$\limsup_{u\to\infty} \frac{1}{\log u} \mathbf{E}_{\mathfrak{D}} \left[R_K \mathbf{1}_{\{T_u<K\}} \right] \leq \epsilon \left(\frac{1}{\Lambda'(\xi)} + \frac{1}{|\Lambda'(0)|} \right). \tag{6.21}$$

But $\epsilon > 0$ was arbitrary. Thus we conclude that (6.18) holds in general.

Part II: Uniqueness. It remains to show that *strict* inequality holds in (6.18) when the family $\{\nu(\cdot)\}$ differs from the dual measure $\mu_{\mathfrak{D}}$. First observe by Jensen's inequality that

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

where, as before, we have suppressed the dependence on (w, q) in the above integrals. Moreover, for any given (w, q), equality holds in this inequality if and only if $\frac{d\nu}{d\mu_{\mathfrak{D}}}$ is equal to a constant a.s. Thus, if $\nu \neq \mu_{\mathfrak{D}}$, then there exists a point (w, q) where

$$\mathbf{E}_{\mathfrak{D}}\left[U_n \,|(W_n, Q_n) = (w, q)\right] = -\Delta, \quad \text{for some } \Delta > 0. \tag{6.22}$$

Our next objective is to use the continuity condition (C_0) to extend (6.22) from a point $w \in [0, 1]$ to an interval. The continuity assumption (C_0) states that

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

is piecewise continuous as a function of w, for any fixed $Q_n \in \{0, 1\}$. We claim that, as a consequence,

$$\mathbf{E}_{\mathfrak{D}}\left[U_n \Big| W_n = w\right] := \mathbf{E}_{\mathfrak{D}}\left[\log\left(\frac{d\nu}{d\mu_{\mathfrak{D}}}(Y_n; W_n, Q_n)\right) \Big| W_n = w\right] \tag{6.23}$$

is also piecewise continuous as a function of w. To establish (6.23), recall that we are assuming $\mu \ll \nu$ and $\nu \ll \mu_{\mathfrak{D}}$. (The latter condition will later be dropped.) Then by the Radon-Nikodym theorem,

$$\log\left(\frac{d\mu}{d\nu}\right) = -\log\left(\frac{d\nu}{d\mu_{\mathfrak{D}}}\right) + \log\left(\frac{d\mu}{d\mu_{\mathfrak{D}}}\right).$$

Since the expectation of the second quantity on the right-hand side is clearly piecewise continuous as a function of w, (6.23) follows immediately from the piecewise continuity in (C_0). Using the piecewise continuity in (6.23), observe that (6.22) implies that this equation holds for all $(w, q) \in G \times \{r\}$, where G is a neighborhood containing the original point w in (6.22) and where $r \in \{0, 1\}$.

We now show that by sharpening the estimate in Jensen's inequality on the set $G \times \{r\}$, we obtain a *strict* inequality in (6.18). First assume that $\log\left(\frac{d\nu}{d\mu_{\mathfrak{D}}}\right)$ is bounded from below by a constant, *i.e.*, that condition $(\mathcal{A})$ holds. Then by repeating our previous arguments, but using the sharper estimate (6.22) (when $(w, q) \in G \times \{r\}$) together with Jensen's inequality (when $(w, q) \notin G \times \{r\}$), we obtain

$$\mathcal{M}_n^* := (U_1^* + \cdots + U_n^*)\, \mathbf{E}_{\mathfrak{D}}\left[\mathbf{1}_{\{T_u < K\}} \big| \mathfrak{F}_n\right]$$

is a supermartingale, where $U_i^* := U_i + \Delta \mathbf{1}_{\{W_n \in G\}} \mathbf{1}_{\{Q_n = r\}}$. Applying the optional sampling theorem to this process $\{\mathcal{M}_n^*\}$, we deduce that

$$\mathbf{E}_{\mathfrak{D}}\left[R_K \mathbf{1}_{\{T_u < K\}}\right] \leq -\Delta\left\{\mathbf{1}_{\{r=0\}} \mathbf{E}_{\mathfrak{D}}\left[\mathcal{O}_u^{(0)}\right] + \mathbf{1}_{\{r=1\}} \mathbf{E}_{\mathfrak{D}}\left[\mathcal{O}_u^{(1)}\right]\right\}, \tag{6.24}$$

where

$$\mathcal{O}_u^{(0)} := \sum_{n=0}^{T_u} \mathbf{1}_{\{W_n \in G\}} \quad \text{and} \quad \mathcal{O}_u^{(1)} := \sum_{n=T_u+1}^{K} \mathbf{1}_{\{W_n \in G\}};$$

that is, $\mathcal{O}_u^{(0)}$ denotes the occupation time 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 T_u and ending at time K.

First assume that $r = 0$, so we are in the first scenario, where we would like to study the occupation measure $\mathcal{O}_u^{(0)}$ of the process $\{\log V_n / \log u\}$ in the set G prior to time T_u or, equivalently, the occupation measure of the process $\{V_n\}$ in the set $G(u) := [u^s, u^t]$ prior to time T_u. First introduce a subinterval $G'(u) := [u^{s'}, u^{t'}]$ strictly contained in $[u^s, u^t]$. Since $\{V_n\}$ is transient in its ξ-shifted measure, by Lemma 2.1 (ii), $\mathbf{P}_\xi\left(T_u < K\right) \to \mathbf{P}_\xi\left(K = \infty\right) = \Theta > 0$ as $u \to \infty$. Thus $\{V_n\}$ enters the region $[u^{s'}, \infty)$ with a positive probability which tends to Θ as $u \to \infty$. We claim that, in fact, $\{V_n\}$ enters the smaller region $G'(u)$ with a positive probability tending to Θ as $u \to \infty$. To establish this claim, observe by Lemma 5.2 that conditional on $\{T_u < K\}$,

$$\log V_{T_v} \Rightarrow \log v + \log \hat{V} \quad \text{as } v \to \infty$$

for a proper random variable $\hat{V}$ supported on $(0, \infty)$. Consequently, setting $v(u) = u^{s'}$, we obtain that

$$\begin{aligned} (6.25)\quad & \mathbf{P}_\xi\left(V_{T_{v(u)}} \in G'(u) \text{ and } T_{v(u)} < K\right) \\ & \quad \to \mathbf{P}_\xi\left(\log \hat{V} \le (t' - s') \log u \text{ and } T_{v(u)} < K\right) \to \Theta \quad \text{as } u \to \infty. \end{aligned}$$

Now consider the process $\{V_n\}$ *after* it enters the region $G'(u)$. In particular, suppose that this process *starts* with an initial state $V_0 \in G'(u)$. Then our objective is to show that the expected number of visits of $\{V_n\}$ to $G(u)$ prior to termination at time T_u behaves, roughly speaking, like a constant multiple of $\log u$ as $u \to \infty$.

Fix $\delta < \min(s' - s, t - t')$. Then observe by a law of large numbers argument that given δ, there exists a positive constant $\eta = \eta(\delta)$ such that

$$(6.26)\qquad \mathbf{P}_\xi\left(|S_n| > \delta \log u, \text{ some } n \le \eta \log u\right) \to 0 \quad \text{as } u \to \infty,$$

where $S_n = \sum_{i=1}^n \log A_i$. Next observe that if $V_{n-1} \ge 0$, then the equation $V_n = A_n \max\left(D_n, V_{n-1}\right) + B_n$ yields the inequalities

$$A_n V_{n-1} - \tilde{B}_n \le V_n \le A_n V_{n-1} + \tilde{B}_n,$$

where $\tilde{B}_n := (A_n|D_n| + |B_n|)$. Iterating these equations yields

$$V_0 - W \le \frac{V_n}{(A_1 \cdots A_n)} \le V_0 + W, \quad \text{where} \quad W := \sum_{i=1}^{\infty} \frac{\tilde{B}_i}{A_1 \cdots A_i}. \tag{6.27}$$

Hence, if $V_0 \in G'(u) := [u^{s'}, u^{t'}]$ and $V_n \notin G(u) := [u^s, u^t]$ for some $n \le \eta \log u$, then we must either have

$$(A_1 \cdots A_n) \notin [u^{-\delta}, u^{\delta}], \quad \text{for some } n \le \eta \log u, \tag{6.28}$$

or

$$W \ge \min\left(u^{s'} - u^{s+\delta}, u^{t-\delta} - u^{t'}\right). \tag{6.29}$$

But in the ξ-shifted measure, W is a perpetuity sequence which converges to a proper random variable (*cf.* the proof of Lemma 5.2 in Collamore and Vidyashankar (2011)). Hence for sufficiently small δ, the probability in (6.29) tends to zero as $u \to \infty$. Similarly, the probability in (6.28) also tends to zero as $u \to \infty$, by (6.26). Thus we conclude

$$\mathbf{P}_\xi\left(V_n \notin G(u), \text{ some } n \le \eta \log u \,|\, V_0 \in G'(u)\right) \to 0 \quad \text{as } u \to \infty, \tag{6.30}$$

uniformly in the initial state $V_0 \in G'(u)$. Combining this result with (6.25) and invoking the strong Markov property, we obtain that

$$\liminf_{u\to\infty} \frac{1}{\log u} \mathbf{E}\left[\mathcal{O}_u^{(0)}\right] \ge \eta > 0. \tag{6.31}$$

Substituting this estimate into (6.24) yields

$$\limsup_{u\to\infty} \frac{1}{\log u} \mathbf{E}_{\mathfrak{D}}\left[R_K \mathbf{1}_{\{T_u < K\}}\right] \le -\Delta\eta < 0. \tag{6.32}$$

Finally, substituting this last estimate and (6.12) into (6.5), we conclude that

$$\liminf_{u\to\infty} \frac{1}{\log u} \mathbf{E}_{\boldsymbol{\nu}}\left[\mathcal{E}_{u,\boldsymbol{\nu}}^2\right] \ge -2\xi + \Delta\eta, \tag{6.33}$$

as required.

Next suppose that $r = 1$. Thus we are in the second scenario, where we would like to study the occupation measure $\mathcal{O}_u^{(1)}$ of $\{V_n\}$ over an interval $G(u) = [u^s, u^t]$ between the time T_u and K. By the Markov property, it is then sufficient to study a process $\{V_n\}$ which *begins* at a level $V_0 > u$ and terminates at time K. Suppose that $V_0/u = v$, for some constant $v > 1$. Then begin by writing (6.27) in a slightly different form, namely,

$$\sup_n \left| V_n - V_0 \prod_{i=1}^n A_i \right| \le W', \text{ where } W' := \sum_{i=1}^\infty \tilde{B}_i \prod_{j=i+1}^\infty A_j, \tag{6.34}$$

provided that $\{V_0, \ldots, V_{n-1}\}$ is nonnegative. Now in the original measure, W' converges to a proper random variable and hence for any $\delta > 0$, $\mathbf{P}\left(W' > u^\delta\right) \to 0$ as $u \to \infty$ (*cf.* the proof of Lemma 5.2 in Collamore and Vidyashankar (2011)). Thus it is sufficient to study the occupation measure of the multiplicative process $V_0 \prod_{i=1}^n A_i$ in a region $G'(u) = [u^{s'}, u^{t'}]$, where $[s', t']$ is strictly contained in $[s, t]$.

Let $\mathfrak{T}_u$ denote the first time that the multiplicative random walk enters the set $(-\infty, u^{s'}]$. Then by Heyde's (1966) a.s. version of the renewal theorem, we have for any $\epsilon > 0$ that

$$\mathbf{P}\left(\mathfrak{T}_u \ge \frac{(1 - s' - \epsilon)\log u}{\mathbf{E}[\log A]} \,\middle|\, \frac{V_0}{u} = v \right) \to 1 \quad \text{as } u \to \infty, \tag{6.35}$$

uniformly in $v > 1$. Similarly, if $\mathfrak{L}_u$ denotes the last exit time of the multiplicative random walk from the interval $[u^{t'}, \infty)$, then by the strong law of large numbers, it follows that

$$\mathbf{P}\left(\mathfrak{L}_u \le \frac{(1 - t' + \epsilon)\log u}{\mathbf{E}[\log A]} \,\middle|\, \frac{V_0}{u} = v \right) \to 0 \quad \text{as } u \to \infty, \tag{6.36}$$

uniformly in $v > 1$. From the last two equations, we conclude that if $\eta \ge (t' - s' - 2\epsilon)/\mathbf{E}[\log A]$, then $\mathbf{P}(N_u^* \ge \eta \log u) \to 1$ as $u \to \infty$, where N_u^* denotes the number of visits of the multiplicative random walk to the set $G'(u)$. Consequently,

$$\liminf_{u\to\infty} \frac{1}{\log u} \mathbf{E}\big[\mathcal{O}_u^{(1)}\big] \ge \eta, \tag{6.37}$$

yielding (6.33), as required.

It remains to consider the case where $(\mathcal{A})$ is violated. But mimicking the argument in Part I, we may set $\nu^{(\epsilon)} = \nu + \epsilon\mu_{\mathfrak{D}}$ and repeat the previous arguments to obtain that (6.33) holds, but with $-2\xi + \Delta\eta + \epsilon\Theta_1$ on the right-hand side, for some constant Θ_1 (*cf.* (6.21)). Since ϵ was arbitrary, the desired result follows.

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,

$$\nu = \nu_a + \nu_s, \quad \text{where} \quad \nu_a \ll \mu_{\mathfrak{D}} \quad \text{and} \quad \nu_s \perp \mu_{\mathfrak{D}}.$$

The proof can now be repeated, replacing everywhere ν with ν_a. The mutually singular measure ν_s plays no role, since $\frac{d\mu_{\mathfrak{D}}}{d\nu}$ is zero on the support of this measure and thus, for example, the expectation on the right of (6.2) can be evaluated equally well with respect to ν_a as ν. Moreover, with respect to the measure ν_a, $\frac{d\mu_{\mathfrak{D}}}{d\nu} = \frac{d\mu_{\mathfrak{D}}}{d\nu_a}$. The proof carries through in exactly the same manner as before, so we omit the details. $\square$

7. Concluding remarks. In this paper, we developed and analyzed a dynamic importance sampling algorithm for rare event simulation of processes generated via SFPEs, encompassing a class of financial time series models and actuarial risk models. Extensions of these ideas to stratified importance sampling and general nonlinear recursions are currently under study.

References.

Alsmeyer, G., A. Iksanov, and U. Rösler (2009). On distributional properties of perpetuities. *J. Theoret. Probab.* **22**, 666–682.

Asmussen, S. (1985). Conjugate processes and the simulation of ruin problems. *Stoch. Proc. Appl.* **20**, 213–229.

Asmussen, S. (2003). *Applied Probability and Queues* (2nd ed.). Berlin: Springer.

Asmussen, S. and P. Glynn (2007). *Stochastic Simulation: Algorithms and Analysis*. Berlin: Springer.

Blanchet, J., H. Lam, and B. Zwart (2011). Rare-event simulation methodology for perpetuities. Submitted.

Blanchet, J. and J. Liu (2010). Efficient importance sampling in ruin problems for multidimensional regularly varying random walks. *J. Appl. Probab.* **47**, 301–322.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **31**, 307–327.

Bucklew, J., P. Ney, and J. Sadowsky (1990). Monte Carlo simulation and large deviations theory for uniformly recurrent Markov chains. *J. Appl. Prob.* **27**, 44–59.

Chan, H. and T. Lai (2007). Efficient importance sampling for Monte Carlo evaluation of exceedance probabilities. *Ann. Appl. Probab.* **17**, 440–473.

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

Collamore, J. F. and A. N. Vidyashankar (2011). Tail estimates for solutions of nonlinear recursions via non-linear renewal theory. Submitted.

Devroye, L. and O. Fawzi (2010). Simulating the Dickman distribution.

Stat. Probab. Lett. **80**, 242–247.

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

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**, 987–1007.

Fill, J. A. and M. L. Huber (2010). Perfect simulation of Vervaat perpetuities. *Electron. J. Probab.* **15**, 96–109.

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

Hammersley, J. M. and D. C. Handscomb (1964). *Monte Carlo Methods.* London: Chapman and Hall.

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

Iscoe, I., P. Ney, and E. Nummelin (1985). Large deviations of uniformly recurrent Markov additive processes. *Adv. Appl. Math.* **6**, 373–412.

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

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

Lehtonen, T. and H. Nyrhinen (1992). Simulating level-crossing probabilities by importance sampling. *Adv. Appl. Prob.* **24**, 858–874.

Letac, G. (1986). A contraction principle for certain Markov chains and its applications. random matrices and their applications. *Proceedings of AMS-IMS-SIAM Joint Summer Research Conference 1984. Contemporary Mathematics* **50**, 263–273.

Meyn, S. and R. Tweedie (1993). *Markov Chains and Stochastic Stability.* Berlin: Springer-Verlag.

Nummelin, E. (1984). *General Irreducible Markov Chains and Non–negative Operators.* Cambridge: Cambridge University Press.

Nummelin, E. and P. Tuominen (1982). Geometric of Harris recurrent Markov chains with applications to renewal theory. *Stoch. Process. Appl.* **12**, 187–202.

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

Siegmund, D. (1985). *Sequential Analysis: Tests and Confidence Intervals.* New York: Springer-Verlag.

Solomon, F. (1972). *Random Walks in Random Environment.* Ph.D. dissertation, Cornell University.

Vervaat, W. (1979). On a stochastic difference equation and a representation of non-negative infinitely divisible random variables. *Adv. Appl. Probab.* **11**, 750–783.

Woodroofe, M. (1982). *Nonlinear Renewal Theory in Sequential Analysis.* Philadelphia: SIAM.

Department of Mathematical Sciences
University of Copenhagen
Universitetsparken 5
DK-2100 Copenhagen Ø
Denmark
E-mail: collamore@math.ku.dk

Department of Statistics
George Mason University
4400 University Drive, MS 4A7
Fairfax, VA 22030
U.S.A.
E-mail: gdiao@gmu.edu

Department of Statistics
George Mason University
4400 University Drive, MS 4A7
Fairfax, VA 22030
U.S.A.
E-mail: avidyash@gmu.edu