STATE-DEPENDENT FEES FOR VARIABLE ANNUITY GUARANTEES

BY

CAROLE BERNARD, MARY HARDY AND ANNE MACKAY

ABSTRACT

For variable annuity policies, management fees for the most basic guarantees are charged at a constant rate throughout the term of the policy. This creates a misalignment of risk and income – the fee income is low when the option value is high, and vice versa. In turn, this may create adverse incentives for policyholders, for example, encouraging surrenders when the options are far out-of-the-money.

In this paper, we explore a new fee structure for variable annuities, where the fee rate supporting the cost of guarantees depends on the moneyness of those guarantees. We derive formulas for calculating the fee rates assuming fees are paid only when the guarantees are in-the-money, or are close to being in-the-money, and we illustrate with some numerical examples. We investigate the effect of this new fee structure on the surrender decision.

KEYWORDS

Variable annuities, pricing, GMMB, GMDB, surrender decision.

1. INTRODUCTION

Variable Annuities (VAs) are generally issued as definite term, single or regular premium, separate account investment policies, with guaranteed minimum payments on death or maturity. Because of the embedded guarantees, and because of advantageous tax treatment of the proceeds, the demand for these products has been strong, even through the recent recession. The technical challenges created by these products include design, pricing, valuation and risk management; of course, each of these is intricately related to all the others (Boyle and Hardy, 2003; Hardy, 2003; Palmer, 2006; Coleman *et al.*, 2007) .

The original design of the contracts issued in North America in the 1990s and early 2000s offered fairly straightforward guarantees – for example, a minimum payment or a return of premium on death or maturity. However, over time the guarantees have become increasingly complex. The range of guaranteed

Astin Bulletin 44(3), 559-585. doi: 10.1017/asb.2014.13 © 2014 by Astin Bulletin. All rights reserved.

https://doi.org/10.1017/asb.2014.13 Published online by Cambridge University Press

minimum benefits is often referred to as "GMxBs" (Bauer *et al.* 2008), to cover GMDBs (death benefits), GMMBs (maturity), GMIBs (income), GMABs (accumulation) and so on. The more complex guarantees have evolved partly for marketing purposes, to distinguish one insurer's product from its competitors', but also, in some cases, to avoid potentially costly lapse and re-entry incentives. Modern products offer a range of income and withdrawal benefits as optional riders, including the plain vanilla 'capital protection' or GMMB.

Guarantees are typically funded by a fixed fee rate. It has been noted (see Bauer *et al.*, 2008 for example) that the fixed fee structure creates incentives for policyholders to surrender when their guarantees are well out-of-the-money. If a policy with a return-of-premium GMMB has an account value very much higher than the guarantee, then it may be worthwhile for the policyholder to lapse the contract and invest directly in the underlying funds; there is little point in continuing to pay for an option with negligible value. Or, the policyholder could lapse and purchase a new policy with the proceeds; the new policy would have a guarantee which is at-the-money, for (perhaps) the same cost as the old policy with an out-of-the-money guarantee. It will not always be rational for the policyholder to lapse and re-enter, for example, if the tax liability exceeds the additional value created by the lapse and re-entry, or if the insurer imposes a surrender charge, but in many cases it may be an optimal strategy for the policyholder (see Milevsky and Salisbury, 2001; Moenig and Bauer, 2012; Bernard and MacKay, 2014 and Bernard *et al.* 2014).

The fixed fee rate is unsatisfactory from a risk management perspective also because there is a misalignment of the fee income and the option cost. When markets fall, the option value is high, but the fee income is reduced. When markets rise, the fee income increases, but there is negligible guarantee liability.

In this paper we investigate a dynamic fee structure for GMMBs and GMDBs, where the fee rate depends in some way on the evolving embedded option value. Specifically, we develop pricing formulas for a contract for which the fee (applied as a proportion of the policyholder's fund) is only payable when the fund value is below the guaranteed benefit – that is, when the embedded option is in-the-money. A similar formula was derived for equity options in Karatzas and Shreve (1984). However, our formula is more general, and we present details that will make it easier to use for the reader. We also extend the results to allow for a threshold for fee payments that is higher than the guaranteed amount, so that the fee begins to be paid when the fund moves close to being in-the-money. For convenience, we refer to this fee structure as 'state-dependent', where the states being considered are simply (1) the policyholder's fund F_t lies above a given barrier B, and (2) F_t lies below the barrier, B. The motivation for this paper was a VA-type contract issued by Prudential UK, whose "flexible investment plan" is an equity-linked policy with an optional GMDB, for which the fee is paid only if the fund value is below the guaranteed level. See Prudential(UK) (2012), page 11, for details. Thus, this is a practical design proposition, at least for the GMDB. The advantages for an optional GMDB rider are clear; once the guarantee is no longer valuable, the policyholder pays the same fees as those

https://doi.org/10.1017/asb.2014.13 Published online by Cambridge University Press

who did not select the GMDB rider. Additionally, the alignment of income and liability value is better managed. One major disadvantage however is that by paying for the option only when the option is in the money, the fee rates at those times will be high, especially for GMMBs, to the extent that the contract will no longer appear competitive with other investment alternatives, and which could also exacerbate the option liability through fee drag. We discuss this further in Section 4.

Although the policyholder's option to surrender is impacted by the fee structure, in this paper we derive and evaluate fair fees assuming, initially, no policy surrenders. It would be straightforward to adapt the formulae to allow for deterministic surrenders. In Section 5 we consider further the impact of the fee structure on the surrender incentive. Many recent works on variable annuities have incorporated optimal surrenders in the pricing analysis (see, for example, Belanger *et al.*, 2009; Bacinello *et al.*, 2011). The complexity added by the state-dependent fee makes this more complex in this case than in the flat fee context, and a full analysis is beyond the scope of this paper. In Section 5 we present a preliminary analysis of the surrender incentive, to support the intuition that the state-dependent fee structure reduces the incentive to surrender for policies that are substantially out-of-the-money.

The rest of this paper is organized as follows. In Section 2 we present the model (which follows the standard Black-Scholes assumptions) and the pricing of a general European option given that the fee is only paid when the fund value is below some critical level. In Section 3 we apply the results to the simple guaranteed minimum maturity and death benefits embedded in a VA contract. Some numerical illustrations can be found in Section 4. A first analysis of the surrender incentive is presented in Section 5. Section 6 studies the robustness of the results to the Black-Scholes model assumption. Section 7 concludes.

2. Pricing with state-dependent fee rates

2.1. Notation

Consider a simple variable annuity contract and denote by F_t the underlying fund value at time t. Assume that there is a single premium P paid at time 0 by the policyholder. Further assume that the premium (minus some fixed initial expense) is fully invested in an equity index S_t which follows a geometric Brownian motion under the risk-neutral measure Q:

$$dS_t = S_t(r\,dt + \sigma\,dW_t).$$

Here r denotes the continuously compounded risk-free rate and σ is the volatility. Furthermore suppose that management fees are paid out of the fund at a constant percentage c whenever the value of the fund, F_t, lies below a given

https://doi.org/10.1017/asb.2014.13 Published online by Cambridge University Press

level B. Let C_t be the total fees paid at time t; its dynamics are given by

$$dC_t = cF_t \mathbb{1}_{\{F_t<B\}}\, dt,$$

where $\mathbb{1}_A$ is the indicator function of the set A. The dynamics of the fund are thus given by

$$\frac{dF_t}{F_t} = \frac{dS_t}{S_t} - c\mathbb{1}_{\{F_t<B\}}\, dt, \quad 0 \leqslant t \leqslant T, \tag{1}$$

with $F_0 = P - e$, where P is the initial premium and e is the fixed expense charge deducted by the insurer at inception. Throughout this paper, we assume $e = 0$. Using Ito's lemma to compute $d\ln(F_t)$, and integrating, we obtain the following representation of (1),

$$F_t = F_0 \exp\left\{\left(r - \frac{\sigma^2}{2}\right)t - c\int_0^t \mathbb{1}_{\{F_s<B\}} ds + \sigma W_t\right\}. \tag{2}$$

2.2. Pricing the VA including guarantees with state-dependent fee rates

We first consider a simple maturity guarantee, with term T-years, and with a payout that depends only on F_T. Let $g(T, F_T)$ be the total payoff of the VA – that is, the policyholder's fund plus any additional payments arising from the embedded option. For example $g(T, F_T) = \max(G, F_T)$ where G is the guaranteed payout at time T. Our first objective in this paper is to compute the following value

$$P_{0,F_0} := E_Q[e^{-rT} g(T, F_T)],$$

which corresponds to the market value (or no-arbitrage price) of the VA at time 0 for its payoff $g(T, F_T)$ paid at time T. As we discuss above, for simplicity, we ignore lapses and surrenders, and value the contract as a definite term investment.

As the fee is only paid when the fund value F_t is below the level B, we introduce the occupation time

$$\Gamma^-_{t,F}(B) = \int_0^t \mathbb{1}_{\{F_s<B\}} ds, \tag{3}$$

which corresponds to the total time in the interval $(0, t)$ for which the fund value lies below the fee barrier B.

Proposition 2.1 (No-Arbitrage Price). *Let* $\theta_t := -\alpha + \frac{c}{\sigma}\mathbb{1}_{\{F_t<B\}}$ *where* $\alpha = \frac{1}{\sigma}(r - \frac{\sigma^2}{2})$ *and let* $K = \frac{1}{\sigma}\ln(\frac{B}{F_0})$. *Define a probability* $\widetilde{Q}$ *by its Radon-Nikodym derivative with respect to the risk-neutral probability*[1] Q

$$\left(\frac{d\widetilde{Q}}{dQ}\right)_T = \exp\left\{\int_0^T \theta_s dW_s - \frac{1}{2}\int_0^T \theta_s^2 ds\right\}. \tag{4}$$

https://doi.org/10.1017/asb.2014.13 Published online by Cambridge University Press

Then,

$$P_{0,F_0} = e^{-rT-\frac{\alpha^2}{2}T} E_{\widetilde{Q}}\left[\exp\left\{\alpha\widetilde{W}_T + \beta\Gamma^-_{T,\widetilde{W}}(K) - \frac{c}{\sigma}\int_0^T \mathbb{1}_{\{\widetilde{W}_s<K\}} d\widetilde{W}_s\right\} g\left(T, F_0 e^{\sigma\widetilde{W}_T}\right)\right], \tag{5}$$

where $\beta = \frac{c}{\sigma^2}(r - \frac{\sigma^2}{2} - \frac{c}{2})$ *and* $\widetilde{W}_t = W_t - \int_0^t \theta_s ds$ *is a* $\widetilde{Q}-$*Brownian motion, and* $\Gamma^-_{T,\widetilde{W}}(K)$ *is defined similarly as in* (3) *and represents the occupation time of* $\widetilde{W}$ *below* K.

Proof. This proof is based on Section 5 of Karatzas and Shreve (1984). Recall that W_t is a standard Brownian motion under the risk-neutral probability measure Q. By Girsanov's Theorem it is clear that $\widetilde{W}_t = W_t - \int_0^t \theta_s ds = W_t + \alpha t - \frac{c}{\sigma}\int_0^t \mathbb{1}_{\{F_s<B\}} ds$ is a standard Brownian motion under the probability measure $\widetilde{Q}$ defined in (4). Furthermore, observe that $F_t = F_0 e^{\sigma\widetilde{W}_t}$, for $0 \leqslant t \leqslant T$, so that the occupation time of the fund below the level B can be rewritten using the Brownian motion $\widetilde{W}$ under the new probability measure $\widetilde{Q}$

$$\{F_t < B\} = \left\{\widetilde{W}_t < K\right\}, \tag{6}$$

where $K := \frac{1}{\sigma}\ln(\frac{B}{F_0})$. In other words, the occupation time $\Gamma^-_{t,F}(B)$ (given in (3)), is also $\Gamma^-_{t,\widetilde{W}}(K) = \int_0^t \mathbb{1}_{\{\widetilde{W}_s<K\}} ds$ (using the equivalence in (6)). Note that we can express the Radon-Nikodym derivative $(\frac{d\widetilde{Q}}{dQ})_T$ as

$$\begin{aligned}\left(\frac{d\widetilde{Q}}{dQ}\right)_T &= \exp\left\{\int_0^T \theta_s(d\widetilde{W}_s + \theta_s ds) - \frac{1}{2}\int_0^T \theta_s^2 ds\right\} \\ &= \exp\left\{\int_0^T \theta_s d\widetilde{W}_s + \frac{1}{2}\int_0^T \theta_s^2 ds\right\} \\ &= \exp\left\{-\alpha\widetilde{W}_T + \frac{c}{\sigma}\int_0^T \mathbb{1}_{\{\widetilde{W}_s<K\}} d\widetilde{W}_s + \frac{\alpha^2}{2}T - \beta\Gamma^-_{T,\widetilde{W}}(K)\right\} \end{aligned} \tag{7}$$

where

$$\beta = \frac{c}{\sigma^2}\left(r - \frac{\sigma^2}{2} - \frac{c}{2}\right).$$

To get the last equality, (7), we used the fact that

$$\theta_s^2 = \alpha^2 - \frac{2c}{\sigma^2}\left(r - \frac{\sigma^2}{2} - \frac{c}{2}\right)\mathbb{1}_{\{\widetilde{W}_s<K\}}.$$

https://doi.org/10.1017/asb.2014.13 Published online by Cambridge University Press

This makes it possible to price a claim $g(T, F_T)$ on the fund at maturity, under the measure $\widetilde{Q}$:

$$P_{0,F_0} = E_Q\left[e^{-rT}g(T,F_T)\right] = E_{\widetilde{Q}}\left[\left(\frac{dQ}{d\widetilde{Q}}\right)_T e^{-rT}g(T,F_T)\right]$$
$$= E_{\widetilde{Q}}\left[e^{-rT+\alpha\widetilde{W}_T-\frac{c}{\sigma}\int_0^T \mathbb{1}_{\{\widetilde{W}_s<K\}}d\widetilde{W}_s-\frac{\alpha^2}{2}T+\beta\Gamma^-_{T,\widetilde{W}}(K)}g\left(T, F_0e^{\sigma\widetilde{W}_T}\right)\right],$$

which ends the proof of Proposition 2.1. □

From Proposition 2.1, it is clear that we need the joint distribution of $(\widetilde{W}_t, \int_0^t \mathbb{1}_{\{\widetilde{W}_s<K\}}d\widetilde{W}_s, \Gamma^-_{t,\widetilde{W}}(K))$ under $\widetilde{Q}$ to evaluate the initial premium P_{0,F_0}. Using the Tanaka formula for the Brownian local time, $L_{t,\widetilde{W}}(K)$ (see Proposition 6.8 of Karatzas and Shreve 1991), we can write

$$\int_0^t \mathbb{1}_{\{\widetilde{W}_s<K\}}d\widetilde{W}_s = L_{t,\widetilde{W}}(K) - (\widetilde{W}_t - K)^- + (0-K)^-, \tag{8}$$

where we recall that $x^- = \max(-x, 0)$ so that $(0-K)^- = \max(K, 0)$. Thus, to price the Variable Annuity contract, we compute the expectation under $\widetilde{Q}$ using the trivariate density[2] $\widetilde{Q}(\widetilde{W}_T \in dx, L_{T,\widetilde{W}} \in dy, \Gamma^-_{T,\widetilde{W}}(K) \in dz)$, which we recall in the following proposition.

Proposition 2.2 (Distribution of $(\widetilde{W}_T, L_{T,\widetilde{W}}(K), \Gamma^-_{T,\widetilde{W}}(K))$ under $\widetilde{Q}$). *Denote by* $f(x,y,z) = \widetilde{Q}(\widetilde{W}_T \in dx, L_{T,\widetilde{W}}(K) \in dy, \Gamma^-_{T,\widetilde{W}}(K) \in dz)$ *and by* $h(s,x) := \frac{|x|}{\sqrt{2\pi s^3}}\exp\{-\frac{x^2}{2s}\}$.

For $K \geqslant 0$, *the joint density of* $(\widetilde{W}_T, L_{T,\widetilde{W}}(K), \Gamma^-_{T,\widetilde{W}}(K))$ *is given by*

$$f(x,y,z) = \begin{cases} 2h(z, y+K)\,h(T-z, y-K+x), & x>K,\ y>0,\ 0<z<T \\ 2h(T-z, y)\,h(z, y+2K-x), & x<K,\ y>0,\ 0<z<T \\ \frac{1}{\sqrt{2\pi T}}\left(e^{-\frac{x^2}{2T}} - e^{-\frac{(x-2K)^2}{2T}}\right), & x<K,\ y=0,\ z=T. \end{cases}$$

For $K < 0$, *the joint density of* $(\widetilde{W}_T, L_{T,\widetilde{W}}(K), \Gamma^-_{T,\widetilde{W}}(K))$ *is given by*

$$f(x,y,z) = \begin{cases} 2h(z,y)\,h(T-z, y+x-2K), & x>K,\ y>0, 0<z<T \\ 2h(T-z, y-K)\,h(z, y-x+K), & x<K,\ y>0, 0<z<T \\ \frac{1}{\sqrt{2\pi T}}\left(e^{-\frac{x^2}{2T}} - e^{-\frac{(x-2K)^2}{2T}}\right), & x>K,\ y=0, z=0. \end{cases}$$

Proof. Section 3 of Karatzas and Shreve (1984) gives the trivariate density of $(Z_T, L_{T,Z}(0), \Gamma^+_{T,Z}(0))$ for a Brownian motion Z with given initial value Z_0. To make use of this result, we express $L_{T,\widetilde{W}}(K)$ and $\Gamma^-_{T,\widetilde{W}}(K)$ in terms of $(Z_T, L_{T,Z}(0), \Gamma^+_{T,Z}(0))$ where Z is a Brownian motion such that $Z_0 = K$. A detailed proof of Proposition 2.2 can be found in Appendix 7. □

https://doi.org/10.1017/asb.2014.13 Published online by Cambridge University Press

From the expression of the price in Proposition 2.1, and from Tanaka's formula in (8), we observe that the trivariate density given in Proposition 2.2 can be used to price any European VA. The following proposition summarizes this result.

Proposition 2.3 (No-arbitrage price). *The no-arbitrage price at time* 0 *of a VA contract with payoff* $g(T, F_T)$ *can be computed as*

$$P_{0,F_0} = e^{-rT-\frac{\alpha^2}{2}T} E_{\widetilde{Q}}\left[e^{\alpha \widetilde{W}_T+\beta\Gamma^-_{T,\widetilde{W}}(K)} - \frac{cL_{T,\widetilde{W}}(K)}{\sigma}+\frac{c}{\sigma}\left((\widetilde{W}_T-K)^- - \max(K,0)\right) g\left(T, F_0 e^{\sigma \widetilde{W}_T}\right)\right]$$

using the trivariate density of $(\widetilde{W}_T, L_{T,\widetilde{W}}(K), \Gamma^-_{T,\widetilde{W}}(K))$ *given in Proposition 2.2.*

3. Examples

In this section we apply Proposition 2.3 to price a VA with level GMMB, and a VA with a level GMDB, with state-dependent fees payable when the policyholder's fund falls below some pre-specified level.

3.1. State-dependent fee rates for a VA with GMMB

The next proposition gives the price of a variable annuity with a guaranteed minimum maturity benefit (GMMB). It pays out the maximum between the value of the fund at maturity, F_T, and a guaranteed amount G at a given maturity date T. The payoff is then

$$g(T, F_T) := \max(G, F_T) = F_T \mathbb{1}_{\{G \leqslant F_T\}} + G\mathbb{1}_{\{F_T < G\}}. \tag{9}$$

Let $P^M_{0,F_0}(T, B, G, c)$ be the no-arbitrage price of this VA at inception assuming that a fee c is taken from the fund continuously, whenever the fund value drops below the level B.

Proposition 3.1 (No-arbitrage price for GMMB with guarantee G). *The initial price* $P^M_{0,F_0}(T, B, G, c)$ *of the GMMB contract is given by*

$$P^M_{0,F_0}(T, B, G, c) = e^{-rT-\frac{\alpha^2}{2}T-\frac{c}{\sigma}\max(0,K)}(F_0 A_T + G D_T), \tag{10}$$

where

$$\begin{cases} A_T = \int_H^{+\infty}\int_0^{+\infty}\int_0^T e^{(\sigma+\alpha)x+\beta z-\frac{c}{\sigma}y+\frac{c}{\sigma}(x-K)^-} f(x, y, z)dz\, dy\, dx \\ D_T = \int_{-\infty}^{H}\int_0^{+\infty}\int_0^T e^{\alpha x+\beta z-\frac{c}{\sigma}y+\frac{c}{\sigma}(x-K)^-} f(x, y, z)dz\, dy\, dx \end{cases}$$

with $H := \frac{1}{\sigma}\ln(\frac{G}{F_0})$. *Recall that* $K := \frac{1}{\sigma}\ln(\frac{B}{F_0})$ *where* B *is the level that triggers the payment of the continuous fee* c.

https://doi.org/10.1017/asb.2014.13 Published online by Cambridge University Press

Note that the maturity benefit is usually conditional on the survival of the policyholder. In Proposition 3.1, we ignore mortality risk. In order to add mortality risk, it would suffice to multiply the value of the benefit by the probability that the policyholder survives to maturity.

Proof. Using Proposition 2.3 and the expression of the payoff $g(T, F_T)$ given in (9), the result follows. □

We now replace the density f by its expression given in Proposition 2.2 and we can simplify the expressions for A_T and D_T. To do so we distinguish four cases. Details are given in Appendix 7.

The fair fee rate $c^\star$ is then computed such that, for a given premium P, the VA value is equal to the premium paid, that is

$$P = P^M_{0,F_0}(T, B, G, c^\star). \tag{11}$$

We ignore here other types of expenses and assume there are no other guarantees attached. This is equivalent to finding the fee rate such that, under the risk neutral probability Q, the expected present value of the aggregate income is equal to that of the aggregate payout. For more details, see Marshall *et al.* (2010).

Proposition 3.1 gives integral expressions for the no-arbitrage price of the GMMB. In the special case when the fee is paid only when the option is in-the-money, in other words when $B = G$, the price of the GMMB is given in the next proposition.

Proposition 3.2 (No-arbitrage price for GMMB with guarantee G when $B = G$). *When the fee is only paid when the option is in-the-money, that is $B = G$, the initial price $P^M_{0,F_0}(T, B, G, c)$ of the GMMB contract is given by*

$$P^M_{0,F_0}(T, B, G, c) = e^{-rT-\frac{\alpha^2}{2}T-\frac{c}{\sigma}\max(0,K)}(F_0 C_1 + G(C_2 + C_3)),$$

where

$$C_3 = \left(\frac{B}{F_0}\right)^{\frac{c}{\sigma^2}} e^{\left(r-\frac{\sigma^2}{2}\right)^2\frac{T}{2\sigma^2}}\left(N\left(\frac{\ln(\frac{B}{F_0})-\eta T}{\sigma\sqrt{T}}\right) - (\tfrac{B}{F_0})^{\frac{2\eta}{\sigma^2}} N\left(\frac{\ln(\frac{F_0}{B})-\eta T}{\sigma\sqrt{T}}\right)\right)$$

with $\eta = r - \frac{\sigma^2}{2} - c$ and $N(\cdot)$ stands for the cdf of the standard normal distribution. For $K \geqslant 0$ (that is $G = B \geqslant F_0$),

$$\begin{cases} C_1 = \int_K^\infty \int_0^\infty \int_0^T \frac{(y+K)(y-K+x)}{\pi\sqrt{z^3(T-z)^3}} e^{(\alpha+\sigma)x+\beta z-\frac{cy}{\sigma}-\frac{(y+K)^2}{2z}-\frac{(y-K+x)^2}{2(T-z)}} dz\, dy\, dx \\ C_2 = \int_{-\infty}^K \int_0^\infty \int_0^T \frac{y(y+2K-x)}{\pi\sqrt{z^3(T-z)^3}} e^{(\alpha-\frac{c}{\sigma})x+\beta z-\frac{cy}{\sigma}+\frac{cK}{\sigma}-\frac{y^2}{2(T-z)}-\frac{(y+2K-x)^2}{2z}} dz\, dy\, dx, \end{cases}$$

https://doi.org/10.1017/asb.2014.13 Published online by Cambridge University Press

and for $K < 0$ *(that is* $G = B < F_0$*)*

$$\begin{cases} C_1 = \int_K^\infty \int_0^\infty \int_0^T \frac{y|y+x-2K|}{\pi\sqrt{z^3(T-z)^3}} e^{(\alpha+\sigma)x+\beta z-\frac{cy}{\sigma}-\frac{y^2}{2z}-\frac{(y-2K+x)^2}{2(T-z)}} dz\, dy\, dx \\ C_2 = \int_{-\infty}^K \int_0^\infty \int_0^T \frac{|y-K|(y-x+K)}{\pi\sqrt{z^3(T-z)^3}} e^{\left(\alpha-\frac{c}{\sigma}\right)x+\beta z-\frac{cy}{\sigma}+\frac{cK}{\sigma}-\frac{(y-K)^2}{2(T-z)}-\frac{(y+K-x)^2}{2z}} dz\, dy\, dx. \end{cases}$$

Details can be found in Appendix 7 in the case when $H = K$.

3.2. State-dependent fee rates for a VA with GMDB

In this section, we price a guaranteed minimum death benefit rider, which guarantees a minimum amount if the policyholder dies before maturity of the contract. Let us first introduce some notation. Let τ_x be the random variable representing the future lifetime of a policyholder aged x and let $\lfloor\tau\rfloor_x$ denote his curtate future lifetime.[3] Let also ${}_tp_x = P(\tau_x > t)$ be the probability that a policyholder aged x survives t years and $q_{x+t} = P(t < \tau_x \leqslant t+1 | \tau_x > t)$ be the probability that a policyholder aged $x+t$ dies during year t. We assume that the GMDB is paid at the end of the year of death, only if death occurs strictly before the maturity T of the contract. The fair price at time 0 of the GMDB is given by

$$E_Q\left[e^{-r(\lfloor\tau\rfloor_x+1)} \max(G_{\lfloor\tau\rfloor_x+1}, F_{\lfloor\tau\rfloor_x+1}) \mathbb{1}_{\{(\lfloor\tau\rfloor_x+1)\leqslant T\}} + e^{-rT} F_T \mathbb{1}_{\{(\lfloor\tau\rfloor_x+1)>T\}}\right],$$

where G_t is the guarantee paid at time t. Assume that mortality rates are deterministic and that death benefits are paid at the end of the year of death. Under those assumptions, the no-arbitrage price of the payoff of the variable annuity with a GMDB rider, denoted by $P^D_{0,F_0}(T, B, c)$, can be computed using the price for a GMMB obtained in the previous section. It can be seen as a weighted sum of GMMB prices. See, for example, Dickson *et al.* (2009) for details.

Proposition 3.3 (Fair price for GMDB). *Under the assumption of deterministic mortality rates, the fair price of a GMDB,* $P^D_{0,F_0}(T, B, c)$ *is given by*

$$\begin{aligned} P^D_{0,F_0}(T, B, c) &= E_Q\left[\sum_{t=0}^{T-1} {}_tp_x\, q_{x+t}\, e^{-r(t+1)} \max(G_t, F_t) + {}_Tp_x\, e^{-rT} F_T\right] \\ &= \sum_{t=0}^{T-1} {}_tp_x\, q_{x+t}\, P^M_{0,F_0}(t, B, G_t, c) + {}_Tp_x\, E_Q\left[e^{-rT} F_T\right], \end{aligned}$$

where $P^M_{0,F_0}(t, B, G_t, c)$ *is the market value of a variable annuity of maturity* t *with a GMMB rider with guaranteed amount* G_t*. We also have that*

$$E_Q\left[e^{-rT} F_T\right] = F_0 e^{-rT-\frac{\alpha^2}{2}T} A^D_T,$$

https://doi.org/10.1017/asb.2014.13 Published online by Cambridge University Press

with

$$A_D^D = \int_{-\infty}^{\infty} \int_0^{\infty} \int_0^T e^{(\alpha+\sigma)x+\beta z-\frac{c}{\sigma}y+\frac{c}{\sigma}\left((x-K)^- - \max(x,0)\right)} f(x, y, z)dz\, dy\, dx,$$

where $f(x, y, z)$ is defined as in Proposition 2.2.

Proof. This result is immediate using (5) and Proposition 2.1. □

The fair fee rate for the GMDB benefit is such that the initial premium is equal to the expected value under Q of the discounted payoff. That is, for a given premium P, and assuming the GMDB is the only guarantee, $c^\star$ satisfies

$$P = P^D(T, B, c^\star). \tag{12}$$

4. Numerical results

In this section, we compare the state-dependent fee rates for VAs with guarantees for a range of parameters and settings. Unless otherwise indicated, we let $r = 0.03$ and $\sigma = 0.2$; the contract term is 10-years. We assume that the initial premium is $P = 100$, with no initial fixed expense, so that $e = 0$. The policyholder is assumed to be aged 50, and mortality is assumed to follow the Gompertz model, with force of mortality $\mu_y = 0.00002\, e^{0.1008y}$.

Let g denote the guaranteed roll-up rate, which relates the guarantee G applying at T, and the initial premium P, as $G = P\, e^{gT}$. When $g = 0$, the guarantee is a 'return-of-premium' guarantee, so that $G = P$. For now, we assume that the fee is paid only below the guaranteed level ($B = G$).

In Figure 1, we show the market value of the VA with a GMMB as a function of the fee rate c. The point on the x-axis at which the curve of the no-arbitrage price of the payout crosses the line $y = 100$ corresponds to the fair fee rate (as computed in (11)). We show different curves for $g = 0\%$, $g = 1\%$ and $g = 2\%$. As expected, the fair fee increases as the guaranteed payoff increases, reaching almost 0.1 when the guaranteed roll-up rate g is 2%.

Figure 1 also illustrates, though, that the relationship between the guarantee level and fee rate is not completely straightforward. We note that the curves for $g = 0\%$ and $g = 1\%$ cross at around $c = 0.09$. At first this observation may seem counter-intuitive; for a conventional, static fee structure, a lower guarantee would lead to a lower fee rate. However, in our case, as the fee is paid only when the option is in the money, the effect of the guarantee roll-up rate g on the state-dependent fee rate c is not so straightforward. A higher guarantee generates a higher occupation time, so that the cost is spread over a longer payment period, which in some cases will lead to a lower fee rate.

Figure 2 illustrates the sensitivity of the dynamic fair fee rate with respect to the volatility σ and the term T. We find that the fair fee rate increases with σ and decreases with T. In relative terms, the maturity of the GMMB does not seem to affect the sensitivity of the fair fee to changes in σ.

https://doi.org/10.1017/asb.2014.13 Published online by Cambridge University Press

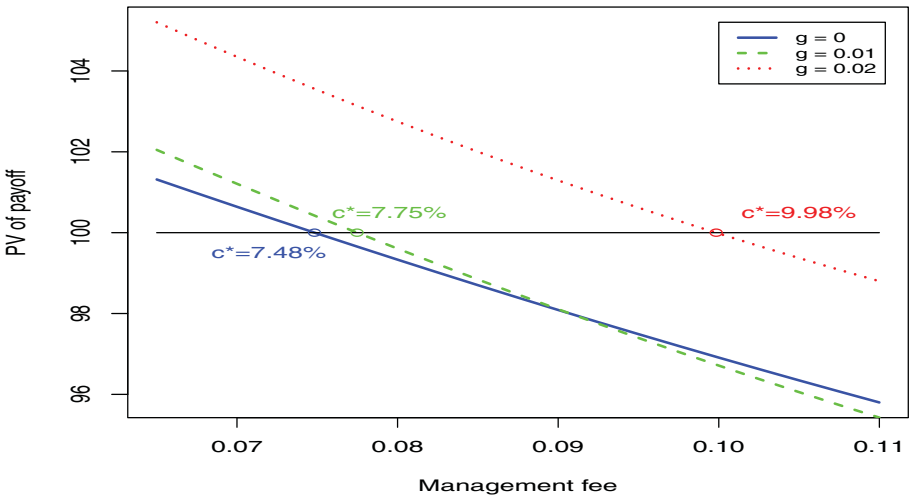


FIGURE 1: No arbitrage price of a VA with a 10-year GMMB as a function of c, with \$100 initial premium and guaranteed roll-up rates $g = 0\%$, 1% and 2%. (Color online)

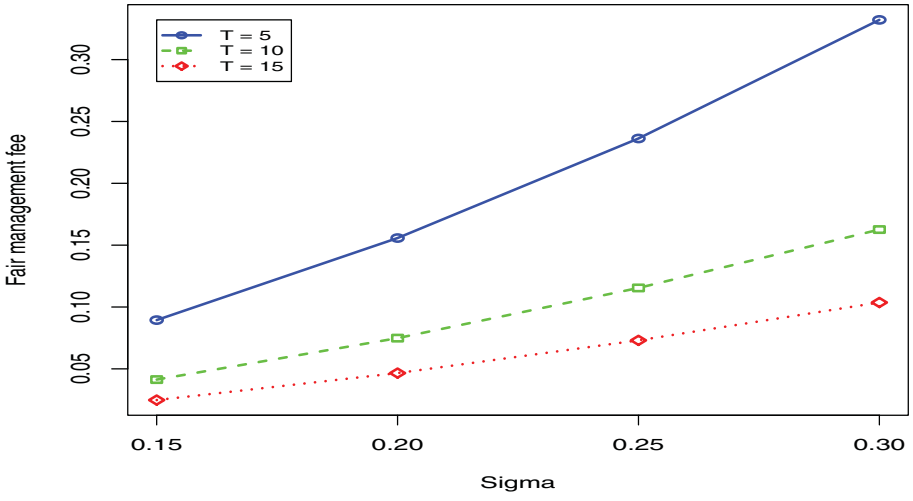


FIGURE 2: Sensitivity of fair fee rates for a GMMB with respect to volatility (from $\sigma = 15\%$ to $\sigma = 30\%$) and contract term, $T = 5$, 10 or 15 years for a contract with $g = 0$. (Color online)

Table 1 considers two contracts; a GMMB and a GMDB, both with fees paid only when the guarantee is in the money. For comparison purposes, results pertaining to contracts with a continuous fee rate are also illustrated. As expected, the fair fee rate for the GMMB is much higher when it is paid only when the guarantee is in the money. The difference is much more significant for shorter maturities. The GMDB results in a lower fair fee rate since the benefit is paid only if the policyholder dies during the life of the contract. In this case, since the

https://doi.org/10.1017/asb.2014.13 Published online by Cambridge University Press

TABLE 1

FAIR FEE RATES (%) FOR THE GMMB AND GMDB WITH RESPECT TO MATURITY T WHEN $B = G$. "(S-D FEE)" REFERS TO STATE-DEPENDENT FEES AND "(CST FEE)" REFERS TO A CONSTANT FEE.

Type of Contract	$T = 5$	$T = 7$	$T = 10$	$T = 12$	$T = 15$
GMMB (s-d fee)	15.58	11.01	7.48	6.08	4.66
GMMB (cst fee)	3.53	2.43	1.58	1.24	0.91
GMDB (s-d fee)	0.1	0.12	0.17	0.21	0.27
GMDB (cst fee)	0.04	0.04	0.06	0.06	0.08

TABLE 2

FAIR FEE RATES (%) FOR THE 10-YEAR GMMB WITH RESPECT TO VOLATILITY σ WHEN $B = G$. "S-D" REFERS TO STATE-DEPENDENT FEES AND "CST" REFERS TO A CONSTANT FEE.

Type of Fee	$\sigma = 0.15$	$\sigma = 0.2$	$\sigma = 0.25$	$\sigma = 0.3$
S-D	4.13	7.48	11.54	16.26
CST	0.86	1.58	2.38	3.22

policyholder is 50 years old, the probability of having to pay the benefit is quite low. Observe that the GMMB fair rate (as computed in (11)) is decreasing with respect to T whereas the GMDB fair rate (as computed in (12)) is increasing with respect to T. This is explained by the fact that the payoff of the GMDB benefit is a weighted sum of GMMB payoffs with different maturities (see Proposition 3.3), with weights linked to the probability that the policyholder dies during a given year. By increasing the maturity of the contract, we extend the period during which the death benefit can be paid, thus increasing the probability that it is paid at any given point before maturity. In addition, since the mortality rates increase with time, the fair fee rate is calculated assuming that less premium will be collected in later years since some policyholders will have died. For this reason, the fee rate needs to be higher for longer maturities.

Table 2 illustrates the sensitivity of the state-dependent fee to the volatility assumption, compared to that of the static fee. Although the state-dependent fair fee is consistently higher, it is not significantly more sensitive to a small increase in volatility. In fact, when the volatility increases from 0.15 to 0.2, the static fee increases by about 84% while the state-dependent fee is 81% higher. However, to keep up with larger increases in volatility, the state-dependent fee rises faster. When the volatility doubles from 0.15 to 0.3, the state-dependent fee is multiplied by 3.94 while its constant counterpart is only 3.74 times higher. Thus, as σ increases, the sensitivity of the state-dependent fee to the volatility also increases, compared to the static fee.

For a simple return-of-premium GMMB, we see in Figure 1 that the fair fee under the dynamic fee structure is around 7.5% of the fund. This is a very high rate, and is not practical in marketing terms. This result though, is not that surprising. The static fee under the same assumptions is around 2%. The result

https://doi.org/10.1017/asb.2014.13 Published online by Cambridge University Press

shows that 2% of the fund paid throughout the policy term has the same value as 7.5% of the fund paid only when the fund is low. Thus, the state-dependent fee is paid for a shorter time, and is paid on a smaller fund.

If we accept that such a high fee is unlikely to be feasible, we can move the payment barrier higher, so that the option being funded at least has significant value, even though it is out-of-the-money. That is, the policyholder would pay for the option when it is close to the money, and cease paying when it moved further away from the money. This adjustment retains the advantage of the differential fee structure with respect to disintermediation risk when the guarantee is far out-of-the-money, but, by increasing the occupation time, offers a more reasonable rate for policies which are close to the money.

Suppose that the dynamic fee is paid when the fund value is below $B = (1 + \kappa)G$, for some $\kappa \geqslant 0$. Figure 3 illustrates the fair fee rate for different values of κ. We see a dramatic effect on the fee rate for relatively low values of κ. Increasing the barrier from G to $1.2G$ causes the fair fee rate to decrease from 7.48% to 3.77%, which may improve the marketability of the contract significantly. Moreover, fixing the barrier at $1.34G$ or higher causes the fair fee to drop below 3.00%, which brings the risk-neutral drift of the fund value process back above 0. Asymptotically, the state-dependent fee rate of the GMMB converges to the static fee rate which is not state-dependent. Increasing the payment barrier decreases the fair fee rate, but it also adds an incentive to surrender. In fact, if the fee is paid while the guarantee is out-of-the-money, it may be optimal for the policyholder to surrender when the fund value is close to the guaranteed amount. This might happen when the expected value of the discounted future fees is more than the value of the financial guarantee. However, even with an increased payment barrier, the incentive to surrender for very high fund values is still eliminated, thus decreasing the value of the surrender option. We discuss this further in Section 5.

5. Analysis of the surrender incentive

One of the motivations of state-dependent fees is to reduce the incentive to surrender the VA contract when the fund value is high, that is when the guarantee is out-of-the-money. It is clear that if the policyholder pays a constant fee, at some point the fund value could become so high that the option is essentially valueless, while the option fees have considerable drag on the fund accumulation. It would be rational to surrender, in order to stop paying the fees. However, if the fee is state-dependent, when the fund value is high the policyholder no longer pays fees, and will gain no benefit from surrendering.

In order to demonstrate this numerically, we consider the relationship between the value of the financial guarantee at maturity and the value of the future fees. At some time t before the maturity date T, the policyholder may surrender, with a payoff of F_t (we assume no surrender charge). If the policyholder holds

https://doi.org/10.1017/asb.2014.13 Published online by Cambridge University Press

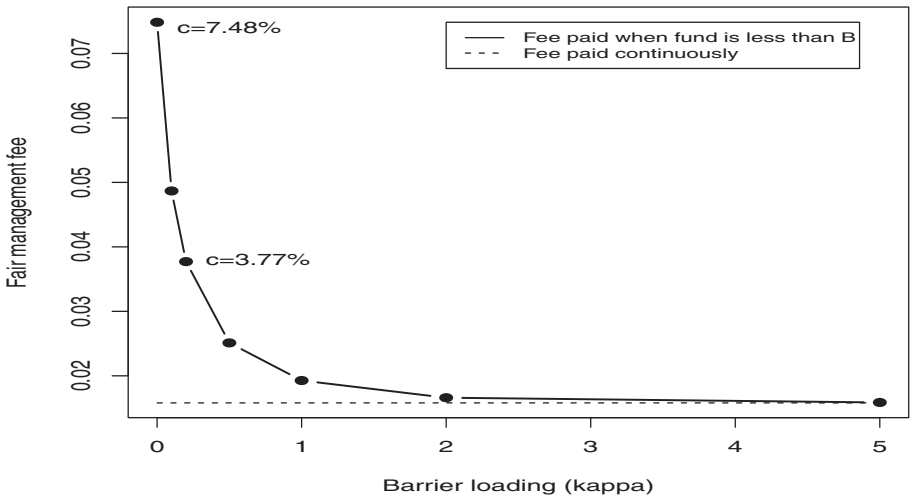


FIGURE 3: State-dependent fee rates for a GMMB, for κ=0 to 5, $r = 0.03$, $T = 10$.

the contract until maturity, she will receive the fund at T plus the additional guarantee payout (if any) – that is, she will receive $F_T + (G - F_T)^+$.

We will show, in Proposition 5.1, that it is optimal to retain the contract at time t if at that time the value of the future fees is less than the value of the maturity guarantee, which is the value of the put option payout $(G - F_T)^+$. However, the reverse is not generally true. That is, it may not be optimal to surrender the contract even when the value of the future fees is greater than the value of the maturity put option. The reason is that the surrender option is American-style, and if the value of the fees is not too much greater than the option value, it may be optimal to postpone the decision to surrender.

The option to postpone the surrender decision when fees are state-dependent is somewhat complex to analyze, putting a full analysis of optimal surrender strategies beyond the scope of this paper. However, a comparison of the value of the maturity put option benefit with the value of the future fees does allow some comparison of surrender incentives. We will demonstrate through an example how the state-dependent fee reduces the surrender incentive for out-of-the-money options, compared with a constant fee.

Proposition 5.1 (Sufficient conditions to retain the contract at time t).
Consider a GMMB contract that can be surrendered and denote the present value of the maturity benefit at time t by $P(t, F_t)$. *A first sufficient condition to retain the contract at time t is given by*

$$P(t, F_t) := E_t[e^{-r(T-t)} \max(F_T, G)] \geqslant F_t. \tag{13}$$

https://doi.org/10.1017/asb.2014.13 Published online by Cambridge University Press

Let $m(t, F_t)$ denote the value at t of the future fees, and let $p(t, F_t)$ denote the value at t of the maturity put option, so that

$$m(t, F_t) := F_t - E_t[e^{-r(T-t)} F_T] \quad \text{and } p(t, F_t) := E_t\left[e^{-r(T-t)}(G - F_T)^+\right]. \tag{14}$$

Then a second sufficient condition to retain the contract at time t is given by

$$p(t, F_t) \geqslant m(t, F_t). \tag{15}$$

Proof. Fix $t \in [0, T)$ and denote the value at time t of the full contract, including the surrender option, by $V(t, F_t)$. The surrender region is $\mathcal{A}$, defined as the region of F_t where $F_t > V(t, F_t)$ (as the surrender benefit is equal to F_t).

The full contract value can never be less than the present value of the maturity benefit, so $V(t, F_t) \geqslant P(t, F_t)$.

Suppose also that (13) holds, so that $P(t, F_t) \geqslant F_t$; then

$$V(t, F_t) \geqslant P(t, F_t) \geqslant F_t \Rightarrow F_t \notin \mathcal{A}$$

that is, F_t is outside the surrender region when $P(t, F_t) \geqslant F_t$. Condition (13) is proved.

Condition (15) follows from (13), as follows

$$\begin{aligned} P(t, F_t) &= E_t[e^{-r(T-t)} F_T] + E_t[e^{-r(T-t)}(G - F_T)^+] \\ &= F_t - m(t, F_t) + p(t, F_t) \\ &\Rightarrow P(t, F_t) - F_t = p(t, F_t) - m(t, F_t) \end{aligned}$$

then

$$P(t, F_t) > F_t \Leftrightarrow p(t, F_t) > m(t, F_t)$$

and condition (15) is proved. □

To illustrate the effect of the fee structure on the incentive to surrender, we consider a 5-year variable annuity contract. All other parameters are the same as in Section 4. We assume that the fee rate is set at the fair rate to fund the maturity benefit without allowing for surrenders.

5.1. Constant fee

We first consider the case where a level fee is paid continuously, regardless of the value of the fund. Figure 4 shows the difference between the value of the financial guarantee and the expected value of the future fees.

As expected, at any time, there is a fund value above which the future fees are worth more than the financial guarantee at maturity. This can be clearly seen in the panels in Figure 4 which give snapshots at $t = 1$, 2 and 4. The curves all cross at a single point. Below this point, the option value is greater than the value of the future fees, and the policyholder should remain in the contract. Above the cross over point, there is another threshold, above which it would be optimal to

https://doi.org/10.1017/asb.2014.13 Published online by Cambridge University Press

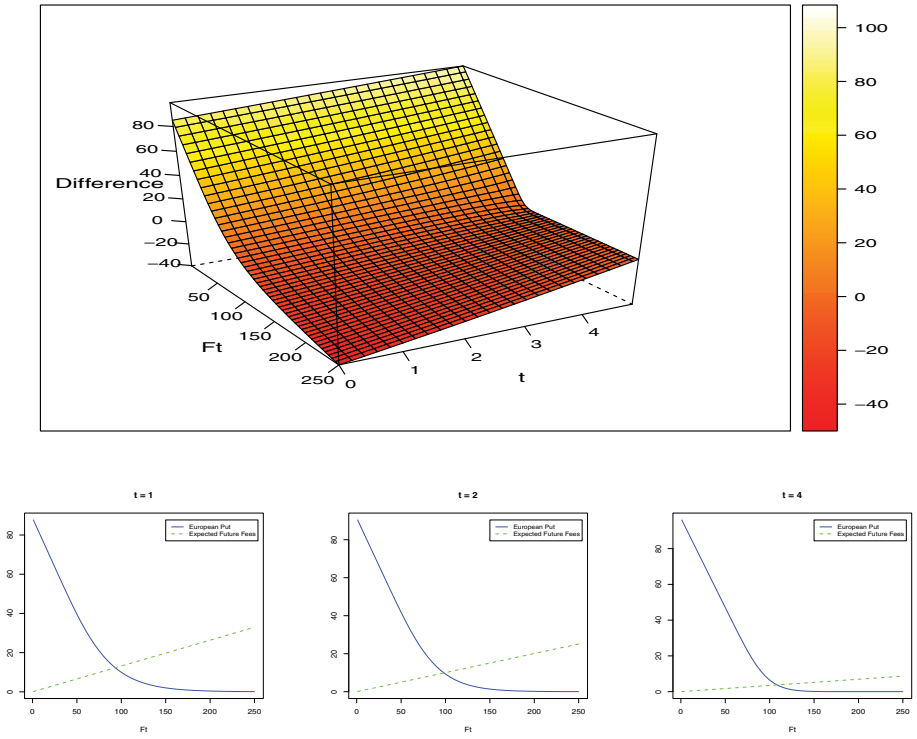


FIGURE 4: Difference at time t (when the fund value is F_t) between the value of the financial guarantee at maturity T and the expected value at t of the discounted future fees for a 5-year GMMB contract with $G = F_0 = 100$ and $\alpha = 3.53\%$ paid continuously. Snapshots at $t = 1$, $t = 2$ and $t = 4$. (Color online)

surrender the policy. Between the cross over point and the surrender threshold, it is optimal to postpone the surrender decision. It is well documented that when the fee is constant, the optimal surrender strategy is to lapse the contract whenever the fund value is above a certain level (see, for example Bernard *et al.* 2014; Milevsky and Salisbury 2001). In this example the threshold is 125.2 at $t = 1$, 126.4 at $t = 2$ and 123.7 at $t = 4$ (these numbers were calculated following the algorithm in Bernard *et al.* 2014). That is, for example, at $t = 1$, the policyholder should surrender if the fund value is greater than 125.2, which has a probability of around 20%.

5.2. State-dependent fee

Next, we consider a contract with a state-dependent fee. We first assume that the payment barrier is $B = G = F_0$. In this case the state-dependent fee is very high. The difference between the value of the financial guarantee and the expected value of the future fees is shown in Figure 5. The smaller panels show the relationship between the future fee value and the option value at $t = 1, 2$ and 4, in terms of F_t.

https://doi.org/10.1017/asb.2014.13 Published online by Cambridge University Press

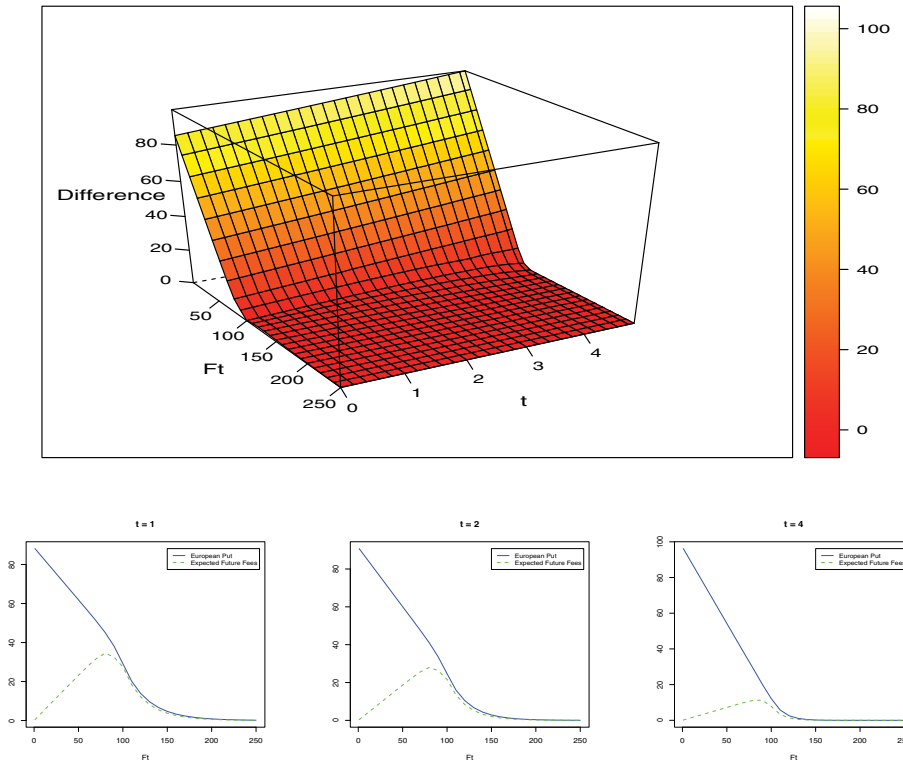


FIGURE 5: Difference at time t (when the fund value is F_t) between the value of the financial guarantee at maturity T and the expected value at t of the discounted future fees, for a 5-year GMMB contract with $B = G$, $G = F_0 = 100$ and $\alpha = 15.58\%$ paid when $F_t < B$. Snapshots at $t = 1$, $t = 2$ and $t = 4$. (Color online)

One major difference in comparison to the constant fee case is that the surface does not drop below 0. That means that in this example, there is no fund value at any date where the value of the future fees exceeds the value of the option. Hence, following Proposition 5.1, it is never optimal to surrender this contract.

To complete this analysis, we consider a similar contract, but with an increased payment barrier B. We let $B = 1.4G$, again assuming $G = F_0$. In this case, the difference between the value of the financial guarantee and the expected future fee can become negative, but returns close to zero for high fund values (see Figure 6).

However, when the payment barrier is increased, there is an interval around the payment barrier where it may become optimal to surrender the contract since the policyholder can expect to pay more than he will receive (see Figure 6). This indicates that the surrender region when the fee is state-dependent has a different form than when the fee is constant; the optimal surrender strategy is no longer based on a simple threshold; there may be a surrender corridor, where surrender is optimal for F_t in the corridor, and not for higher or lower values of

https://doi.org/10.1017/asb.2014.13 Published online by Cambridge University Press

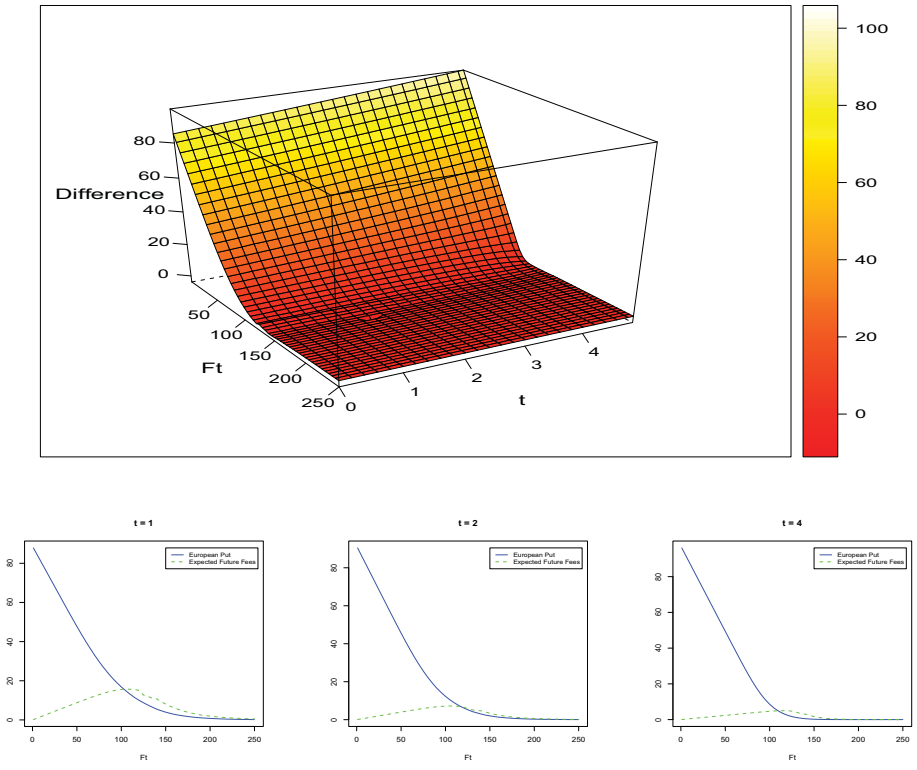


FIGURE 6: Difference at time t (when the fund value if F_t) between the value of the financial guarantee at maturity T and the expected value at t of the discounted future fees for a 5-year GMMB contract with $B = 1.4G$, $G = F_0 = 100$ and $\alpha = 4.84\%$ paid when $F_t < B$. Snapshots at $t = 1$, $t = 2$ and $t = 4$. (Color online)

F_t. However, it is apparent that the area where there is no surrender incentive, because the option value is greater than or equal to the future fee value, is much larger than for the constant fee case.

We have demonstrated through an example that the optimal surrender area for the constant fee contract is much larger than for the state-dependent fee contract. However, we have not included a full analysis of optimal surrenders, nor have we priced the contracts assuming optimal (or deterministic) surrenders. As mentioned before, the small panels of Figure 6 indicate that the optimal surrender strategy is not of the threshold type. This makes a full analysis of the optimal surrender strategy more complex than for the traditional constant fee[4].

6. MODEL RISK

In the previous sections we have assumed that fund values may be modelled by a geometric Brownian motion, which allows us to obtain solutions in integral form. However, it is well-known that this model does not provide a good fit to

https://doi.org/10.1017/asb.2014.13 Published online by Cambridge University Press

TABLE 3

REGIME-SWITCHING LOG-NORMAL PARAMETERS USED FOR MONTE CARLO SIMULATIONS.

Parameter	σ_1	σ_2	$p_{1,2}$	$p_{2,1}$
Value	0.035	0.0748	0.0398	0.3798

the empirical distribution of stock returns over longer terms. Our results in the previous section may thus be viewed as approximations for real market values. In order to test their sensitivity to model risk, we find the fair fee rate when the stock returns are assumed to follow a regime-switching log-normal (RSLN) model. Hardy (2001) shows that this model provides a better fit for the distribution of long-term stock returns, allowing for heavier tails. It also reproduces the volatility clustering observed in empirical data.

The RSLN model is based on an underlying state variable, the value of which is governed by a transition matrix P (of size 2×2 if there are 2 possible regimes). The elements of this matrix, denoted $p_{i,j}$ represent the probability that the state variable moves to state j given that it is currently in state i. Between each transition, stock returns follow a log-normal distribution whose parameters are determined by the regime indicated by the state variable. Hardy (2001) presents a two-regime model in which transitions occur monthly. These characteristics are very suitable for our purposes. In fact, while we assumed earlier that management fees are withdrawn in a continuous manner, many insurance companies collect them monthly. This also has an impact on the fair fee rate that should be charged.

In this section, in addition to testing the sensitivity of the fair fee rate to model changes, we also analyze the impact of discrete fee collection. When fees are withdrawn in discrete time, getting an analytical expression for the present value of the guarantee payoff is not always possible. For this reason, we use Monte Carlo simulations to obtain the results given in this section. We consider GMMB guarantees with maturities ranging from 5 to 15 years. To quantify the model impact and fee discretization separately, we find the fair fee rates under three models. The first case is identical to the setting presented in the previous sections; the value of the fund follows a geometric Brownian motion (GBM) and fees are paid continuously. In the second case, the return on the fund is also assumed to be log-normal, but the fees are paid monthly. The last case assumes that fund returns follow a regime-switching log-normal model and that the fees are paid monthly. We use a risk-free rate of 0.03. For the GBM, we use $\sigma = 0.14029$ to match Hardy (2001). For the RSLN, we used the parameters given in Table 3, which are taken from Hardy (2001). The volatility parameters σ_1 and σ_2 are expressed per month.

Table 4 presents the fair fee rates obtained using 5×10^6 simulations. The column RSLN corresponds to the fair rates computed by Monte Carlo simulations when the underlying index price follows a regime switching model and fees are paid monthly when the fund is in-the-money, assuming a return-of-premium guarantee. The column GBM-D corresponds to the fair rates computed by

https://doi.org/10.1017/asb.2014.13 Published online by Cambridge University Press

TABLE 4

FAIR FEE RATES (%) IN THE REGIME SWITCHING MODEL (RSLN) AND IN THE BLACK-SCHOLES MODEL WHEN FEES ARE PAID MONTHLY (GBM-D) AND WHEN FEES ARE PAID CONTINUOUSLY (GBM-C).

T	RSLN	GBM-D	GBM-C
5	7.18	7.27	7.82
10	3.43	3.44	3.57
15	2.07	2.06	2.11

Monte Carlo when the underlying index price evolves as in the Black-Scholes model and fees are paid monthly when the fund is below the return-of-premium guarantee. Finally GBM-C corresponds to the fair rates computed using Proposition 3.2 and solving (11) when the fee is paid continuously, conditional on the fund value lying below the return-of-premium guarantee.

We observe from Table 4 that the RSLN fees are close to the Black-Scholes fees, given monthly payments, which indicates that the results appear to be fairly robust with respect to model risk, particularly with respect to fat tails. A possible explanation for this may be the fact that when the fund does not perform well, the guarantee is more likely to be in the money and the fee is collected. If the guarantee is in the money more often, the fee is also paid for a longer period of time. Thus, although fatter tails may lead to potentially higher benefits to pay, they can drag the fund down and cause the fee to be paid more often. In this case, a slightly lower fee rate can still be sufficient to cover the guarantee. We also note that the Black-Scholes model with continuous fees generates slightly higher fee rates than for the monthly fees, which arises from the difference between the occupation time for the discrete case and the continuous case, with lower expected occupation time where the process is continuous.

7. CONCLUDING REMARKS

This paper finds the fair fee rate for European-type guaranteed benefits in variable annuities when the fee payment is contingent on the position of the value of the underlying fund relative to some critical level B.

This is a first step towards a dynamic state-dependent charging structure, where the fee rate depends on the fund value and the dynamic value of the embedded guarantees. We have considered pricing here, but we recognize that the fee structure could also have an important impact on hedging performance and on the surrender rates. In future research we will investigate valuation and risk management allowing for dynamic, state-dependent surrenders, and also consider the impact of a non-level fee threshold.

ACKNOWLEDGEMENTS

All authors acknowledge support from NSERC and from the Society of Actuaries Centers of Actuarial Excellence Research Grant. A. MacKay also

https://doi.org/10.1017/asb.2014.13 Published online by Cambridge University Press

acknowledges the support of the FQRNT and of the Hickman scholarship of the Society of Actuaries. The authors would like to thank A. Kolkiewicz, I. Karatzas, F. Ruez, J. Russ and A. Niemeyer for helpful discussions, as well as Daniel Bauer and three anonymous referees for their comments, which greatly improved this work.

NOTES

1. This means that $\widetilde{Q}(A) = E[\frac{d\widetilde{Q}}{dQ}\mathbb{1}_A]$ for any measurable set A.

2. Note that the terminology "density" is used but it is degenerate as there could be some probability mass on $\{L_{T,\widetilde{W}}(K) = 0, \Gamma^-_{T,\widetilde{W}}(K) = T\}$ or on $\{L_{T,\widetilde{W}}(K) = 0, \Gamma^-_{T,\widetilde{W}}(K) = 0\}$.

3. Here, $\lfloor\cdot\rfloor$ denotes the floor function.

4. Because of the shape of the surrender region, we cannot use the integral representation techniques of Kim and Yu (1996), and least squares Monte Carlo techniques (used, for example, by Bacinello *et al.* 2011) break down due to the fact that the difference between the surrender benefit and the continuation option is so small for much of the range of F_t, that numerical errors are too significant. The problem is manageable using partial differential equations, as in Dai *et al.* (2008), Chen *et al.* (2008) and Belanger *et al.* (2009), but the technical challenges involved take such analysis beyond the scope of this paper.

REFERENCES

BACINELLO, A.R., MILLOSSOVICH, P., OLIVIERI, A. and PITACCO, E. (2011) Variable annuities: A unifying valuation approach. *Insurance: Mathematics and Economics*, **49**(3), 285–297.

BAUER, D., KLING, A. and RUSS, J. (2008) A universal pricing framework for guaranteed minimum benefits in variable annuities. *Astin Bulletin*, (38), 621–651

BELANGER, A., FORSYTH, P. and LABAHN, G. (2009) Valuing the guaranteed minimum death benefit clause with partial withdrawals. *Applied Mathematical Finance*, **16**(6), 451–496.

BERNARD, C. and MACKAY, A. (2014) Reducing surrender incentives through fee structure in variable annuities. In *Proceedings of the Risk Management Reloaded Conference*.

BERNARD, C., MACKAY, A. and MUEHLBEYER, M. (2014) Optimal surrender policy for variable annuity guarantees. *Insurance: Mathematics and Economics* **55**, 116–128.

BOYLE, P.P. and HARDY, M.R. (2003) Guaranteed annuity options. *Astin Bulletin*, **33**(2), 125–152.

CHEN, Z., VETZAL, K. and FORSYTH, P.A. (2008) "The effect of modelling parameters on the value of GMWB guarantees. *Insurance: Mathematics and Economics*, **43**(1), 165–173.

COLEMAN, T., KIM, Y., LI, Y. and M. PATRON, M. (2007) Robustly hedging variable annuities with guarantees under jump and volatility risks. *Journal of Risk and Insurance*, **74**(2), 347–376.

DAI, M., KUEN KWOK, Y. and ZONG, J. (2008) Guaranteed minimum withdrawal benefit in variable annuities. *Mathematical Finance*, **18**(4), 595–611.

DICKSON, D.C.M., HARDY, M.R. and WATERS, H. (2009) *Actuarial Mathematics for Life Contingent Risks*. Cambridge, England: Cambridge University Press.

HARDY, M.R. (2001) A regime-switching model of long-term stock returns. *North American Actuarial Journal*, **5**(2), 41–53.

HARDY, M.R. (2003) *Investment Guarantees: Modeling and Risk Management for Equity-Linked Life Insurance*. Hoboken, New Jersey: John Wiley & Sons, Inc.

KARATZAS, I. and SHREVE, S.E. (1984) Trivariate density of Brownian motion, its local and occupation times, with application to stochastic control. *Annals of Probability*, **12**(3), 819–828.

KARATZAS, I. and SHREVE, S.E. (1991) *Brownian Motion and Stochastic Calculus*, 2nd ed. New York: Springer.

KIM, I.J. and YU, G.G. (1996) An alternative approach to the valuation of American options and applications. *Review of Derivatives Research*, **1**(1), 61–85.

https://doi.org/10.1017/asb.2014.13 Published online by Cambridge University Press

MARSHALL, C., HARDY, M.R. and SAUNDERS, D. (2010) Valuation of a guaranteed minimum income benefit. *North American Actuarial Journal*, **14**(1), 38–58.
MILEVSKY, M.A. and SALISBURY, T.S. (2001) The real option to Lapse a variable annuity: Can surrender charges complete the market. In *Conference Proceedings of the 11th Annual International AFIR Colloquium*.
MOENIG, T. and BAUER, D. (2012) *Revisiting the risk-neutral approach to optimal policyholder behavior: A study of withdrawal guarantees in variable annuities*, Working Paper, Georgia State University.
PALMER, B. (2006) "*Equity-indexed annuities: Fundamental concepts and issues*, Working Paper.
PRUDENTIAL(UK) (2012) Key Features of the Flexible Investment Plan (No Initial Charge and Initial Charge option) - Additional Investments, www.pru.co.uk/pdf/FIPK10080.pdf; Retrieved February 27, 2013.

CAROLE BERNARD
Department of Statistics and Actuarial Science
University of Waterloo
Waterloo, Ontario, Canada
E-Mail: c3bernar@uwaterloo.ca

MARY HARDY
Department of Statistics and Actuarial Science
University of Waterloo
Waterloo, Ontario, Canada
E-Mail: mrhardy@uwaterloo.ca

ANNE MACKAY (Corresponding author)
Department of Statistics and Actuarial Science
University of Waterloo
Waterloo, Ontario, Canada
E-Mail: a3mackay@uwaterloo.ca

APPENDIX

A. PROOF OF PROPOSITION 2.2

Proof. We first prove the case where $K \geqslant 0$. Observe that

$$\begin{aligned}
&\widetilde{Q}\big(\widetilde{W}_T \in dx,\, L_{T,\widetilde{W}}(K) \in dy,\, \Gamma^-_{T,\widetilde{W}}(K) \in dz\big) \\
&\quad = \widetilde{Q}\big(-\widetilde{W}_T \in -dx,\, L_{T,\widetilde{W}}(K) \in dy,\, \Gamma^-_{T,\widetilde{W}}(K) \in dz\big) \\
&\quad = \widetilde{Q}\big(Z_T \in K - dx,\, L_{T,\widetilde{W}}(K) \in dy,\, \Gamma^-_{T,\widetilde{W}}(K) \in dz\big),
\end{aligned}$$

where $Z_t = K - \widetilde{W}_t$ is a $\widetilde{Q}$-Brownian motion starting at K.

Next, we need to express $L_{T,\widetilde{W}}(K)$ and $\Gamma^-_{T,\widetilde{W}}(K)$ in terms of $L_{T,Z}(0)$ and $\Gamma^+_{T,Z}(0)$ (where $\Gamma^+_{T,Z}(0)$ denotes the occupation time of Z above 0). Then, it will

https://doi.org/10.1017/asb.2014.13 Published online by Cambridge University Press

be possible to use the results of Karatzas and Shreve (1984) to obtain the desired distribution. Note that

$$\Gamma^{-}_{T,\widetilde{W}}(K) = \int_0^T \mathbb{1}_{\{\widetilde{W}_s \leqslant K\}} ds = \int_0^T \mathbb{1}_{\{Z_s \geqslant 0\}} ds = \Gamma^{+}_{T,Z}(0). \tag{A1}$$

Using the Tanaka formula, we also have

$$\begin{aligned} L_{T,\widetilde{W}}(K) &= (\widetilde{W}_T - K)^- - (0-K)^- + \int_0^T \mathbb{1}_{\{\widetilde{W}_s \leqslant K\}} d\widetilde{W}_s \\ &= (-Z_T)^- - K - \int_0^T \mathbb{1}_{\{Z_s \geqslant 0\}} dZ_s \\ &= (-Z_T)^- - K - \int_0^T \left(1 - \mathbb{1}_{\{Z_s < 0\}}\right) dZ_s \\ &= (-Z_T)^- - K - Z_T + K + \int_0^T \mathbb{1}_{\{Z_s < 0\}} dZ_s \\ &= (-Z_T)^+ + \int_0^T \mathbb{1}_{\{Z_s < 0\}} dZ_s, \end{aligned} \tag{A2}$$

where we use $\int_0^T dZ_s = Z_T - Z_0 = Z_T - K$ to obtain the second to last equation. The Tanaka formula also allows us to express $L_{T,Z}(0)$ as

$$L_{T,Z}(0) = Z_T^- - K^- + \int_0^T \mathbb{1}_{\{Z_s < 0\}} dZ_s, \tag{A3}$$

where $K^- = 0$ as $K \geqslant 0$. Re-arranging (A2) using (A3), we obtain

$$L_{T,\widetilde{W}}(K) = L_{T,Z}(0) + (-Z_T)^+ - Z_T^-.$$

Since $(-Z_T)^+ - Z_T^- = 0$, we get

$$L_{T,\widetilde{W}}(K) = L_{T,Z}(0). \tag{A4}$$

Then using (A1) and (A4), we can write

$$\begin{aligned} &\widetilde{Q}\left(Z_T \in K - dx,\, L_{T,\widetilde{W}}(K) \in dy,\, \Gamma^{-}_{T,\widetilde{W}}(K) \in dz\right) \\ &\quad = \widetilde{Q}\left(Z_T \in K - dx,\, L_{T,Z}(0) \in dy,\, \Gamma^{+}_{T,Z}(0) \in dz\right). \end{aligned} \tag{A5}$$

Since Z_t is a Brownian motion starting at K, we can use the trivariate density of $(Z_T,\, L_{T,Z}(0),\, \Gamma^{+}_{T,Z}(0))$ given in Section 4 of Karatzas and Shreve (1984) and get the desired result.

https://doi.org/10.1017/asb.2014.13 Published online by Cambridge University Press

Similarly, for $K < 0$, we can observe that

$$\begin{aligned}
&\widetilde{Q}\big(\widetilde{W}_T \in dx, L_{T,\widetilde{W}}(K) \in dy, \Gamma^-_{T,\widetilde{W}}(K) \in dz\big) \\
&\quad = \widetilde{Q}\big(\widetilde{W}_T - K \in dx - K, L_{T,\widetilde{W}}(K) \in dy, \Gamma^-_{T,\widetilde{W}}(K) \in dz\big) \\
&\quad = \widetilde{Q}\big(H_T \in dx - K, L_{T,\widetilde{W}}(K) \in dy, \Gamma^-_{T,\widetilde{W}}(K) \in dz\big),
\end{aligned}$$

where $H_t = \widetilde{W}_t - K$ be a $\widetilde{Q}$-Brownian motion starting at $-K$. We now express $L_{T,\widetilde{W}}(K)$ and $\Gamma^-_{T,\widetilde{W}}(K)$ in terms of $L_{T,H}(0)$ and $\Gamma^+_{T,H}(0)$. We have

$$\Gamma^-_{T,\widetilde{W}}(K) = \int_0^T \mathbb{1}_{\{\widetilde{W}_s \leqslant K\}} ds = \int_0^T \mathbb{1}_{\{H_s \leqslant 0\}} ds = \int_0^T 1 - \mathbb{1}_{\{H_s \geqslant 0\}} ds = T - \Gamma^+_{T,H}(0).$$

We also have

$$\begin{aligned}
L_{T,\widetilde{W}}(K) &= (\widetilde{W}_T - K)^- - (0 - K)^- + \int_0^T \mathbb{1}_{\{\widetilde{W}_s \leqslant K\}} d\widetilde{W}_s \\
&= H_T^- + \int_0^T \mathbb{1}_{\{H_s \leqslant 0\}} dH_s \\
&= L_{T,H}(0). \qquad \text{(A6)}
\end{aligned}$$

Then we can write

$$\begin{aligned}
&\widetilde{Q}\big(H_T \in dx - K, L_{T,\widetilde{W}}(K) \in dy, \Gamma^-_{T,\widetilde{W}}(K) \in dz\big) \\
&= \widetilde{Q}\left(H_T \in dx - K, L_{T,H}(0) \in dy, \Gamma^+_{T,H}(0) \in T - dz\right). \qquad \text{(A7)}
\end{aligned}$$

Since H_t is a Brownian motion starting at $-K$ (remember that $K < 0$), the result follows from Section 4 of Karatzas and Shreve (1984). ■

B. Details for the GMMB price

Note that A_T and D_T in the expression (10) in Proposition 3.1 depend on the trivariate density established in Proposition 2.2, which depends on K and B. We now discuss all possible cases needed to implement this formula.

https://doi.org/10.1017/asb.2014.13 Published online by Cambridge University Press

Assume $K \geqslant 0$ ***and*** $B \geqslant G$**.**
In this case $K \geqslant H$ and we have $A_T = A_T^{(1)} + A_T^{(2)} + A_T^{(3)}$ where

$$\begin{cases} A_T^{(1)} = \int_H^{K} \int_0^{+\infty} \int_0^{T} e^{(\sigma+\alpha)x+\beta z-\frac{c}{\sigma}y+\frac{c}{\sigma}(K-x)}\, 2h(T-z, y)\, h(z, y+2K-x)dzdydx \\ A_T^{(2)} = \int_K^{+\infty} \int_0^{+\infty} \int_0^{T} e^{(\sigma+\alpha)x+\beta z-\frac{c}{\sigma}y}\, 2h(z, y+K)\, h(T-z, y-K+x)dzdydx \\ A_T^{(3)} = \frac{e^{\beta T+\frac{cK}{\sigma}}}{\sqrt{2\pi T}} \int_H^{K} e^{(\alpha+\sigma-\frac{c}{\sigma})x} \left(e^{\frac{-x^2}{2T}} - e^{\frac{-(x-2K)^2}{2T}}\right) dx. \end{cases}$$

We also have $D_T = D_T^{(1)} + D_T^{(2)}$ where

$$\begin{cases} D_T^{(1)} = \int_{-\infty}^{H} \int_0^{+\infty} \int_0^{T} e^{\alpha x+\beta z-\frac{c}{\sigma}y+\frac{c}{\sigma}(K-x)}\, 2h(T-z, y)\, h(z, y+2K-x)dzdydx \\ D_T^{(2)} = \frac{e^{\beta T+\frac{cK}{\sigma}}}{\sqrt{2\pi T}} \int_{-\infty}^{H} e^{(\alpha-\frac{c}{\sigma})x} \left(e^{\frac{-x^2}{2T}} - e^{\frac{-(x-2K)^2}{2T}}\right) dx. \end{cases}$$

Assume $K < 0$ ***and*** $B \geqslant G$**.**
In this case $K \geqslant H$ and we have $A_T = A_T^{(4)} + A_T^{(5)} + A_T^{(6)}$ where

$$\begin{cases} A_T^{(4)} = \int_H^{K} \int_0^{+\infty} \int_0^{T} e^{(\sigma+\alpha)x+\beta z-\frac{c}{\sigma}y+\frac{c}{\sigma}(K-x)}\, 2h(T-z, y-K)\, h(z, y-x+K)dzdydx \\ A_T^{(5)} = \int_K^{+\infty} \int_0^{+\infty} \int_0^{T} e^{(\sigma+\alpha)x+\beta z-\frac{c}{\sigma}y}\, 2h(z, y)\, h(T-z, y+x-2K)dzdydx \\ A_T^{(6)} = \frac{e^{\frac{cK}{\sigma}}}{\sqrt{2\pi T}} \int_H^{K} e^{(\alpha+\sigma-\frac{c}{\sigma})x} \left(e^{\frac{-x^2}{2T}} - e^{\frac{-(x-2K)^2}{2T}}\right) dx. \end{cases}$$

We also have $D_T = D_T^{(3)} + D_T^{(4)}$ where

$$\begin{cases} D_T^{(3)} = \int_{-\infty}^{H} \int_0^{+\infty} \int_0^{T} e^{\alpha x+\beta z-\frac{c}{\sigma}y+\frac{c}{\sigma}(K-x)}\, 2h(T-z, y-K)\, h(z, y-x+K)dzdydx \\ D_T^{(4)} = \frac{e^{\frac{cK}{\sigma}}}{\sqrt{2\pi T}} \int_{-\infty}^{H} e^{(\alpha-\frac{c}{\sigma})x} \left(e^{\frac{-x^2}{2T}} - e^{\frac{-(x-2K)^2}{2T}}\right) dx. \end{cases}$$

Assume $K \geqslant 0$ ***and*** $B < G$**.**
In this case $K < H$ and we have

$$A_T = \int_H^{+\infty} \int_0^{+\infty} \int_0^{T} e^{(\sigma+\alpha)x+\beta z-\frac{c}{\sigma}y}\, 2h(z, y+K)\, h(T-z, y-K+x)dzdydx.$$

https://doi.org/10.1017/asb.2014.13 Published online by Cambridge University Press

We also have $D_T = \bar{D}_T^{(1)} + \bar{D}_T^{(2)} + \bar{D}_T^{(3)}$ where

$$\begin{cases} \bar{D}_T^{(1)} = \int_{-\infty}^{K} \int_{0}^{+\infty} \int_{0}^{T} e^{\alpha x + \beta z - \frac{c}{\sigma} y + \frac{c}{\sigma}(K-x)} 2h(T-z, y)\, h(z, y+2K-x) dz dy dx \\ \bar{D}_T^{(2)} = \int_{K}^{H} \int_{0}^{+\infty} \int_{0}^{T} e^{\alpha x + \beta z - \frac{c}{\sigma} y} 2h(z, y+K)\, h(T-z, y-K+x) dz dy dx \\ \bar{D}_T^{(3)} = \frac{e^{\beta T + \frac{cK}{\sigma}}}{\sqrt{2\pi T}} \int_{-\infty}^{K} e^{(\alpha - \frac{c}{\sigma})x} \left(e^{\frac{-x^2}{2T}} - e^{\frac{-(x-2K)^2}{2T}} \right) dx. \end{cases}$$

Assume $K < 0$ ***and*** $B < G$**.**
In this case $K < H$ and we have

$$A_T = \int_{H}^{+\infty} \int_{0}^{+\infty} \int_{0}^{T} e^{(\sigma+\alpha)x + \beta z - \frac{c}{\sigma} y} 2h(z, y)\, h(T-z, y+x-2K) dz dy dx$$

We also have that $D_T = \bar{D}_T^{(4)} + \bar{D}_T^{(5)} + \bar{D}_T^{(6)}$ where

$$\begin{cases} \bar{D}_T^{(4)} = \int_{-\infty}^{K} \int_{0}^{+\infty} \int_{0}^{T} e^{\alpha x + \beta z - \frac{c}{\sigma} y + \frac{c}{\sigma}(K-x)} 2h(T-z, y-K)\, h(z, y-x+K) dz dy dx \\ \bar{D}_T^{(5)} = \int_{K}^{H} \int_{0}^{+\infty} \int_{0}^{T} e^{\alpha x + \beta z - \frac{c}{\sigma} y} 2h(z, y)\, h(T-z, y+x-2K) dz dy dx \\ \bar{D}_T^{(6)} = \frac{e^{\frac{cK}{\sigma}}}{\sqrt{2\pi T}} \int_{-\infty}^{K} e^{(\alpha - \frac{c}{\sigma})x} \left(e^{\frac{-x^2}{2T}} - e^{\frac{-(x-2K)^2}{2T}} \right) dx. \end{cases}$$

To compute $A_T^{(3)}$, we observe that for any $a \in \mathbb{R}$, we have the following identity

$$\begin{aligned} &\frac{1}{\sqrt{2\pi T}} \int_{H}^{K} e^{ax} \left(e^{-\frac{x^2}{2T}} - e^{-\frac{(x-2K)^2}{2T}} \right) dx \\ &= e^{\frac{a^2 T}{2}} \left[N\left(\frac{aT - H}{\sqrt{T}} \right) - N\left(\frac{aT - K}{\sqrt{T}} \right) \right. \\ &\left. - e^{2aK} N\left(\frac{2K - H + aT}{\sqrt{T}} \right) + e^{2aK} N\left(\frac{K + aT}{\sqrt{T}} \right) \right]. \end{aligned}$$

https://doi.org/10.1017/asb.2014.13 Published online by Cambridge University Press

After replacing $a = \alpha + \sigma - \frac{c}{\sigma} = \frac{r-c}{\sigma} + \frac{\sigma}{2}$, and simplify, we obtain

$$A_T^{(3)} = \left(\frac{B}{F_0}\right)^{\frac{c}{\sigma^2}} e^{\frac{T}{2\sigma^2}(\gamma^2 - c(c-2r+\sigma^2))} \left[N\left(\frac{\ln\left(\frac{F_0}{G}\right) + \gamma T}{\sigma\sqrt{T}}\right) - N\left(\frac{\ln\left(\frac{F_0}{B}\right) + \gamma T}{\sigma\sqrt{T}}\right) \right.$$
$$\left. - \left(\frac{B}{F_0}\right)^{\frac{2\gamma}{\sigma^2}} N\left(\frac{\ln\left(\frac{B^2}{F_0 G}\right) + \gamma T}{\sigma\sqrt{T}}\right) + \left(\frac{B}{F_0}\right)^{\frac{2\gamma}{\sigma^2}} N\left(\frac{\ln\left(\frac{B}{F_0}\right) + \gamma T}{\sigma\sqrt{T}}\right) \right],$$

where $\gamma = r + \frac{\sigma^2}{2} - c$.

Define $\mathcal{I}(a, b)$ as follows

$$\mathcal{I}(a, b) := \frac{1}{\sqrt{2\pi T}} \int_{-\infty}^{b} e^{ax} (e^{-\frac{x^2}{2T}} - e^{-\frac{(x-2K)^2}{2T}}) dx.$$

We use the following identity to get closed-form expressions for $\bar{D}_T^{(3)}$ and $D_T^{(2)}$:

$$\mathcal{I}(a, b) = e^{\frac{a^2 T}{2}} \left[N\left(\frac{b - aT}{\sqrt{T}}\right) - e^{2aK} N\left(\frac{b - aT - 2K}{\sqrt{T}}\right) \right].$$

It is clear that $\bar{D}_T^{(3)} = e^{\beta T} (\frac{B}{F_0})^{\frac{c}{\sigma^2}} \mathcal{I}(\alpha - \frac{c}{\sigma}, K)$ and $D_T^{(2)} = e^{\beta T} (\frac{B}{F_0})^{\frac{c}{\sigma^2}} \mathcal{I}(\alpha - \frac{c}{\sigma}, H)$. After simplifications, we obtain $\bar{D}_T^{(3)}$

$$\bar{D}_T^{(3)} = \left(\frac{B}{F_0}\right)^{\frac{c}{\sigma^2}} e^{\left(r - \frac{\sigma^2}{2}\right)^2 \frac{T}{2\sigma^2}} \left(N\left(\frac{\ln(\frac{B}{F_0}) - \eta T}{\sigma\sqrt{T}}\right) - \left(\frac{B}{F_0}\right)^{\frac{2\eta}{\sigma^2}} N\left(\frac{\ln(\frac{F_0}{B}) - \eta T}{\sigma\sqrt{T}}\right) \right),$$

where $\eta = r - \frac{\sigma^2}{2} - c$. Note that $D_T^{(2)}$ can be obtained similarly as the only difference is that H replaces K. Thus we only need to replace B by G in the above expression to obtain $D_T^{(2)}$. □

https://doi.org/10.1017/asb.2014.13 Published online by Cambridge University Press