A Hybrid Estimate for the Finite-Time Ruin Probability in a Bivariate Autoregressive Risk Model with Application to Portfolio Optimization

Qihe Tang* and Zhongyi Yuan†

Abstract

Consider a discrete-time risk model in which the insurer is allowed to invest a proportion of its wealth in a risky stock and keep the rest in a risk-free bond. Assume that the claim amounts within individual periods follow an autoregressive process with heavy-tailed innovations and that the log-returns of the stock follow another autoregressive process, independent of the former one. We derive an asymptotic formula for the finite-time ruin probability and propose a hybrid method, combining simulation with asymptotics, to compute this ruin probability more efficiently. As an application, we consider a portfolio optimization problem in which we determine the proportion invested in the risky stock that maximizes the expected terminal wealth subject to a constraint on the ruin probability.

1. Introduction

Classical risk theory may be questioned on two principal grounds. First, it often assumes that claim sizes are independent and identically distributed (i.i.d.) random variables. This assumption is far unrealistic because claims from the same insurance policy during successive reference periods or claims from different policies during the same reference period are generated in very similar situations, and, hence, they should be dependent on each other. Second, it is only recently that researchers have started to account for investments while modeling insurance business. At this stage of the study, some strong and unrealistic conditions have to be assumed on the investment portfolio. For instance, log-return rates are often assumed to follow a multivariate normal distribution. This assumption is dubious because data in insurance and finance often show pronounced asymmetry and heavy-tailedness, both violating the normality assumption. In particular, the multivariate normal model implies pairwise asymptotic independence, while data in insurance and finance often show strong asymptotic dependence, indicating a much riskier situation than described by the multivariate normal model. See Bingham et al. (2002, 2003) for related discussions.

As a consequence, there is a critical need for modifications that can reflect the skewness, heavy-tailedness, and asymptotic dependence of insurance and financial risks. In the absence of such modifications, the study could end up with naive solutions or potentially fatal mistakes.

We propose a bivariate autoregressive (AR) risk model in discrete time, which enables us to accommodate the modifications mentioned above. In this model, one-period claim amounts are assumed to follow an AR process with i.i.d. heavy-tailed innovations. The insurer is allowed to make investments in

---

* Qihe Tang is Professor of Actuarial Science at the Department of Statistics and Actuarial Science, University of Iowa, 241 Schaeffer Hall, Iowa City, IA 52242, qihe-tang@uiowa.edu.
† Zhongyi Yuan is Ph.D. candidate in Statistics with concentration in Actuarial Science/Financial Mathematics at the Department of Statistics and Actuarial Science, University of Iowa, 241 Schaeffer Hall, Iowa City, IA 52242, zhongyi-yuan@uiowa.edu.
a discrete-time financial market consisting of a risk-free bond and a risky stock. The log-return rates of the risky stock are assumed to follow another AR process, independent of the former one, with i.i.d., but not necessarily normal, innovations. To keep the paper short, we assume that both AR processes are of order 1, but we would like to point out that our study can be extended to AR models of higher order or even to autoregressive moving average (ARMA) models.

AR (or, more generally, ARMA) models have been extensively applied to actuarial science in the literature. In this regard we refer the reader to the following references, among many others. Gerber (1981, 1982) used AR and ARMA structures for one-period net gains (premiums minus claims). Wilkie (1986, 1987) discussed in great detail the relevance of mean-reverting AR structures in stochastic investment models. Mikosch and Samorodnitsky (2000) modeled one-period net losses by a two-sided linear process and studied the asymptotic behavior of the ultimate ruin probability for the heavy-tailed case. Cai (2002) assumed that the return rates of a risk-free investment possess an AR structure. Yang and Zhang (2003) modeled claim amounts by an AR process under a constant risk-free rate. The book Hardy (2003) contains a short review of AR models in the risk management context.

Our idea of assuming a bivariate discrete-time risk model for the stochastic structures of both claim amounts and investment returns stems from the works of Nyrhinen (1999, 2001) and Tang and Tsitsiashvili (2003a, 2004). These authors earlier considered ruin problems in the presence of both insurance processes and investment returns in a discrete-time setting but under different assumptions. The study of ruin in a continuous-time setting that integrates insurance processes with investment returns has a long history. We refer the reader to Norberg (1999), Frolova et al. (2002), Kalashnikov and Norberg (2002), Paulsen (2008), Chen (2011), and Hao and Tang (2012), to name a few. Chapter VIII of Asmussen and Albrecher (2010) contains a review of the study in both continuous-time and discrete-time models.

The contribution of this paper is threefold. First, we derive an asymptotic formula for the finite-time ruin probability, which quantitatively captures the impact of dependence of the claim amounts and of the investment returns. Second, we propose a hybrid method to compute the finite-time ruin probability. This method makes a compromise between the crude Monte Carlo estimate and the asymptotic estimate, and it effectively overcomes their deficiencies. Finally, as an application, we consider a portfolio optimization problem in which we determine the proportion invested in the stock that maximizes the expected terminal wealth of the insurer subject to a solvency constraint.

Two possible extensions of this study were proposed by anonymous referees. The first is to model claim amounts and investment returns by nonlinear AR models, such as threshold autoregressive models; for details of threshold autoregressive models, see the pioneering work by Tong and Lim (1980) and a recent revisit by Tong (2011). The second is to introduce dependence between the two AR processes. We shall pursue these interesting and promising extensions in a future project.

The rest of the paper consists of five sections. Section 2 elaborates on the bivariate AR risk model, Section 3 explains our idea of the hybrid method in a general setting, Section 4 first derives an asymptotic estimate for the finite-time ruin probability and then applies the hybrid method to improve the computational accuracy and efficiency, Section 5 is devoted to an application of our result and method to a portfolio optimization problem, and Section 6 is composed of proofs.

2. The Bivariate Autoregressive Risk Model

Consider the following discrete-time insurance risk model. Denote the claim amount (plus other expenses) paid by an insurer within period $i$ by a nonnegative random variable $B_i$, $i \in \mathbb{N}$, and assume that these claim amounts form an AR process of order 1, written as AR(1). Formally, starting with a deterministic value $B_0 \geq 0$, they evolve according to

$$B_i = \rho B_{i-1} + X_i, \quad i \in \mathbb{N},$$

where the autoregressive coefficient $\rho$ takes value in $[0, 1)$ and the innovations $X_i$, $i \in \mathbb{N}$, are i.i.d. copies of a nonnegative random variable $X$. Denote by $\mu_X$ the mean of $X$ and by $\mu_B$ the mean of the
stationary solution $B_{m}$. An advantage of the AR(1) model is that it can capture asymptotic dependence between claim amounts; see Proposition 6.1.

For simplicity, we assume that the claims are paid out at the end of each period and that the premium amounts during these periods are a deterministic constant $a > 0$. Then the one-period net losses are $B_{i} - a$, $i \in \mathbb{N}$. In insurance practice, premiums are usually collected at the beginning of each period. We have ignored insignificant value changes due to investments on premiums.

Note that in (2.1), to ensure the nonnegativity of the claim amounts, the autoregressive coefficient $\rho$ is not allowed to be negative. This is a slight flaw of the AR(1) process (2.1) from the modeling point of view despite the fact that a negative autoregressive coefficient is not usual for claim amounts. See Yang and Zhang (2003) for a similar remark. Therefore, it seems more reasonable to assume an autoregressive structure directly for real-valued net loss variables $\tilde{B}_{i}$, $i \in \mathbb{N}$, that is,

$$\tilde{B}_{i} - \mu_{i} = \rho(\tilde{B}_{i-1} - \mu_{i}) + \tilde{X}_{i}, \quad i \in \mathbb{N},$$

(2.2)

where the autoregressive coefficient $\rho$ takes value in $(-1, 1)$, the innovations $\tilde{X}_{i}$, $i \in \mathbb{N}$, are i.i.d. real-valued random variables with mean 0, and the constant $\mu_{i}$ is the mean of the stationary solution $\tilde{B}_{i}$. In this paper we follow the style of (2.1) to model claim amounts, but we remark that our study can be easily extended to the setting of (2.2).

Suppose that there is a discrete-time financial market consisting of a risk-free bond with a deterministic continuously compounded rate of interest $r > 0$ and a risky stock with a stochastic log-return rate $R_{i} \in \mathbb{R}$ during period $i$, $i \in \mathbb{N}$. These log-return rates are also assumed to follow an AR(1) process. Formally, starting with a deterministic value $R_{0}$, they evolve according to

$$(R_{i} - \mu_{i}) = \gamma(R_{i-1} - \mu_{i}) + Y_{i}, \quad i \in \mathbb{N},$$

(2.3)

or, equivalently,

$$R_{i} = d + \gamma R_{i-1} + Y_{i}, \quad i \in \mathbb{N},$$

(2.4)

where the autoregressive coefficient $\gamma$ takes value in $(-1, 1)$, the innovations $Y_{i}$, $i \in \mathbb{N}$, are i.i.d. copies of a real-valued random variable $Y$ with mean 0, the constant $\mu_{i}$ is the mean of the stationary solution $R_{n}$, and the constant $d$ is equal to $(1 - \gamma)\mu_{i}$. Assume that $\{X, X_{i}, i \in \mathbb{N}\}$ and $\{Y, Y_{i}, i \in \mathbb{N}\}$ are mutually independent, and, hence, so are the two AR(1) processes (2.1) and (2.3).

Suppose that, at the beginning of each period $i$, the insurer invests a fixed proportion $\pi \in [0, 1]$ of its current wealth in the stock and keeps the rest in the bond. We remark that the proportion $\pi$ is assumed to be fixed just for ease of presentation and the extension to a time-dependent investment is rather straightforward. Denote by $W_{m}$ the insurer’s wealth at time $m \in \mathbb{N}$. Then with a deterministic initial value $W_{0} = x > 0$, the wealth process $\{W_{m}, m \in \mathbb{N}\}$ evolves according to

$$W_{m} = ((1 - \pi)e^{r} + \pi e^{R_{m}})W_{m-1} - (B_{m} - a), \quad m \in \mathbb{N}.$$  

(2.5)

Iterating (2.5) yields

$$W_{m} = x \prod_{j=1}^{m} ((1 - \pi)e^{r} + \pi e^{R_{j}}) - \sum_{i=1}^{m} (B_{i} - a) \prod_{j=i+1}^{m} ((1 - \pi)e^{r} + \pi e^{R_{j}}),$$

(2.6)

where, and throughout this paper, multiplication over an empty index set produces value 1 by convention. Note that investments with a time-invariant $\pi \in [0, 1]$ often appear in continuous-time finance; see, for instance, Merton (1990).

Let $n \in \mathbb{N}$ be the time horizon of interest. As usual, the probability of ruin by time $n$ is defined as

$$\psi(x; n) = \Pr \left( \min_{0 \leq m \leq n} W_{m} < 0 | W_{0} = x \right).$$

To comply with certain risk reserve regulations such as EU Solvency II, the insurer has to hold enough risk reserve, the so-called Solvency Capital Requirement, so that it is able to meet its future obligations
with a high probability. This makes the notion of ruin theory particularly relevant. Mathematically, the risk reserve $x$ should be large enough so that the ruin probability satisfies

$$\psi(x; n) \leq 1 - q$$  \hspace{1cm} (2.7)$$

for some $0 < q < 1$ close to 1, for instance, $q = 0.995$. Therefore, an accurate estimate for the ruin probability is crucially important for determining the Solvency Capital Requirement.

Introduce

$$\delta = (1 - \pi)e^r \quad \text{and} \quad \theta_i = \prod_{j=1}^{i} \frac{1}{\delta + \pi e^{B_j}}, \quad i \in \mathbb{N}, \hspace{1cm} (2.8)$$

where $\theta_i$ can be interpreted as the overall stochastic discount factor over the time interval $(0, i]$. Then we have

$$\psi(x; n) = \Pr \left( \min_{0 \leq m \leq n} \left( x \prod_{j=1}^{m} (\delta + \pi e^{B_j}) - \sum_{i=1}^{m} (B_i - \alpha) \prod_{j=i+1}^{m} (\delta + \pi e^{B_j}) \right) < 0 \right)$$

$$= \Pr \left( \min_{0 \leq m \leq n} \left( x - \sum_{i=1}^{m} \theta_i (B_i - \alpha) \right) < 0 \right)$$

$$= \Pr \left( \max_{1 \leq m \leq n} \sum_{i=1}^{m} \theta_i (B_i - \alpha) > x \right). \hspace{1cm} (2.9)$$

At the core of this work is a hybrid method that we create to compute the ruin probability $\psi(x; n)$, as is shown in the next two sections.

### 3. The Hybrid Method

In this section we expound our idea of the hybrid method in a general setting. Suppose that we are to estimate

$$u(x) = \mathbb{E}[U(x)], \quad x \in \mathbb{R},$$

where $U(x)$ is a stochastic quantity presumably associated with a rare event so that $u(x) \to 0$ as $x \to \infty$. A prototype of $u(x)$ is the tail probability $\Pr(L > x)$ of a loss variable $L$, where $U(x)$ corresponds to $1_{(L > x)}$, the indicator function of $(L > x)$. Such a quantity is important in risk management because most risk measures, such as Value-at-Risk and Conditional Tail Expectation, target the tail area of the underlying risk variable. Because of the complex stochastic structure of $U(x)$, very often it is not possible to obtain the exact value of $u(x)$, and an estimate will be needed.

A common way to estimate $u(x)$ is to use the crude Monte Carlo (CMC) simulation, which gives an estimate as

$$u_1(x) = \frac{1}{N} \sum_{k=1}^{N} U_k(x),$$

where $U_k(x), k = 1, \ldots, N,$ are i.i.d. samples from $U(x)$ and $N$ is the sample size. The CMC simulation is indeed a powerful method prevailing in various computational problems in insurance and finance. However, taking the cost of simulation into consideration, we find that the estimate $u_1(x)$ in many cases is efficient only when the true value of $u(x)$ is not too small, or, equivalently, when $x$ is not too large.

Let us demonstrate this by considering how many samples are needed for $u_1(x)$ to achieve a required accuracy. In statistical terms, accuracy is usually characterized as follows. For some $h > 0$ close to 0 and $0 < p < 1$ close to 1, the simulated estimate $u_1(x)$ is within $100h\%$ of the true value $u(x)$ with probability not smaller than $p$; that is,
Throughout this paper, we always choose \( h = 0.05 \) and \( p = 0.95 \).

Consider again the example of estimating the tail probability \( u(x) = \Pr(L > x) \) of a loss variable \( L \). By the central limit theorem, approximately (3.1) means that

\[
\frac{h u(x)}{\sqrt{\frac{1}{N} u(x)(1 - u(x))}} \geq \varphi_p,
\]

or, equivalently,

\[
N \geq \left( \frac{\varphi_p}{h} \right)^2 \frac{1 - u(x)}{u(x)},
\]

where \( \varphi_p = \Phi^{-1}((1 + p)/2) \) is the quantile of the standard normal distribution at \((1 + p)/2\). Thus, the sample size \( N \) needed for the required accuracy depends heavily on the value of \( u(x) \). To offset the negative effect of \( u(x) \) being extremely small, which often arises in rare event simulation, the sample size \( N \) has to be extremely large. For instance, for \( u(x) = 10^{-6} \), according to (3.3) we need as many as \( 1.5 \times 10^9 \) samples to obtain the required accuracy. For simulations of such a scale, both computational time and memory allocation will be big issues. Related discussions can be found in Asmussen and Glynn (2007).

Another question to ask is, given an affordable sample size \( N \), for what values of \( u(x) = \Pr(L > x) \) the simulated estimate \( u_1(x) \) can attain the required accuracy. This can be answered also by the formulation in (3.2). Straightforwardly, the range for \( u(x) \) is

\[
u(x) \geq \left( N \left( \frac{h}{\varphi_p} \right)^2 + 1 \right)^{-1} = u_0.
\]

As an example, for \( N = 500,000 \) the simulated estimate \( u_1(x) \) meets the required accuracy if \( u(x) \geq 0.003 \), but it does not if \( u(x) < 0.003 \).

Whenever the CMC method breaks down for small values of \( u(x) \), or, equivalently, for large values of \( x \), the asymptotic method emerges naturally as an alternative. There are variance reduction methods, such as importance sampling and antithetic variables, which can help improve the accuracy of MC estimation. In this paper, however, our focus is on the asymptotic method. By this method, one can usually hope for a transparent analytical formula \( u_2(x) \) such that

\[
\lim_{x \to \infty} \frac{u_2(x)}{u(x)} = 1.
\]

This means that \( u_2(x) \) will be an accurate estimate for \( u(x) \) for large values of \( x \). In this paper we are satisfied with \( |u_2(x)/u(x) - 1| \leq 0.05 \).

Both estimates \( u_1(x) \) and \( u_2(x) \) have their advantages and disadvantages. The simulated estimate \( u_1(x) \) may meet the accuracy requirement only for relatively large values of \( u(x) \). On the contrary, the asymptotic estimate \( u_2(x) \) performs satisfactorily only for small values of \( u(x) \). An overall assessment suggests that we should make a compromise between \( u_1(x) \) and \( u_2(x) \). Once \( u_1(x) \) breaks down for small values of \( u(x) \), we seek a threshold \( u_0 \in (0, 1) \) such that \( u_2(x) \) starts to achieve the required accuracy when \( u(x) < u_0 \). Based on this threshold \( u_0 \), we decide the sample size \( N \) through (3.3) so that \( u_1(x) \) achieves the required accuracy for \( u(x) \geq u_0 \). Then it is natural to use \( u_1(x) \) when \( u(x) \geq u_0 \) and use \( u_2(x) \) otherwise. In our practice, the condition \( u(x) \geq u_0 \) is checked through the value of \( u_1(x) \).

We define the hybrid estimate as

\[
\bar{u}(x) = u_1(x)1_{(u_1(x) \geq u_0)} + u_2(x)1_{(u_1(x) < u_0)}.
\]
Of vital importance for this method to work is the availability of an asymptotic estimate \( u_\alpha(x) \) for \( u(x) \) that starts to work efficiently when \( u(x) \) is moderately small. This is usually feasible for heavy-tailed cases. In such cases, losses occur with a relatively small probability but are typically accompanied by disastrous consequences, which we often call rare events, or, more fashionably, black swan events.

In the next section, we use estimation of the ruin probability as an example to illustrate the procedure of the hybrid method. Nonetheless we would like to stress that the idea behind should work for various estimation problems involving asymptotics.

4. A Hybrid Estimate for the Ruin Probability

Hereafter all limit relationships are according to \( x \to \infty \) unless otherwise stated. For two positive functions \( f(\cdot) \) and \( g(\cdot) \), we write \( f(x) \sim g(x) \) if \( \lim f(x)/g(x) = 1 \) and write \( f(x) \preceq g(x) \) if \( \lim \sup f(x)/g(x) \leq 1 \). For two positive bivariate functions \( f(\cdot, \cdot) \) and \( g(\cdot, \cdot) \), we say that the asymptotic relation \( f(x, y) \sim g(x, y) \) holds uniformly for \( y \) in a nonempty set \( \Delta \) if

\[
\lim \sup_{x \to \infty} \sup_{y \in \Delta} \left| \frac{f(x, y)}{g(x, y)} - 1 \right| = 0.
\]

Let us return to the bivariate AR risk model introduced in Section 2, in which the claim amounts follow the AR(1) process (2.1) and the log-return rates of the stock follow the AR(1) process (2.3).

4.1 The Crude Monte Carlo Estimate

For the ruin probability \( \psi(x; n) \) given by (2.9), the CMC estimate is

\[
\psi_1(x; n) = \frac{1}{N} \sum_{k=1}^{N} l(x_k > x),
\]

where \( L_k, k = 1, \ldots, N \), are i.i.d. samples from

\[
L = \max_{1 \leq m \leq n} \sum_{i=1}^{m} \theta_i (B_i - \alpha).
\]

As stated in Section 3, the CMC method works only if \( \psi(x; n) \) is not too small. For instance, suppose that \( N = 500,000 \) is the maximal sample size one is willing to work with. Then by (3.4), the threshold \( u_0 \) is 0.003. This means that the estimate \( \psi_1(x; n) \) given by (4.1) is accurate only if \( \psi(x; n) \geq 0.003 \) (practically, \( \psi_1(x; n) \geq 0.003 \)). For large values of \( x \) such that \( \psi(x; n) < 0.003 \), we shall rely on the asymptotic estimate constructed in the next subsection.

4.2 The Asymptotic Estimate

Now assume that the common distribution function \( F \) of the innovations in the AR(1) process (2.1) is heavy tailed. More specifically, there is some \( \alpha > 0 \) such that the relation

\[
\lim_{x \to \infty} \frac{F(xy)}{F(x)} = y^{-\alpha}
\]

(4.2)

holds for all \( y > 0 \). Denote by \( \mathcal{R}_{-\alpha} \) the class of all distribution functions \( F \) on \( \mathbb{R} \) satisfying the regular variation property described by relation (4.2). The definition immediately indicates that \( F \in \mathcal{R}_{-\alpha} \) if and only if \( \tilde{F}(\cdot) \) can be expressed as

\[
\tilde{F}(x) = x^{-\alpha} l(x), \quad x > 0,
\]

(4.3)

for some slowly varying function \( l(\cdot) \). Hence, a smaller value of \( \alpha \) means a heavier right tail of \( F \). This class contains many popular heavy-tailed distributions such as Pareto, log gamma, Burr and Student’s \( t \) distributions, and it has been extensively used to model heavy-tail phenomena in insurance and fi-
nance. For background knowledge of regular variation, the reader is referred to Bingham et al. (1987), Resnick (1987, 2007), and Embrechts et al. (1997).

Throughout the paper, we write $M_Y(s) = E[e^{st}]$, $s \in \mathbb{R}$, as the moment-generating function of the generic innovation $Y$ of the AR(1) process (2.3), and define $0^+$ to be 1 as usual. The asymptotic formula obtained by the following result will be used to construct the desired asymptotic estimate for the finite-time ruin probability.

**Theorem 4.1**

Consider the bivariate AR risk model introduced in Section 2. Assume that $F \in \mathcal{R}_{-\alpha}$ for some $\alpha > 0$. Then for every $n \in \mathbb{N}$, the relation

$$
\psi(x; n) \sim E\left[ \tilde{F} \left( x + \sum_{i=1}^{n} \theta_i (x - \rho B) \right) \sum_{j=1}^{n} \left( \sum_{i\neq j} \theta_i \rho^{i-j} \right)^\alpha \right]
$$

holds under either of the following two conditions:

a. $\pi \in [0, 1]$;

b. $\pi = 1$ and $M_Y(s_1) < \infty$ for some $s_1 < -2\alpha(1 \vee (1 - \gamma)^{-1})$.

**Proof**

Let us first verify the conditions of Lemma 6.4. If $\pi \in [0, 1)$, then each $\theta_i$ defined by (2.8) is bounded by $e^{-\pi i} < \infty$. If $\pi = 1$, then for each $i \in \mathbb{N}$, by (2.4) there is some constant $c_i > 0$ such that

$$
\theta_i = \exp \left\{ -\sum_{j=1}^{\infty} R_j \right\} = c_i \exp \left\{ -\sum_{j=1}^{\infty} \frac{1 - \gamma^j}{1 - \gamma} Y_j \right\}.
$$

By the moment condition on $Y$, we have $E[\theta_i^{\beta}] < \infty$ for $\beta = s_1/2$ $(1 \wedge (1 - \gamma)) > \alpha$. Thus, Lemmas 6.4 and 6.5 are applicable for both cases. Applying these two lemmas we complete the proof of Theorem 4.1.

The asymptotic formula (4.4) can certainly be simplified to

$$
\psi(x; n) \sim \tilde{F}(x) \sum_{j=1}^{n} E \left[ \left( \sum_{i\neq j} \theta_i \rho^{i-j} \right)^\alpha \right].
$$

However, this completely ignores the sum $\sum_{i=1}^{\infty} \theta_i (x - \rho B)$, which is an indispensable term in the expression for $W_n$ as shown in our derivation of (6.11). Hence, the simplified version would be less accurate than (4.4), as remarked after Lemma 6.3.

The expectation on the right-hand side of (4.4) is still estimated by the CMC method; that is,

$$
\psi_2(x; n) = \frac{1}{N} \sum_{k=1}^{N} \tilde{F}(x + \Sigma_{0k}) \Sigma_k,
$$

where $(\Sigma_{0k}, \Sigma_k)$, $k = 1, \ldots, N$, are i.i.d. samples from

$$(\Sigma_0, \Sigma) = \left( \sum_{i=1}^{n} \theta_i (x - \rho B), \sum_{j=1}^{n} \left( \sum_{i\neq j} \theta_i \rho^{i-j} \right)^\alpha \right).$$

However, we would like to remark that this step of estimation incurs little extra error, as explained below. Consider the relative error of $\psi_2(x; n)$ defined by

$$
\sqrt{\frac{\text{var} \{ \psi_2(x; n) \}}{\text{E}[\psi_2(x; n)]^2}} = \frac{1}{\sqrt{N}} \sqrt{\frac{\text{E}[\tilde{F}(x + \Sigma_0)\Sigma]^2}{\text{E}[\tilde{F}(x + \Sigma_0)\Sigma]^2} - 1}.
$$
Let the conditions of Theorem 4.1 be valid except that the moment condition on \( Y \) in (b) is strengthened to \( M_Y(s_1) < \infty \) for some \( s_1 < -4\alpha \sqrt{(1 \vee (1 - \gamma)^{-1})} \). It is easy to verify all the conditions of Lemma 6.6, by which we have
\[
\lim_{n \to \infty} \sqrt{\frac{\text{var} [\psi_2(x; n)]}{(\text{E}[\psi_2(x; n)])^2}} = \lim_{n \to \infty} \frac{1}{N} \sqrt{\frac{(\bar{F}(x))^2\text{E}[\Sigma]^{2}}{(\bar{F}(x))^2(\text{E}[\Sigma])^2}} - 1 = \frac{1}{N} \sqrt{\frac{\text{var} [\Sigma]}{\text{E}[\Sigma]}} < \infty.
\]
Thus, the CMC estimate \( \psi_2(x; n) \) has a bounded relative error, which is usually the best that one can desire for in rare-event simulation.

### 4.3 Accuracy of the Asymptotic Estimate

To answer whether or not \( \psi_2(x; n) \) given by (4.5) is a good estimate for \( \psi(x; n) \), we carry out numerical studies to examine its accuracy by checking how close to 1 the ratio \( \psi_2(x; n)/\psi_1(x; n) \) is. Throughout the rest of this paper, all numerical studies are conducted in the R software environment. For a reference manual on R, see R Development Core Team (2008).

For ease of computation, we assume that the generic innovation \( X \) of the AR(1) process (2.1) follows a Pareto distribution with shape parameter \( \alpha > 1 \) and mean \( \mu_X \), and that the generic innovation \( Y \) of the AR(1) process (2.3) follows a normal distribution with mean 0 and variance \( \sigma^2 \).

In Figures 1(a)–(c), the parameters are set to \( \pi = 0.2, 0.5, \) or 0.8, \( n = 4, \alpha = 1.1, 1.3, \) or 1.5, \( \rho = 0.3 \) or 0.5, \( B_0 = 5.0, \mu_B = 5.0, \mu_X = (1 - \rho)\mu_B, \alpha = 5.5, r = 1.242\% \) (so that \( e^r - 1 = 1.25\%) \), \( \gamma = 0.5 \) or 0.8, \( R_0 = 1.5\%, \mu_R = 1.5\%, \) and \( \sigma = 0.2 \). By setting \( n = 4 \) we mean that each period is a season. The parameters \( r, R_0, \mu_R, \) and \( \sigma \) are chosen so that the annual risk-free rate and the annual return rates are within a reasonable range. The values of \( \mu_B \) and \( \alpha \) are chosen to yield a safety loading of 10%. The sample size is \( N = 1,000,000 \) for both \( \psi_1(x; n) \) and \( \psi_2(x; n) \) despite the fact that it is more than enough for \( \psi_2(x; n) \). We compare the two estimates on the left and show their ratio on the right.

From these graphs we see that, for each set of parameters, the ratio \( \psi_2(x; n)/\psi_1(x; n) \) quickly converges to 1 as the initial wealth \( x \) increases. After staying around 1 for a while, the ratio starts to fluctuate. An explanation of the fluctuation is that, as \( x \) becomes large, the true value of \( \psi_2(x; n) \) becomes too small, leading to too low of an accuracy of \( \psi_1(x; n) \). In other words, the fluctuation may result only from the low accuracy of \( \psi_1(x; n) \). Notice that the larger the parameter \( \alpha \) is, the smaller the ruin probability \( \psi(x; n) \) becomes and the more fluctuation the ratio \( \psi_2(x; n)/\psi_1(x; n) \) exhibits. Moreover, the convergence of the ratio is robust with respect to \( \pi, \rho, \) and \( \gamma \).

In the far right area, the ruin probability \( \psi(x; n) \) decreases quickly as \( \alpha \) increases. Hence, not surprisingly, a larger value of \( \alpha \) leads to a less accurate \( \psi_1(x; n) \). For instance, let us look at Figure 2(a), for which the parameters are set identical to those in Figures 1(a)–(c) except \( \pi = 0.5, \alpha = 2.0, \rho = 0.3, \) and \( \gamma = 0.8 \). In this graph the ratio \( \psi_2(x; n)/\psi_1(x; n) \) fluctuates dramatically and does not seem to converge to 1 at all. One might start to doubt the validity of the asymptotic estimate.

Actually, this is an illusion because of the poor performance of the simulated estimate \( \psi_1(x; n) \). We repeat the simulation with the sample size \( N \) increased from 1,000,000 to 10,000,000 and draw Figure 2(b). Then we observe a much improved convergence of the ratio \( \psi_2(x; n)/\psi_1(x; n) \). This confirms that the only source of fluctuation is the awful estimate \( \psi_1(x; n) \) in the far right area. There is nothing wrong with the asymptotic estimate \( \psi_2(x; n) \). For a large value of \( \alpha \), the sample size \( N \) usually has to be very large to make the simulated estimate \( \psi_1(x; n) \) acceptable.

### 4.4 The Hybrid Estimate for the Ruin Probability

Now that a good asymptotic estimate for \( \psi(x; n) \) is available, we explain in detail how the hybrid method introduced in Section 3 can be implemented in the current situation.

In our practice, with a reasonably large sample size \( N^* \), we explore the value of \( \psi_1(x; n) \) for which the ratio \( \psi_2(x; n)/\psi_1(x; n) \) starts to fall into the strip between 0.95 and 1.05. Such a value serves as the threshold \( u_0 \) in our hybrid method. Note that, for a fixed \( \alpha \), the threshold \( u_0 \) is determined once
Figure 1

Accuracy of the Asymptotic Estimate $\psi_2$: (a) for $\pi = 0.2$, (b) for $\pi = 0.5$, (c) for $\pi = 0.8$
and for all owing to the robustness of the asymptotic estimate $\psi_2(x; n)$ with respect to the parameters $\pi$, $\rho$, and $\gamma$. In our numerical study, with $N^\circ = 1,000,000$, $\pi = 0.2$, $0.5$ or $0.8$, $n = 4$, $\rho = 0.3$ or $0.5$, $B_0 = 5.0$, $\mu_R = 5.0$, $\mu_X = (1 - \rho)\mu_R$, $\alpha = 5.5$, $r = 1.242\%$, $\gamma = 0.5$ or $0.8$, $R_0 = 1.5\%$, $\mu_R = 1.5\%$, and $\sigma = 0.2$, we see that the threshold $u_0$ can be chosen to be $u_0 = 0.01$ for $\alpha = 1.1$, $u_0 = 0.003$ for $\alpha = 1.3$, $u_0 = 0.003$ for $\alpha = 1.5$, and $u_0 = 0.0005$ for $\alpha = 2.0$.

Next, we calibrate the sample size $N$ according to (3.3) with $u(x)$ replaced by $u_0$. For instance, in our numerical study, the sample size needed for the required accuracy of $\psi_1(x; n)$ is about $N = 140,000$ for $\alpha = 1.1$, $N = 500,000$ for $\alpha = 1.3$, $N = 500,000$ for $\alpha = 1.5$, and $N = 3,000,000$ for $\alpha = 2.0$. We compute $\psi_1(x; n)$ and $\psi_2(x; n)$ again based on the calibrated sample size $N$.

Finally, define
\[
\tilde{\psi}(x; n) = \psi_1(x; n)1_{(\phi_1(x; n) \geq u_0)} + \psi_2(x; n)1_{(\phi_1(x; n) < u_0)},
\]
which is the desired hybrid estimate for $\psi(x; n)$. It is easily understood that $\tilde{\psi}(x; n)$ gives a globally accurate estimation for $\psi(x; n)$.
5. Portfolio Optimization

Generally speaking, investors aim at maximizing gains while minimizing risks. How to balance between gains and risks is a question that permeates many areas of finance. Since the seminal work of Merton (1971), many variants of Merton’s problem have been proposed in the finance literature to address different interests and concerns.

Let us now restrict ourselves to portfolio optimization problems in the insurance context. Due to the increasing prudence of insurance regulations, a solvency constraint needs to be imposed in portfolio optimization problems. Such a constraint is relevant in practice, especially after the 2008 financial crisis. Related discussions on portfolio optimization with a solvency constraint can be found in Paulsen (2003), Irgens and Paulsen (2005), Dickson and Drekic (2006), Kostadinova (2007), He et al. (2008), and Liang and Huang (2011), among others.

Consider the bivariate AR risk model introduced in Section 2. We now aim to determine the proportion invested in the stock that maximizes the expected terminal wealth subject to a solvency constraint. We stick to all conditions of Theorem 4.1 but require \( \alpha > 1 \) and strengthen the moment condition on \( Y \) to \( M_{s_1}(s_2) < \infty \) for some \( s_1 < -2\alpha (1 \vee (1 - \gamma)^{-1}) \) and \( M_{s_1}(s_2) < \infty \) for \( s_2 = 1 \vee (1 - \gamma)^{-1} \) so as to ensure the finiteness of the expected terminal wealth. By the convexity of \( M(\cdot) \), it follows that \( M_{s_1}(s) < \infty \) for \( s \in [s_1, s_2] \).

By relation (2.6), the terminal wealth \( W_n \) is

\[
W_n = x \prod_{j=1}^{n} (\delta + \pi e^{R_j}) - \sum_{i=1}^{n} (B_i - a) \prod_{j=i+1}^{n} (\delta + \pi e^{R_j}).
\]

Our goal is to determine a value of \( \pi \in [0, 1] \) that maximizes \( E[W_n] \) subject to the constraint on \( \psi(x; n) \) given by (2.7), namely, \( \psi(x; n) \leq 1 - q \) for some \( 0 < q < 1 \) close to 1.

We employ a recursive procedure to compute \( E[W_n] \). Write

\[
I(s, i, n) = E \left[ e^{\pi R_n} \prod_{j=i+1}^{n} (\delta + \pi e^{R_j}) \right], \quad s \in \mathbb{R}, i = 0, 1, \ldots, n,
\]

so that

\[
E[W_n] = xI(0, 0, n) - \sum_{i=1}^{n} (E[B_i] - a)I(0, i, n).
\]

The expectations \( E[B_i], i = 1, \ldots, n, \) are computed by using (6.2). The terms \( I(s, i, n) \) for \( i = 0, 1, \ldots, n \) can be computed recursively. Denote by \( \mathcal{F}_0 \) the trivial \( \sigma \)-field specifying the initial values \( B_0, R_0, \) and \( W_0, \) and by \( \mathcal{F}_n \) the \( \sigma \)-field generated by \( (B_1, R_1), \ldots, (B_m, R_m) \). Thus, \( \{\mathcal{F}_0, \mathcal{F}_n, m \in \mathbb{N}\} \) serves as the natural filtration of the bivariate AR model. For \( i = 0, 1, \ldots, n - 1, \) by conditioning on \( \mathcal{F}_{n-1} \) and using (2.4), we have

\[
I(s, i, n) = \delta e^{\pi R_{n-1}} I(s, i, n - 1) + \pi e^{\pi R_{n-1}} M_{s_1}(s_2) + \pi e^{\pi R_{n-1}} M_{s_1}(s_2) I(s, i, n - 1).
\]

(5.1)
Continue this recursive procedure until the third argument decreases from $n$ to $i$. We see that the computation of $I(s, i, n)$ is reduced to that of $I(s, i, i) = \mathbb{E}[e^{sR}]$ for $i = 0, 1, \ldots, n$. Similarly, the latter can be computed recursively through

$$I(s, 0, 0) = e^{sR_0}, \quad I(s, i, i) = e^{s\mu}M_Y(s)I(s\gamma, i - 1, i - 1), \quad i = 1, \ldots, n.$$ 

We remark that, for computing $\mathbb{E}[W_n]$, it suffices to set the argument $s$ in the first step of (5.1) to 0. In this case the moment-generating function of $Y$ appearing in these recursive formulas is always finite.

In our numerical study, the same as in Subsection 4.3, we assume for ease of computation that the generic innovation $X$ of the AR(1) process (2.1) follows a Pareto distribution with shape parameter $\alpha > 1$ and mean $\mu_X$, and that the generic innovation $Y$ of the AR(1) process (2.3) follows a normal distribution with mean 0 and variance $\sigma^2$.

We determine the optimal value of $\pi$ by a grid search over the interval $[0, 1]$. First, we find a region, called the admissible region, of $\pi$ satisfying the constraint described by (2.7), where the ruin probability is computed according to the hybrid estimate (4.6). Then we compute $\mathbb{E}[W_n]$ for each $\pi$ in the admissible region using the recursive formulas above. The optimal $\pi$ is the one that corresponds to the maximal value of $\mathbb{E}[W_n]$.

The optimization algorithm is also implemented in R, but with C++ integrated to speed up. The powerful R package Rcpp makes the integration seamless. As stated by Eddelbuettel and François (2011), Rcpp defines a class hierarchy and maps every R type (such as vector, function, and environment) to a dedicated class. For instance, numeric vectors are represented as instances of the Rcpp::NumericVector class. Data passing between R and C++, which used to be cumbersome, is now made surprisingly straightforward by the highly flexible templates Rcpp::wrap and Rcpp::as. For more detailed descriptions on Rcpp, the reader is referred to Eddelbuettel and François (2011). The package RcppArmadillo is an integration of R and the templated C++ linear algebra library Armadillo. See Sanderson (2010) and François et al. (2011) for more details of Armadillo and RcppArmadillo, respectively. For the purpose of integrating R and C++, the package inline can be a good assistant as it enables one to define R functions dynamically with inlined C++ code. For instance, in our numerical study, an R function that takes certain parameters and returns an optimal value of $\pi$ is defined via the function cxxfunction in the package inline. See Sklyar et al. (2010) for more details of inline.

---

**Figure 3**

**Optimal $\pi$ for Different Values of the Initial Wealth**

---
Figure 3 shows how the optimal \( \pi \) changes with respect to the initial wealth \( x \). The parameters are set to \( n = 4, \alpha = 1.1, 1.5, \) or \( 2.0, \rho = 0.3, B_0 = 5.0, \mu_\mu = 5.0, \mu_x = (1 - \rho)\mu_\mu, \alpha = 5.5, r = 1.242\%, \gamma = 0.8, R_0 = 1.5\%, \mu_R = 1.5\% \) or \( 3\% \), and \( \sigma = 0.2 \), most being the same as in Subsection 4.4. According to the threshold \( u_0 \) determined in Subsection 4.4, we set \( N = 200,000 \) for \( \alpha = 1.1, N = 500,000 \) for \( \alpha = 1.5, \) and \( N = 5,000,000 \) for \( \alpha = 2.0 \), with \( N \) increasing to appropriately offset the negative effect of a small value of the ruin probability. It shows a clear trend of the optimal \( \pi \) increasing in \( x \) and \( \mu_R \), all being consistent with our anticipation.

6. PROOFS

First, let us collect some useful preliminaries on the distribution class \( R_{-\alpha} \). The first two below are known as max-sum equivalence and Potter’s bounds, respectively.

**Lemma 6.1**

Let \( F, F_1, \) and \( F_2 \) be three distribution functions on \( \mathbb{R} \), all belonging to the class \( R_{-\alpha} \) for some \( \alpha > 0 \). We have the following:

a. The convolution \( F_1 \ast F_2 \) still belongs to the class \( R_{-\alpha} \) and \( F_1 \ast F_2(x) = F_1(x) + F_2(x) \).

b. For every \( \varepsilon > 0 \), there is some \( c_0 > 0 \) such that, for all \( x, y \geq x_0 \),

\[
(1 - \varepsilon)((x/y)^{a+\varepsilon} \land (x/y)^{a-\varepsilon}) \leq \frac{F(y)}{F(x)} \leq (1 + \varepsilon)((x/y)^{a+\varepsilon} \lor (x/y)^{a-\varepsilon}).
\]

c. The relations \( x^{-a-\varepsilon} = o(F(x)) \) and \( F(x) = o(x^{-a+\varepsilon}) \) hold for every \( \varepsilon > 0 \).

d. For every \( C > 1 \), the relation \( F(xy) \sim y^{-a}F(x) \) holds uniformly for \( y \in [1/C, C] \).

**Proof**

a. See Lemma 1.3.1 of Embrechts et al. (1997).

b. See Theorem 1.5.6 of Bingham et al. (1987).

c. This can be proved by applying Proposition 1.3.6 of Bingham et al. (1987) to the representation for \( F(x) \) given by (4.3).

d. By Theorem 1.5.2 of Bingham et al. (1987), the convergence in (4.2) holds uniformly for \( y \in [1/C, \infty) \), namely,

\[
\lim_{x \to \infty} \sup_{y \in [1/C, \infty]} \left| \frac{F(xy)}{F(x)} - y^{-a} \right| = 0.
\]

Then one easily sees the uniformity of the relation \( F(xy) \sim y^{-a}F(x) \) for \( y \in [1/C, C] \).

It is sometimes convenient to consider the following augmented version of the AR(1) process (2.1):

\[
B_i = \rho B_{i-1} + X_i, \quad i \in \mathbb{Z},
\]

(6.1)

where, the same as before, the autoregressive coefficient \( \rho \) takes value in \([0, 1)\) and the innovations \( X_i, i \in \mathbb{Z} \), are i.i.d. copies of a nonnegative random variable \( X \). The stationary solution of (6.1) is given by

\[
B_i = \sum_{j=0}^{\infty} \rho^j X_{i-j}, \quad i \in \mathbb{Z}.
\]

By plugging the starting value \( B_0 \) in the above we obtain

\[
B_i = \sum_{j=0}^{i-1} \rho^j X_{i-j} + \rho^i \sum_{j=0}^{\infty} \rho^j X_{i-j} \equiv \sum_{j=0}^{i-1} \rho^j X_{i-j} + \rho^i B_0, \quad i \in \mathbb{N}.
\]

(6.2)
Proposition 6.1

Consider the AR(1) process (2.1). If $F \in \mathcal{F}_{-\alpha}$ for some $\alpha > 0$, then the relation

$$\lim_{x \to \infty} \Pr(B_{i_2} > x | B_{i_1} > x) = \mu^{i_2-i_1}$$  \hspace{1cm} (6.3)$$

holds for all $i_1, i_2 \in \mathbb{N}$.

Proof

Relation (6.3) holds trivially for $\mu = 0$ or $i_1 = i_2$. Moreover, the proofs for $i_1 < i_2$ and $i_1 > i_2$ are similar. Hence, we prove relation (6.3) only for $0 < \mu < 1$ and $i_1 < i_2$.

Similar to (6.2), it holds that

$$B_{i_2} = \sum_{j=0}^{i_2-i_1-1} \mu^i X_{i_2-j} + \mu^{i_2-i_1} B_{i_1} = I_1 + I_2,$$

where $B_{i_1}$ is expressed by (6.2) with $i$ replaced by $i_1$. Note that the two terms $I_1$ and $I_2$ are independent. Applying Lemma 6.1(a) and extending by induction, we see that both $I_1$ and $I_2$ follow distributions belonging to the class $\mathcal{F}_{-\alpha}$.

Now we deal with the joint probability $\Pr(B_{i_2} > x, B_{i_1} > x)$. On the one hand, we have

$$\Pr(B_{i_2} > x, B_{i_1} > x) = \Pr(I_1 + I_2 > x, I_2 > \mu^{i_2-i_1} x)$$

$$\geq \Pr(I_1 + I_2 > x, I_2 > x)$$

$$= \Pr(I_2 > x).$$

(6.4)

On the other hand, still starting from (6.4) we have

$$\Pr(B_{i_2} > x, B_{i_1} > x) \leq \Pr(I_1 + I_2 > x) - \Pr(I_1 > x) \Pr(I_2 \leq \mu^{i_2-i_1} x)$$

$$\sim \Pr(I_1 > x) + \Pr(I_2 > x) - \Pr(I_1 > x)$$

$$= \Pr(I_2 > x),$$

where in the second step we used Lemma 6.1(a). It follows that $\Pr(B_{i_2} > x, B_{i_1} > x) \sim \Pr(\mu^{i_2-i_1} B_{i_1} > x)$. Therefore,

$$\lim_{x \to \infty} \Pr(B_{i_2} > x | B_{i_1} > x) = \lim_{x \to \infty} \frac{\Pr(\mu^{i_2-i_1} B_{i_1} > x)}{\Pr(B_{i_1} > x)} = \mu^{i_2-i_1}.$$

This proves relation (6.3) for $i_1 < i_2$.

The following elementary result is used in proving Lemma 6.3.

Lemma 6.2

Let $X$ be a random variable distributed by $F \in \mathcal{F}_{-\alpha}$ for some $\alpha > 0$, let $\vartheta$ be a positive random variable with $E[\vartheta^\beta] < \infty$ for some $\beta > \alpha$, let $\{\Delta_t, t \in \mathcal{T}\}$ be a set of random events satisfying $\lim_{t \to t_0} \Pr(\Delta_t) = 0$ for some $t_0$ in the closure of the index set $\mathcal{T}$, and let $\{\vartheta, \{\Delta_t, t \in \mathcal{T}\}\}$ and $X$ be independent. Then

$$\lim_{t \to t_0} \limsup_{x \to \infty} \frac{\Pr(\vartheta X > x, \Delta_t)}{F(x)} = 0.$$ 

Proof

Choose some $\alpha/\beta < c < 1$. Then split $\Pr(\vartheta X > x, \Delta_t)$ into two parts as

$$I_1(x, t) + I_2(x, t) = \Pr(\vartheta X > x, \Delta_t, 0 < \vartheta \leq \vartheta^c) + \Pr(\vartheta X > x, \Delta_t, \vartheta > \vartheta^c).$$
By Lemma 6.1(b), it holds for arbitrarily fixed $0 < \varepsilon < \alpha \land (\beta - \alpha)$ and all large $x$ that

$$I_1(x, t) = \int_0^{x+\varepsilon} \Pr(sX > x)\Pr(\tilde{\theta} \in ds, \Delta_t)$$

$$\leq (1 + \varepsilon)\bar{F}(x) \int_0^{x+\varepsilon} (s^{a+\varepsilon} \lor s^{a-\varepsilon})\Pr(\tilde{\theta} \in ds, \Delta_t)$$

$$\leq (1 + \varepsilon)\bar{F}(x)\mathbb{E}[((\tilde{\theta}^{a+\varepsilon} \lor \tilde{\theta}^{a-\varepsilon})1_{\Delta_t}].$$

By Markov’s inequality and Lemma 6.1(c), we have

$$I_2(x, t) \leq \Pr(\tilde{\theta} > x^c) \leq x^{-c}\mathbb{E}[\tilde{\theta}^\beta] = o(\bar{F}(x)).$$

Putting the two bounds together yields

$$\lim_{t \to t_0} \lim_{x \to \infty} \frac{\Pr(\tilde{\Theta}_X > x, \Delta_t)}{\bar{F}(x)} \leq (1 + \varepsilon) \lim_{t \to t_0} \mathbb{E}[((\tilde{\theta}^{a+\varepsilon} \lor \tilde{\theta}^{a-\varepsilon})1_{\Delta_t}] = 0,$$

where in the last step we used the dominated convergence theorem.

The following result plays a key role in establishing our main result, and it is interesting in its own right.

**Lemma 6.3**

Let $\{X_i, i \in \mathbb{N}\}$ be a sequence of i.i.d. random variables with common distribution function $F$ on $\mathbb{R}$, let $\{\tilde{\theta}_0, \tilde{\theta}_i, i \in \mathbb{N}\}$ be another sequence of not necessarily independent random variables with $\tilde{\theta}_0$ real valued and $\tilde{\theta}_i$ positive for $i \in \mathbb{N}$, and let the two sequences be independent. If $F \in \mathcal{R}_-\alpha$ for some $\alpha > 0$, $\Pr(\tilde{\theta}_0 \leq -x) = o(\bar{F}(x))$, $\mathbb{E}[\tilde{\theta}_0^\beta] < \infty$ for some $\beta > \alpha$ and $\mathbb{E}[\tilde{\theta}_i^\alpha 1_{(\tilde{\theta}_0, -x)}] = o(\bar{F}(x))$, $i \in \mathbb{N}$, then it holds for every $n \in \mathbb{N}$ that

$$\Pr\left(\sum_{i=1}^n \tilde{\theta}_i X_i - \tilde{\theta}_0 > x\right) \sim \mathbb{E}\left[\bar{F}(x + \tilde{\theta}_0) \sum_{i=1}^n \tilde{\theta}_i^\alpha\right] - \bar{F}(x) \sum_{i=1}^n \mathbb{E}[\tilde{\theta}_i^\alpha]. \quad (6.5)$$

Lemma 6.3 gives two asymptotics, $\mathbb{E}[\bar{F}(x + \tilde{\theta}_0) \sum_{i=1}^n \tilde{\theta}_i^\alpha]$ and $\bar{F}(x) \sum_{i=1}^n \mathbb{E}[\tilde{\theta}_i^\alpha]$. We remark that the former should work better for computation purposes because it captures the impact of $\tilde{\theta}_0$. For this reason, Lemma 6.4 and Theorem 4.1 are tied to the former rather than the latter.

**Proof**

The proof of the second relation in (6.5) is easy. Actually, for arbitrarily fixed $0 < c < 1$, we split $\mathbb{E}[\bar{F}(x + \tilde{\theta}_0) \sum_{i=1}^n \tilde{\theta}_i^\alpha]$ into two parts as

$$I_1(x) + I_2(x) = \mathbb{E}\left[\bar{F}(x + \tilde{\theta}_0) \sum_{i=1}^n \tilde{\theta}_i^\alpha 1_{(\tilde{\theta}_0, -cx)}\cup(\tilde{\theta}_0, c-x)\right].$$

By the condition $\mathbb{E}[\tilde{\theta}_i^\alpha 1_{(\tilde{\theta}_0, -cx)}] = o(\bar{F}(x))$ for $i \in \mathbb{N}$, it holds that

$$I_1(x) \leq \sum_{i=1}^n \mathbb{E}[\tilde{\theta}_i^\alpha 1_{(\tilde{\theta}_0, -cx)}] = o(\bar{F}(x)) = o(\bar{F}(x)).$$

For $I_2(x)$, noting that $\bar{F}(x + \tilde{\theta}_0) \leq \bar{F}((1 - c)x) \sim (1 - c)^{-c}\bar{F}(x)$, we have, by the dominated convergence theorem,

$$\lim_{x \to \infty} \frac{I_2(x)}{\bar{F}(x)} = \mathbb{E}\left[\lim_{x \to \infty} \frac{\bar{F}(x + \tilde{\theta}_0)}{\bar{F}(x)} 1_{(\tilde{\theta}_0, c-x)}\right] \sum_{i=1}^n \tilde{\theta}_i^\alpha = \sum_{i=1}^n \mathbb{E}[\tilde{\theta}_i^\alpha].$$

This proves the second relation in (6.5).
Now we turn to the first relation in (6.5). Write \( \overline{s} = (s_1, \ldots, s_n) \) and \( \overline{\vartheta} = (\vartheta_1, \ldots, \vartheta_n) \). Arbitrarily choose \( 0 < c < 1 \) and \( C > 1 \) such that \( \Pr(\overline{\vartheta} \in [1/C, C]^n) > 0 \). By Proposition 5.1 of Tang and Tsitsiashvili (2003b), the relation

\[
\Pr \left( \sum_{i=1}^{n} s_i X_i > x + s_0 \right) \sim \sum_{i=1}^{n} \Pr(s_i X_i > x + s_0)
\] (6.6)

holds uniformly for \( s_0 > -cx \) and \( \overline{s} \in [1/C, C]^n \). According to these constants \( c \) and \( C \), we split the space \( \mathbb{R}^{n+1} \) into three disjoint parts as

\[
\Delta_1 \cup \Delta_2 \cup \Delta_3 = (|s_0| > cx) \cup (|s_0| \leq cx, \overline{s} \notin [1/C, C]^n) \cup (|s_0| \leq cx, \overline{s} \in [1/C, C]^n).
\]

Then the tail probability \( \Pr(\sum_{i=1}^{n} \vartheta_i X_i - \vartheta_0 > x) \) is equal to

\[
\sum_{j=1}^{3} J_j(x) = \sum_{j=1}^{3} \Pr \left( \sum_{i=1}^{n} \vartheta_i X_i - \vartheta_0 > x, (\vartheta_0, \overline{\vartheta}) \in \Delta_j \right).
\] (6.7)

By Markov’s inequality and Lemma 6.2, it holds for arbitrarily fixed \( \varepsilon > 0 \) and all large \( x \) that

\[
J_1(x) \leq \Pr \left( \sum_{i=1}^{n} \vartheta_i X_i > x, \vartheta_0 > cx \right) + \Pr(\vartheta_0 \leq -cx)
\leq \sum_{i=1}^{n} \Pr \left( \vartheta_i X_i > \frac{x}{n}, \vartheta_0 > cx \right) + o(\tilde{F}(cx))
\leq \varepsilon \tilde{F}(x).
\] (6.8)

Similarly, by Lemma 6.2 again, it holds for all large \( x \) and \( C \) that

\[
J_2(x) \leq \sum_{i=1}^{n} \Pr \left( \vartheta_i X_i > \frac{1-c}{n} x, (\vartheta_0, \overline{\vartheta}) \in \Delta_2 \right) \leq \varepsilon \tilde{F}(x).
\] (6.9)

By conditioning on \( (\vartheta_0, \overline{\vartheta}) \) and applying the uniform asymptotic relation (6.6),

\[
J_3(x) \sim \sum_{i=1}^{n} \int_{\Delta_3} \cdots \int \Pr(s_i X_i > x + s_0) \Pr \left( \bigcap_{i=0}^{n} (\vartheta_i \in ds_i) \right)
\sim \sum_{i=1}^{n} \int_{\Delta_3} \cdots \int s_i^c \tilde{F}(x + s_0) \Pr \left( \bigcap_{i=0}^{n} (\vartheta_i \in ds_i) \right)
\sim \sum_{i=1}^{n} \mathbb{E}[\tilde{F}(x + \vartheta_0) \vartheta_i^c] - \sum_{i=1}^{n} \mathbb{E}[\tilde{F}(x + \vartheta_0) \vartheta_i^c 1_{(\vartheta_0, \overline{\vartheta}) \in \Delta_1 \cup \Delta_2}],
\] (6.10)

where in the second step we applied the uniform asymptotic result in Lemma 6.1(d). The last sum in (6.10) is negligible. Actually, it holds for all large \( x \) and \( C \) that

\[
\mathbb{E}[\tilde{F}(x + \vartheta_0) \vartheta_i^c 1_{(\vartheta_0, \overline{\vartheta}) \in \Delta_1 \cup \Delta_2}]
\leq \mathbb{E}[\vartheta_i^c 1_{(\vartheta_0 \leq -cx)}] + \tilde{F}(x) \mathbb{E}[\vartheta_i^c 1_{(\vartheta_0 > cx)}] + \tilde{F}((1-c)x) \mathbb{E}[\vartheta_i^c 1_{(\vartheta_0, \overline{\vartheta}) \in \Delta_2}]
\leq \varepsilon \tilde{F}(x).
\]

Substituting this into (6.10), then combining (6.7) with (6.8), (6.9), and (6.10), we obtain

\[
\mathbb{E} \left[ \tilde{F}(x + \vartheta_0) \sum_{i=1}^{n} \vartheta_i^c \right] - n \varepsilon \tilde{F}(x) \leq \Pr \left( \sum_{i=1}^{n} \vartheta_i X_i - \vartheta_0 > x \right) \leq \mathbb{E} \left[ \tilde{F}(x + \vartheta_0) \sum_{i=1}^{n} \vartheta_i^c \right] + 2 \varepsilon \tilde{F}(x).
\]
The first relation in (6.5) follows by the arbitrariness of \( \varepsilon \) and the second relation in (6.5). \( \blacksquare \)

Recall the bivariate AR model introduced in Section 2, in which \( B_i \) denotes the claim amount, \( a \) denotes the premium amount, and, hence, \( B_i - a \) denotes the net loss during period \( i \). In particular, as shown in (2.9), the ruin probability \( \psi(x; n) \) is the tail probability of the maximum of \( n \) randomly weighted sums. The following result targets the tail probability of a randomly weighted sum, and it forms the essence of the proof of Theorem 4.1.

**Lemma 6.4**

Let the notation in Section 2 be valid and consider the randomly weighted sum

\[
S_n = \sum_{i=1}^{n} \theta_i (B_i - a), \quad n \in \mathbb{N}.
\]

If the i.i.d. innovations \( X_i, i \in \mathbb{N} \), of the AR(1) process (2.1) follow a common distribution function \( F \in \mathcal{R}_a \) for some \( \alpha > 0 \) and the random weights satisfy \( \mathbb{E}[\theta_i^\beta] < \infty \) for some \( \beta > \alpha \) and all \( i \in \mathbb{N} \), then for every \( n \in \mathbb{N} \), the sum \( S_n \) still follows a distribution in the class \( \mathcal{R}_a \) and the relation

\[
\Pr(S_n > x) \sim \mathbb{E} \left[ F \left( x + \sum_{i=1}^{n} \theta_i (\alpha - \rho_i B_0) \right) \right. \left. \sum_{j=1}^{n} \left( \sum_{i=1}^{n} \theta_i \rho_i^{-j} \right)^{\alpha} \right]
\]

holds. In case \( \alpha - \rho_i B_0 \geq 0 \), the moment condition can be weakened to \( \mathbb{E}[\theta_i^\beta] < \infty \) for some \( \beta > \alpha \) and all \( i \in \mathbb{N} \).

**Proof**

By (6.2) we have

\[
S_n = \sum_{i=1}^{n} \theta_i \left( \sum_{j=0}^{i-1} \rho_i^j X_{i-j} + \rho_i^i B_0 \right) - a \sum_{i=1}^{n} \theta_i
\]

\[
= \sum_{j=1}^{n} \left( \sum_{i=1}^{n} \theta_i \rho_i^{-j} \right) X_j - \sum_{i=1}^{n} \theta_i (\alpha - \rho_i B_0)
\]

\[
= \sum_{j=1}^{n} \theta_j X_j - \theta_0. \quad (6.11)
\]

It is easy to verify all the conditions of Lemma 6.3. For instance, \( \mathbb{E}[\theta_i^\alpha 1_{(\theta_i \leq -x)}] = o(\tilde{F}(x)) \) holds because

\[
\mathbb{E}[\theta_i^\alpha 1_{(\theta_i \leq -x)}] \leq x^{-\beta} \mathbb{E}[|\theta_i|^\beta \theta_i^\alpha],
\]

while \( x^{-\beta} = o(\tilde{F}(x)) \) by Lemma 6.1(e) and, for some \( C > 0 \),

\[
\mathbb{E}[|\theta_i|^\beta \theta_i^\alpha] \leq C \sum_{i=1}^{n} \mathbb{E}[\theta_i^{2\alpha} + \theta_i^{2\alpha}] < \infty
\]

by \( c_i \)-inequality. Finally, an application of Lemma 6.3 concludes the proof. \( \blacksquare \)

The following lemma is interesting in its own right.

**Lemma 6.5**

Let \( Z_1, \ldots, Z_n \) be \( n \) not necessarily independent random variables with partial sums \( S_m = \sum_{i=1}^{m} Z_i \) for \( m = 1, \ldots, n \). If the last sum \( S_n \) follows a distribution in the class \( \mathcal{R}_a \) for some \( \alpha > 0 \) and \( \Pr(Z_i \leq -x) = o(\Pr(S_n > x)) \) for each \( i = 1, \ldots, n \), then

\[
\Pr \left( \max_{1 \leq m \leq n} S_m > x \right) \sim \Pr(S_n > x).
\]
Proof
It suffices to prove the asymptotic inequality \( \Pr(\max_{1 \leq m \leq n} S_m > x) \leq \Pr(S_n > x) \) because the other one is trivial. For \( A \subset \mathcal{J} = \{1, \ldots, n\} \), define
\[
\Omega_A = \{Z_i > 0 \text{ whenever } i \in A \text{ and } Z_i \leq 0 \text{ whenever } i \in \mathcal{J} \setminus A\}.
\]
Then, for every \( 0 < \varepsilon < 1/n \),
\[
\Pr \left( \max_{1 \leq m \leq n} S_m > x \right)
\leq \sum_{A \subset \mathcal{J}} \Pr \left( \sum_{i \in A} Z_i > x, \Omega_A \right)
= \sum_{A \subset \mathcal{J}} \Pr \left( \sum_{i \in A} Z_i > x, \Omega_A, \bigcap_{i \in \mathcal{J} \setminus A} (Z_i \leq -\varepsilon x) \right) + \sum_{A \subset \mathcal{J}} \Pr \left( \sum_{i \in \mathcal{J} \setminus A} Z_i > x, \Omega_A, \bigcup_{i \in \mathcal{J} \setminus A} (Z_i \leq -\varepsilon x) \right)
\leq \sum_{A \subset \mathcal{J}} \Pr(S_n > (1 - n\varepsilon)x, \Omega_A) + \sum_{A \subset \mathcal{J}} \sum_{i \in \mathcal{J} \setminus A} \Pr(Z_i \leq -\varepsilon x)
= \Pr(S_n > (1 - n\varepsilon)x) + o(1)\Pr(S_n > \varepsilon x)
\sim (1 - n\varepsilon)^{-\alpha} \Pr(S_n > x).
\]
The arbitrariness of \( \varepsilon \) concludes the proof.

Lemma 6.6
Let \( F \in \mathcal{R}_{-\alpha} \) for some \( \alpha > 0 \) and let \( \Sigma_0 \) and \( \Sigma \) be two random variables with \( \Sigma_0 \) real valued and \( \Sigma \) positive.

a. If \( E[\Sigma] < \infty \) and \( E[\Sigma_1(\Sigma_0 = -x)] = o(\tilde{F}(x)) \), then \( E[\tilde{F}(x + \Sigma_0)\Sigma] \sim \tilde{F}(x)E[\Sigma] \).

b. If \( E[\Sigma^2] < \infty \) and \( E[\Sigma^2_1(\Sigma_0 = -x)] = o((\tilde{F}(x))^2) \), then \( E[(\tilde{F}(x + \Sigma_0)\Sigma)^2] \sim (\tilde{F}(x))^2E[\Sigma^2] \).

Proof
Both can be proved by following the proof of the second relation in (6.5).

7. Acknowledgments
The authors thank the two anonymous referees for their helpful comments on an earlier version of this paper. The research for this project is sponsored by The Actuarial Foundation and the Society of Actuaries through the 2011 Individual Grants Competition. The source code for the numerical studies in this paper is available on both authors’ homepages or upon request.

References


Discussions on this paper can be submitted until January 1, 2013. The authors reserve the right to reply to any discussion. Please see the Submission Guidelines for Authors on the inside back cover for instructions on the submission of discussions.