GPH 파산확률 계산방법의 실험적 검토

Various experiments for the GPH-based ruin probability computaion method

윤복식

홍익대학교 기초과학과

bsyoon@hongik.ac.kr

초록

보험요율 및 정책 결정의 기본이 되는 파산확률은 계산이 매우 복잡하여 보통 근사적인 방법을 사용하게 된다. 윤복식(2007)에서는 다양한 상황에서 정확하게 파산확률을 계산할 수 있는 방법이 GPH 분포에 기초하여 제안된 바 있다. 본 연구에서는 이 방법의 타당성을 다양한 실험을 통해 검증하고. 기존의 근사적 방법들과의 비교하는 것이 목적이다. 실험을 통해 이 방법이 일상적인 상황에서 뿐 아니라 클레임 분포가 비정규적인 대재해 상황에서도 정확하게 파산확률을 계산해 주는 것을 관찰할 수 있었고, 계산시간 또한 도 매우 짧아서 실용성을 겸비함을 확인할 수 있었다. 또한 이 결과를 근거로, 기초적인 관측 자료만 입력하면 중간에 분포모델 설정단계를 거치지 않고 바로 분석 결과를 얻는 접근법이 제안된다.

1. 서론

보험요율 및 정책 결정의 기본이 되는 파산확률의 계산은 매우 복잡하여 보통 근사적인 방법을 사용하고 있다. 본 연구는 일상적인 상황에서 뿐 아니라 클레임 분포가 비정규적인 대재해 상황 등 다양한 상황에서 정확하게 파산확률을 계산할 수 있게 해주는 GPH 분포에 근거한 방법(윤복식, 2007)의 타당성을 실험적으로 검증해 보는 것이 목적이다. 이 방법은 서비스분포가 특이한 heavy-tail 분포일 경우의 M/G/1 대기시스템에서 체류시간 분포계산을 위해서도 사용될 수 있다

2. 보험에서의 파산확률

본 연구에서는 접근의 단순성을 위해 이자의 발생을 고려하지 않고 클레임의 발생을 Poisson 과정으로 한정하여 파산확률의 근사화 및 분석을 시도한다. 이 가정은 추후 완화될 것이다.

시간 t까지의 클레임의 발생과정 $\{N(t), t \geq 0\}$가 발생률 α인 포아송 과정 따르고 각 클레임의 규모 $Z_i, i=1,2,\ldots$는 iid F를 따르고 독립이고 $N(t)$와도 독립이라고 가정한다. $(0,t]$사이의 누적클레임은

$$Y_t = \sum_{i=1}^{N(t)} Z_i$$

로 나타낼 수 있고 보험료율이 p로 일정하다면 t에서의 누적보험료와 누적클레임의 차이는

$$X_t = pt - Y_t$$

로 나타낼 수 있다. 초기$(t=0)$에 위험유보금을 u로 준비했다면 t에서의 위험유보금은

$$R_t = u + pt - \sum_{k=1}^{N_t} Z_k = u + X_t \tag{1}$$

로 정의된다.

Z_i의 평균을 μ, 분산을 σ^2로 표기하면

$$ER_t = u + pt - \alpha\mu t = u + (p - \alpha\mu)t \ .$$

이때 $\eta = \dfrac{p-\alpha\mu}{\alpha\mu}$는 안전할증률(safety loading)이라 불린다. $\rho = \alpha\mu$라고 정의하면 $\eta = \dfrac{p-\rho}{\rho}$가 되고 특히 $p=1$일 때는 $\eta = \dfrac{1-\rho}{\rho}$, $1+\eta = \dfrac{1}{\rho}$의 관계가 된다. 앞으로 $p=1$로 가정하고 파산이 확실히 일어나는 경우는 고려 대상에서 제외하여 $\eta > 0$ $(0 < \rho < 1)$로 설정한다.

이제 파산확률은 고려 대상시간이 무한대일 때는

$$\Psi(u) = P\{ \inf_{t \geq 0} R_t < 0 \}$$

유한시간 T일 때는

$$\Psi(u, T) = P\{ \inf_{0 \leq t \leq T} R_t < 0 \}$$

로 정의된다. 본 연구에서는 일단 무한시간에 초점을 맞춘다.

이제 F의 누적꼬리분포(integrated tail

distribution) 또는 평형분포(equilibrium distribution)를

$$F_e(z) \equiv \frac{1}{\mu}\int_0^z \overline{F}(x)dx \ , \ \ \overline{F}(x) = 1 - F(x)$$

라고 정의하면 파산확률의 기본 공식인 Pollaczeck-Kinchine 공식을 다음과 같이 얻을 수 있다.

$$\Psi(u) = \frac{\eta}{1+\eta}\sum_{n=0}^{\infty}\left(\frac{1}{1+\eta}\right)^n \overline{F_e^{(n)}}(u), \qquad (2)$$

여기서 $F_e^{(n)}(u)$는 F_e의 n중 합성곱(n-fold convolution)이다. $p=1$일 때 (2)를 다시 표현하면

$$\Psi(u) = (1-\rho)\sum_{n=0}^{\infty}\rho^n \overline{F}_e^{(n)}(u) \qquad (3)$$

가 된다. (2)에서

$$\Psi(0) = \frac{\alpha\mu}{p} = \frac{1}{1+\eta} \quad (\eta > 0\text{일 때}) \qquad (4)$$

임을 관찰을 할 수 있고, 특히 $p=1$일 때는 $\Psi(0)=\rho$가 됨을 알 수 있다.

F가 $\exp(\frac{1}{\mu})$일 때는 (2)는 다음과 같이 정확하게 계산 가능하다.

$$\Psi(u) = \frac{1}{1+\eta}e^{-(\eta u/\mu(1+\eta))} \qquad (5)$$

이때 $p=1$이면 $\delta = \frac{1}{\mu}$라 놓을 때

$$\Psi(u) = \rho e^{-(\delta-\alpha)u} \qquad (6)$$

가 얻어진다.

3. GPH 근사화

3.1. GPH(λ,g) 분포의 성질

$(E_n)_0^{\infty}$을 평균이 $1/\lambda$인 지수 분포를 따르는 확률변수 열이라 하고(단, E_0=0 w.p.1), L을 분포함수 G와 확률질량함수(pmf, probability mass function) g를 갖는 이산 확률변수라고 할 때, $X=\sum_{n=0}^{L}E_n$의 분포 함수 F는 다음과 같이 쓸 수 있다.

$$F_X(x) = \sum_{n=0}^{\infty}G(n)\frac{e^{-\lambda x}(\lambda x)^n}{n!}, \ \ x \geq 0 \qquad (7)$$

이 분포를 $GPH(\lambda,g)$로 정의한다. 이것을 x로 미분하면 pmf는

$$f_X(x) = \sum_{n=0}^{\infty}g(n+1)\frac{e^{-\lambda x}(\lambda x)^n}{n!}\lambda, \ \ x \geq 0$$

로 얻어지고 Laplace-Stieltjes 변환(LST)는

$$\tilde{F}_X(s) \equiv \int_0^{\infty}e^{-sx}dF_X(x) = \hat{G}\left(\frac{\lambda}{\lambda+s}\right)$$

이 되는데 여기서 $s>0$ 이고

$$\hat{G}(z) \equiv \sum_{n=0}^{\infty}g(n)z^n, |z|<1,$$

즉 L의 확률생성함수(pgf)이다. 또한 X의 생존함수(survival function) $\overline{F_X}(x) \equiv 1-F(x)$는

$$\overline{F_X}(x) = \sum_{n=0}^{\infty}\overline{G}(n)\frac{e^{-\lambda x}(\lambda x)^n}{n!}, \ \ x>0$$

로 표현된다. 여기서

$$\overline{G}(n) = 1-G(n), n=0,1,2,\ldots$$

이다.

또한 F가 연속인 $GPH(\lambda,g)$ 분포라면 평형분포 F_e는

$$\begin{aligned}
F_e(x) &\equiv \frac{1}{E(X)}\int_0^x(1-F(t))\,dt \\
&= \frac{1}{E(X)}\int_0^x\sum_{k=0}^{\infty}\overline{G}(k)\frac{e^{-\lambda t}(\lambda t)^k}{k!}dt \\
&= \frac{1}{E(X)}\sum_{k=0}^{\infty}\overline{G}(k)\int_0^x\frac{e^{-\lambda t}(\lambda t)^k}{k!}dt \\
&= \frac{1}{\lambda E(X)}\sum_{k=0}^{\infty}\overline{G}(k)\sum_{n=k+1}^{\infty}\frac{e^{-\lambda t}(\lambda t)^n}{n!} \\
&= \frac{1}{E(L)}\sum_{n=1}^{\infty}\left(\sum_{k=0}^{n-1}\overline{G}(k)\right)\frac{e^{-\lambda t}(\lambda t)^n}{n!} \\
&= \sum_{n=1}^{\infty}G_e(k)\frac{e^{-\lambda t}(\lambda t)^n}{n!}
\end{aligned}$$

을 통해 역시 $GPH(\lambda,g_e)$ 분포임을 알 수 있다. 여기서 $G_e(n) = \frac{1}{E[L]}\sum_{k=0}^{n-1}\overline{G}(k), n=1,2,\cdots$.

한편 GPH 분포는 합성곱(convolution)에 닫혀 있다. 이 사실은 쉽게 파악할 수 있는데 X_1을 $GPH(\lambda,g_1)$, X_2를 $GPH(\lambda,g_2)$ 분포를 $G_e(n) = \frac{1}{E(L)}\sum_{k=0}^{n-1}\overline{G}(k), n=1,2,\ldots$따르는 독립인 두개의 확률 변수라고 하고, L_1과 L_2를 각각 g_1, g_2를 pmf로 갖는 확률변수라고 하자. $X_1 \sim \sum_{n=0}^{L_1}E_n$이고, $X_2 \sim \sum_{n=L_1+1}^{L_1+L_2}E_n$이기 때문에, $X=X_1+X_2$라 하면, $X \sim \sum_{n=0}^{L_1+L_2}E_n$가 되어, X는 $GPH(\lambda, g_1{*}g_2)$ 분포를 따르게 된다. 여기서 $g_1{*}g_2(n) \equiv \sum_{k=0}^{n}g(n-k)g(k)$, 이산 합성곱을 의미하고 기호 ~는 같은 분포를 따르는 관계를 의미한다. 여기서, 보듯이 연속 변수인

X_1과 X_2의 연속 합성곱(적분)이 이산 합성곱(합)으로 바뀌게 되는데 이 성질은 계산적인 응용을 가능하게 하는 중요한 성질이다. 물론, 이와 같은 연산이 가능하기 위해서는, 두 GPH 분포의 λ값을 동일하게 설정해야 한다.

3.2 GPH분포의 추정

3.2.1 분포함수가 알려진 경우

일반적으로 F가 $[0, \infty]$에서 정의된 유계함수이면 다음의 GPH 형태의 함수는

$$F_\lambda(t) = \sum_{n=0}^{\infty} F\left(\frac{n}{\lambda}\right) \frac{e^{-\lambda t}(\lambda t)^n}{n!} \qquad (8)$$
$$t \geq 0, \quad \lambda > 0$$

$\lambda \to \infty$일 때 $F(t)$로 수렴함을 보일 수 있다. 따라서 $[0, \infty]$에서 정의된 분포는 모두 GPH 형태로 근사화 시킬 수 있다.

3.2.2 데이터로부터의 추정

샘플규모가 m인 자료 $U_i (i = 1, 2, \cdots, m)$에 대해 분포함수를 $F(x)$, $x \geq 0$라고 하면 F의 경험분포함수 $F_m(x)$는 다음과 같이 구해진다.

$$F_m(x) = \sum_{i=1}^{m} \frac{I(U_i \leq x)}{m}$$

$$\text{여기서 } I(A) = \begin{cases} 1, & A\text{가 참이면} \\ 0, & \text{그 외의 경우} \end{cases}$$

$m \to \infty$일 때 $F_m(x) \to F(x)$ $(w.p.\ 1)$이므로 (8)에서 $F(x)$대신 $F_m(x)$를 대입하면

$$F_{m,\lambda}(x) = \sum_{n=0}^{\infty} F_m\left(\frac{n}{\lambda}\right) \frac{e^{-\lambda x}(\lambda x)^n}{n!} \quad (9)$$

이 얻어진다. (9)는 모든 고정된 m에 대해, $\lambda \to \infty$일 때 $F_{m,\lambda}(x) \to F_m(x)$ 이 되므로 그 다음에 m을 크게 하면 결국 $F_{m,\lambda}(x) \to F(x)$ (w.p.1)임을 관찰할 수 있다. 따라서 (9)를 이용하면 경험분포로부터 직접 GPH 형태의 분포를 추정할 수 있다. 여기에 대한 다양한 실험적 연구는 윤복식 외(1994)에서 찾아볼 수 있다.

3.2.3 GPH분포의 효과적 계산방법

식(8)은 무한급수이므로 계산을 위해서는 적절한 절단이 필요하다. 우선

$$\sum_{n=0}^{K} \frac{e^{-\lambda t}(\lambda t)^n}{n!} \geq 1 - \epsilon$$

인 최소의 K를 구하면

$$\sum_{n=K+1}^{\infty} F\left(\frac{n}{\lambda}\right) \frac{e^{-\lambda t}(\lambda t)^n}{n!} \leq \sum_{n=K+1}^{\infty} \frac{e^{-\lambda t}(\lambda t)^n}{n!} \leq \epsilon$$

이 된다. 따라서

$$F_\lambda(t) \cong \sum_{n=0}^{K} F\left(\frac{n}{\lambda}\right) \frac{e^{-\lambda t}(\lambda t)^n}{n!} \quad (10)$$

로 계산하면 절단오차를 ϵ이하로 유지할 수 있다. λ가 충분히 큰 경우 $\epsilon = 10^{-7}$ 일 때 대략

$$K \cong \lambda t + 5\sqrt{\lambda t}$$

이면 충분하다. 이를 적용하여 (10)을 계산할 때 계산의 안정성과 효율을 위해 다음과 같은 방법을 사용할 수 있다.

우선 $k = \lceil \lambda t \rceil$ 일 때

$$P_n = \frac{e^{-\lambda t}(\lambda t)^n}{n!}$$

를 먼저 계산하고 관계식

$$P_{n+1} = \frac{\lambda t}{n+1} P_n$$

을 이용하여 $P_{n+1}, P_{n+2}, \cdots, P_{n+\lceil 5\sqrt{\lambda t} \rceil}$ 를 차례로 계산하고 마찬가지로

$$P_{n-1} = \frac{n}{\lambda t} P_n$$

를 이용하여

$$P_{n-1},\ P_{n-2},\ \cdots,\ P_{n-\lceil 5\sqrt{\lambda t} \rceil}$$

를 차례로 구하여

$$\sum_{n=\lceil \lambda t \rceil - \lceil 5\sqrt{\lambda t} \rceil}^{\lceil \lambda t \rceil + \lceil 5\sqrt{\lambda t} \rceil} F\left(\frac{n}{\lambda}\right) P_n$$

와 같이 $F_\lambda(t)$의 근사값을 계산한다.

4. 파산확률의 계산

4.1 M/GPH/1의 이용

Pollaczeck-Kinchine 공식 (3)을 다시 쓰면

$$\Psi(u) = (1-\rho) \sum_{n=0}^{\infty} \rho^n \overline{F}_e^{(n)}(u) \qquad (11)$$
$$= 1 - (1-\rho) \sum_{n=0}^{\infty} \rho^n F_e^{(n)}(u)$$

이 된다. (11)의 식은 다중 합성곱 때문에 일부 특수한 경우를 제외하고는 사실상 계산이 불가능하다. 그러나 F가 GPH 라면 합성곱을 이산화 할 수 있다. 클레임의 분포 F가 $GPH(\lambda, g)$라고 가정하면 F_e는 $GPH(\lambda, g_e)$이고 따라서 $F_e^{(n)}$도 LST가 $(G_e(\frac{\lambda}{\lambda+s}))^n$ 인 GPH가 되고 그것들의 기하분포에 의한 혼합(mixture)인 $1-\Psi(u)$도 GPH가 된다. $1-\Psi(u)$를 $GPH(\lambda, g_Q)$이라 하면 이제 g_Q을 효율적으로 계산하는 방법이 필요하다.

실제로 $1-\Psi(u)$는 FIFO-M/G/1 대기열시스

템에서 정상상태의 대기시간(delay) 분포로 해석할 수 있다(Asmussen,1987). 보다 구체적으로 고객의 도착과정이 도착률 α인 포아손과정이고 서비스 시간이 평균 μ인 분포F를 따른다고 하고 $\rho=\alpha\mu<1$ 라고 가정하자. 이때 F가 $GPH(\lambda,g)$ 라고 가정하면 앞의 설명과 같이 정상상태의 대기시간의 분포 F_Q도 GPH 형태가 되는데 이것을 $GPH(\lambda,g_Q)$로 표시하자. g_Q를 구하기 위해 우선 대기시간이 0일 확률은 $1-\rho$이므로

$$g_Q(0)=F_Q(0)=1-\rho \qquad (12)$$

이다. 가상지연과정에서 수준 x를 교차하여 내려가고 올라가는 발생율(rate)이 같아야 하므로

$$f_Q(x)=\alpha\int_0^x (1-F(x-t))\,dF_Q(t),\ x>0$$

이 되는데 좌변을 GPH 형태로 나타내면

$$f_Q(x)=\sum_{n=0}^{\infty} g_Q(n+1)\frac{e^{-\lambda x}(\lambda x)^n}{n!}\lambda \qquad (13)$$

이고, 우변에 GPH 형태를 대입하여 정리하면

$$\alpha\sum_{n=0}^{\infty}\left[\sum_{k=0}^{n}\widehat{G}(n-k)g_Q(k)\right]\frac{e^{-\lambda x}(\lambda x)^n}{n!} \qquad (14)$$

이 된다. (13), (14)를 등호로 연결하고 비교하면 $n=0,1,\cdots$에 대해

$$g_Q(n+1)=\left(\frac{\alpha}{\lambda}\right)\sum_{r=0}^{n}\overline{G}(n-r)g_Q(r) \qquad (15)$$

과 같은 되돌림 관계를 얻을 수 있다.

(12)와 (15)를 이용하여 $g_Q(n), n=0,1,\cdots$를 구한 후

$$G_Q(n)=\sum_{k=0}^{n} g_Q(k),\ n=0,1,\cdots$$

를 구하여

$$F_Q(x)=\sum_{n=0}^{\infty} G_Q(n)\frac{e^{-\lambda x}(\lambda x)^n}{n!}$$

와 같이 대기시간 분포를 계산하여 최종적으로 파산확률을

$$\Psi(u)=1-F_Q(u)$$

로 구할 수 있다.

클레임 분포함수가 F로 주어진 경우는 3.2.1의 근사화를 사용하여

$$G(n)=F\left(\frac{n}{\lambda}\right)$$

로 놓으면 된다.

4.2. 계산 예

Willmot(1998)에 주어진 예을 사용하여 파산확률을 계산하고 정확성을 비교해보자. 그 논문과 같이 클레임 규모의 pdf가

$$f(x)=b(bx)^{a-1}e^{-bx}/\Gamma(a)$$

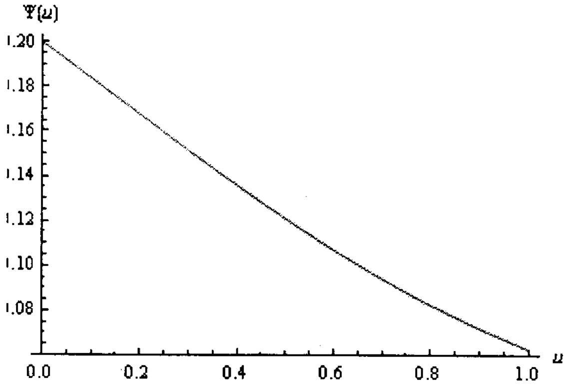
즉, Gamma(a,b) 분포인 경우를 우선 고려해보자. $a=3, b=3$인 경우(이때 평균은 $\mu=a/b=1$, 분산은 $\sigma^2=a/b^2=1/3$, 제곱변동계수는 $c^2=\sigma^2/\mu^2=1/a=1/3$)와 $a=0.5, b=0.5$ (이때 $\mu=1, c^2=2$)인 경우에 $\rho=0.2, 0.5, 0.8$ 일 때의 파산확률의 계산결과가 표 1과 표 2에 주어져 있다. 표에서 GPH 방법의 정확성을 확인할 수 있다. 또한 GPH 방법을 이용하여 구한 Gamma(3,3), $\rho=0.2$일 때의 u에 대한 $\Psi(u)$의 그래프가 그림 1에 주어져 있다.

표 1. c^2=1/3 인 경우의 파산확률
(GPH: $\lambda=256$)

ρ	u	GPH	Willmot	참값
0.2	0.1	0.1839	0.1822	0.1839
	0.25	0.1595	0.1596	0.1594
	0.5	0.1209	0.1217	0.1209
	0.75	0.0883	0.0886	0.0882
	1	0.0627	0.0627	0.0626
0.5	0.1	0.4744	0.4728	0.4744
	0.25	0.4342	0.4344	0.4342
	0.5	0.3665	0.3672	0.3664
	0.75	0.3035	0.3037	0.3033
	1	0.2485	0.2484	0.2484
0.8	0.1	0.7834	0.7826	0.7834
	0.25	0.7562	0.7564	0.7562
	0.5	0.7075	0.7074	0.7074
	0.75	0.6578	0.6578	0.6577
	1	0.6098	0.6096	0.6097

표 2. c^2=2 인 경우의 파산확률

ρ	u	GPH	Willmot	참값
0.2	0.1	0.1865	0.1849	0.1865
	0.25	0.1697	0.1685	0.1697
	0.5	0.1465	0.1461	0.1466
	0.75	0.1274	0.1275	0.1276
	1	0.1113	0.1117	0.1115
0.5	0.1	0.4786	0.4768	0.4787
	0.25	0.4511	0.4499	0.4512
	0.5	0.4113	0.4111	0.4114
	0.75	0.3765	0.3769	0.3768
	1	0.3455	0.3462	0.3458
0.8	0.1	0.7861	0.7852	0.7862
	0.25	0.7677	0.7672	0.7677
	0.5	0.7396	0.7396	0.7397
	0.75	0.7136	0.7140	0.7139
	1	0.6892	0.6896	0.6894



5. 결 론

본 연구를 통해 GPH 분포에 기반한 파산확률 계산방법의 타당성과 정확성을 부분적으로 확인할 수 있었다. 향후 두꺼운 꼬리를 가진 경우 등 보다 다양한 클레임 분포에 대한 실험이 수행될 것이고 클레임 데이터로부터 직접 파산확률 얻는 일관된 방법에 대한 연구가 계속될 것이다. 또한 이자율이 고려되는 상황이나 포아손 도착을 완화하는 방향으로의 확장이 이루어질 것이다.

참고문헌

윤복식, 박광우, 이창훈, "*GPH분포에 의한 확률적 근사화*", 한국 경영과학회지 제19권, 1994.

Asmussen, S., *Ruin Probability*, World Scientific, Singapore, 2000.

Beard R. E., T. Penetikäinen and E. Pesonen, *Risk Theory: The Stochastic Basis of Insurance*, Chapman and Hall, London, 1984

Brekelmans, R. and A. De Waegenaere, Approximation the finite-time ruin probability under interest force, *Insurance: Mathematics and Economics*, 29, 217-229, 2001

Chen, Y and K.W. Ng, The ruin probability of the renewal model with constant interest force and negatively dependent heavy-tailed claims, *Insurance: Mathematics and Economics*, 40, 415-423,2007.

Dickson, D.C.M and H.R. Waters, Reinsurance and ruin, *Insurance: Mathematics and Economics*, 19, 61-80, 1994.

dos Reis, A.D.E., On the moments of ruin and recovery times, *Insurance: Mathematics and Economics*, 27 (2000) 331-343.

Embrechts, P. and G. Samorodnitsky, Ruin problem and how fast stochastic processes mix, *The Annals of Applied Probability*, 13, 1-36, 2003.

Grandell, J., Simple approximations of ruin probabilities, *Insurance: Mathematics and Economics*, 26, 157-173, 2000

Ignatov, Z.G, V.K. Kaishev, and R.S. Krachunov, An improved finite-time ruin probability formula and its Mathematica implementation, *Insurance: Mathematics and Economics*, 29 (2001) 375-386.

Kalashnikov, V. and D. Konstantinides, Ruin under interest force and subexponential claims: a simple treatment, *Insurance: Mathematics and Economics*, 27, 145-149, 2000. (subexponential)

Kalashnikov, V. and R. Norberg, Power tailed ruin probabilities in the presence of risky investments, Stochastic Processes and their Applications 98 (2002) 211-228.

Shanthikumar J. G. , "Bilateral Phase-Type Distributions," *Naval Research Logistics Quarterly*, vol. 32, 1985, pp. 119-136.

Sundt, B. and J.L. Teugels, The adjustment function in ruin estimates under interest force, *Insurance: Mathematics and Economics*, 19 (1997) 85-94.

Willmot, G.E., On a class of approximations for ruin and waiting time probabilities, *Operations Resaerch Letters*, 22 (1998) 27-32.