The Real Option to Lapse
and the Valuation of Death-Protected
Investments

Moshe A. Milevsky and[1] Tom S. Salisbury
York University, Toronto
Ontario, M3J 1P3, Canada

January 2002

Contents

The Real Option to Lapse and
the Valuation of Death-Protected Investments
ABSTRACT

Variable Annuity (VA), similar to their mutual fund cousins, are managed pools of investments whose gains are tax-deferred. In addition, they provide money-back guarantees on invested principal. These guarantees – which can be viewed as non-separable put options with a possibly increasing strike price – are structured to mature upon death of the contract owner and/or at a pre-specified time horizon. The protection is paid for by installments – as opposed to up front – and is funded with a proportional insurance charge that is deducted from the underlying fund on a periodic basis.

Of great importance – and critical to any formal analysis of this product – is the fact that the holder can *lapse the contract* and instantaneously repurchase an identical investment to reestablish a new basis for the guarantee. In the absence of transaction costs, this would be optimal each and every time the value of the account reaches a new high. We classify this strategy as the Real Option to lapse. In this paper we analyze this product, by focusing on the *optimal time* to exercise the Real Option to lapse.

Our paper's conceptual contribution lies in highlighting the critical importance of the deferred surrender charge (DSC) – essentially transaction costs – in *completing* the market and allowing the claim to be hedged. Technically, we formulate the valuation exercise as a *optimal stopping* problem to provide a closed-form analytic solution and complete analysis when hazard rates are constant. Some numerical examples are provided to confirm that most annuity vendors are substantially overcharging for this guarantee.

JEL Codes: G13, G23, D91, G22

Keywords: *Optimal control, option pricing, insurance.*

1 Introduction.

Most of the existing literature on derivative pricing – starting with Black-Scholes/Merton (1973) – tacitly assumes that the price of an a option will be paid in advance, at the time of acquisition. In practice, however, many financial and especially insurance contracts, that contain embedded options, are purchased by a series of installments. So, in fact, a very minor payment is made up front, and the remainder is due over an extended period of time. Sometimes, the magnitude of the installment payments themselves is proportional to the value of the underlying security on which the option is written. Clearly, it is incorrect to amortize the cost of the option over the (expected) life of the contract. This is because the installment plan endows the holder (long position) with an additional *Real Option*[1] to terminate payments, or lapse the contract without having paid the full value of the embedded option to the seller (short position). Most likely – and rationally – the contract will be lapsed, or terminated, when the embedded options are out-of-the-money. Exercising this option is akin to defaulting on a swap contract when it becomes a liability, or a corporation abandoning a mine when it becomes uneconomical, although with less distasteful legal (and ethical) consequences.

One of the most common example of this *personal abandonment option* occurs with (tax sheltered) Variable Annuities, in the U.S. and Segregated Funds in Canada. These insurance-based products are quite similar[2] to their mutual fund cousins, in all financial, legal and accounting aspects, except that they contain an additional money-back guarantee (to the estate or heirs) upon death, or at some fixed horizon[3]. If the fund has lost money – net of any withdrawals – the insurance company will refund the difference. In fact, it is now possible to purchase almost any well known mutual fund product,

---

[1]We are using the term real option in the personal, as opposed to corporate finance, sense of word. This is strictly different from the classical use of the term in the literature. We refer the interested reader to the work by Berk (1999), Amram and Kulatilaka (1999), Trigeorgis (1996), Ross (1995), Ingersoll and Ross (1992) and Hubbard (1994) for additional information about real options.

[2]The *raison detre* of the Variable Annuity (VA) is the convenient deferral of taxation on all investment gains until the funds are withdrawn or annuitized. In this paper we will not concern ourselves with the tax aspects of the product, other than to mention that one can, in fact, lapse a VA contract, and purchase another one, without incurring any tax consequences.

[3]The insurance lingo for this feature is a guaranteed minimum accumulation benefit (GMAB) which is available as an additional rider on most variable annuity policies.

with, or without, the added insurance protection. The embedded protection is paid for by installments – as opposed to up front – and funded via a continuous insurance charge that is deducted from the underlying fund on a periodic basis. In other words, the management fees are higher on these products, compared to regular mutual funds.[4] The size of this market is not trivial. According to recent estimates by Moody's, there is approximately $1 trillion U.S. invested in variable annuity policies with explicit maturity guarantees. In fact, Moody's recently issued a 'special comment' which was entirely devoted to the risks of these products.[5]

The impetus for our paper, is that the holder of these products – collectively known as *Maturity-Protected Investments* (MPIs) – can at any time sell and then repurchase the investment to reestablish a new basis for the guarantee. This, in theory, would be optimal each and every time the value of the account reaches a new high. By continuously engaging in this out-and-in transaction, the individual could theoretically convert a money back guarantee (i.e. a vanilla put) into a highest-value-achieved (i.e. lookback put option) guarantee. Of course, transaction costs, in the form of loads and surrender charges complicate this simple scheme, which is the essence of our analysis. We classify this feature as the Real Option (RO) to lapse.

For example, assume that a contract owner, who is 55 years old, invests $10,000 in a variable annuity with a basic money-back guarantee at death, or in ten years, whichever comes first. The guarantee is funded with an additional 100 basis points annual fee that is charged to assets on a daily basis. If, for example, the investment doubles in value to $20,000 over the next year, the contract owner is still paying 100 basis points for a put option which is out-of-the-money by 50%. Clearly, there is a huge incentive to sell the fund, and then re-purchase the exact same investment, to re-establish the guarantee at the new $20,000 level. The companies providing the guarantee are aware of this and therefore impose a contingent deferred surrender charge (DSC) to 'force' investors to stay in the fund, or at the very least recoup some of the costs if investors decide to lapse. Even so – and despite the transaction costs – it may be optimal to swap the old out-of-the-money put, in

---

[4]According to Morningstar Inc., the average expense ratio on the universe of 8,200 U.S. mutual funds is 137 basis points, while the 6,600 variable annuity sub-accounts have an average expense of 211 basis points. The 74 b.p. difference can be viewed as payment for the option.

[5]See *Bells and Whistles: Credit Implcations of the New Variable Annuities.* Moody's Investors Service, October 2000.

a exchange for the new at-the-money put. This paper will derive the optimal lapsation policy, in the presence of a particular DSC schedule as well as the appropriate continuous insurance charge that funds the maturity/death guarantee. Our paper links the DSC, the continuous insurance charge and the optimal lapsation strategy in a parsimonious and tractable manner.

From an academic point of view, a large and diverse body of research has been published over the last 20 years on the topic of maturity-protected investments. In the finance and insurance literature these products have been analyzed under the title of equity-linked, or unit-linked, insurance policies. The landmark contribution in the field, was the Brennan and Schwartz (1976), or Boyle and Schwartz (1977) extension of the Black-Scholes/Merton formula (1973) to equity-linked insurance contracts. They assumed a market that is complete in both financial and mortality risk. Therefore, all derivative prices can be expressed as suitable expectations with respect to an appropriate probability measure. Building on the complete markets framework, various researchers extended the analysis to stochastic interest rates and exotic payoff structures. See, for example, Bacinello and Ortu (1993, 1996), Ekern and Persson (1996), Nielsen and Sandman (1996), Persson and Aase (1997), Aase and Persson (1994) as well as Miltersen and Persson (1999). Without exceptions, all of the above mentioned papers on equity-linked insurance have focused on locating the *single initial premium* that funds, or pays for, the maturity benefit. In practice, of course, the guarantee is always paid by installments, which, as we have argued, completely changes the nature of the problem. More recently, and more practically, Windcliff, Forsyth and Vetzal (2000), as well as Boyle, Kolkiewicz and Tan (1999) have looked at the 'reset' features available in some of the variable annuities using Monte Carlo and numerical PDE approaches. Milevsky and Posner (2000), provided theoretical and empirical evidence on the cost structure of variable annuity contracts.

Of course, within the context of insurance, a complete market assumption implies that vendors can completely diversify their mortality risk by selling enough policies. In contrast to these assumptions, *our main argument is that when option premiums are paid by installments – even in the presence of complete mortality and financial markets – the ability to 'lapse' de facto creates an incomplete market in which the contingent claim can not be hedged.* Therefore, to salvage the hedge, our theoretical contribution is to identify the contingent deferred surrender charges (DSC) – properly calibrated to the optimal lapsation policy – that will complete the market.

5

Usually, one thinks of transaction costs, commissions and trading fees as impediments to pricing *via* risk-neutral expectations. Indeed, theoretical research by Garman and Ohlson (1981) and Dermody and Rockafellar (1991) as well as empirical work by Ronn (1987) has shown that 'frictions' induce a multitude of non-unique valuation operators, which is the essence of incomplete financial markets. In contrast, we will demonstrate that our particular contingent claim can only be hedged in the presence of a 'friction', which is the contingent deferred surrender charge (DSC). This is not unlike the ideas introduced by Dammon and Green (1987), and Prisman (1986), where tax (frictions) can induce equilibrium in the bond market.

From a slightly different perspective, our research is similar to Geske's (1979) *compound option*, or Carr's (1988) *sequential exchange opportunity* where the holder is granted the right (but not the obligation) to acquire another option at some future point in time. In our case, the contract owner has the right to continue holding the fund – while paying additional expenses – and maintaining the 'old' money back guarantee. Likewise, our problem can also be positioned within the context of the vulnerable (defaultable) options literature, for example Johnson and Stulz (1987), but where the default emanates with the buyer, instead of the seller.

The technical contribution of this paper is to formulate[6] the personal lapsation decision as an optimal stopping problem. The Real Option is then valued as an American contingent claim, with the strike price being equal to the underlying account value net of any deferred surrender charges. From that point, standard American option pricing techniques, such as Jack (1991), Kim (1990), or Huang, Subrahmanyam and Yu (1996), or the more recent Ju (1998) or Carr (1998), can be applied to locate the optimal lapsation boundary, and by inversion, the proper continuous insurance charge.

Conveniently, when the population hazard rate is assumed constant over time (exponential death) the free-boundary problem can be simplified *a la* McKean (1965), and the corresponding ODE can be solved, to obtain closed form analytic expressions for all quantities of interest. We provide expressions

---

[6]Optimal stopping techniques, *vis a vis* the decision to surrender or lapse an insurance contract, have been employed in the academic literature. See for example Grosen and Jorgensen (1997) as well as Albizzati and Geman (1994). However, the options analyzed in those papers involve explicit interest rate guarantees that are added to insurance products and therefore more properly treated as financial option, rather than Real Options. More importantly, in contrast to those papers, our guarantees are paid by installments, which is the *essence* of our market incompleteness problem.

for the optimal level at which to lapse the contract, assuming a fixed asset and deferred surrender charge.

By means of an example we pre-empt our numerical results with the following case study; assuming the current 6% interest rate environment. A 60 year-old individual, with a life expectancy of 20 years[7], that purchases a death-protected mutual fund with a (historical broad-based U.S. equity) volatility of 20%, an additional *continuous insurance charge* of 30 basis points per annum, and a 2% contingent deferred surrender charge, should optimally lapse the contract as soon as the account appreciates by exactly 57%. The owner should sell the position – incurring the 2% DSC – and then repurchase the exact same contract, which obviously generates a new at-the-money guarantee. From the insurer's point of view, if the population exercises too early, the insurer pockets more surrender charges than are needed to cover the hedge. If the population exercises too late, then its guarantee – at least for a time – is at a lower level than the insurer is hedged against. So, when people die, the insurer does not have to pay out as much.

Stated differently, a 30 basis point *continuous insurance charge* together with a 2% contingent deferred surrender charge will completely fund the money back guarantee. However, the same model also indicates that *if* the insurance company charges less than 16 basis points for the guarantee, the mortality-contingent claim is essentially un-hedgeable, regardless of how high the contingent deferred surrender charge is set. Finally, if the insurance company charges more than 167 basis points for the guarantee, regardless of how low the deferred surrender charge is set, a rational individual will never purchase the product.

The remainder of this paper is organized as follows. Section 2 will introduce notation, terminology and then derive the general model for the optimal lapsing time and the value of the American contingent claim. Section 3 continues by making some specific assumptions about hazard rates and maturity guarantees which then allows for a closed-form analytic solution. Section 4 provides some numerical examples, while Section 5 concludes the paper with some general remarks and directions for further research. All technical proofs are relegated to the appendix.

---

[7]This is according to the U.S. Decennial Life Tables for 1989-1991, complied by the U.S. Department of Health and Human Services in conjunction with the National Center for Health Statistics

2 The General Model

2.1 Underlying Asset and Dynamics

The contract owner pays exactly \$1, at time $t = 0$, to acquire a long position in the mutual fund, together with a non-separable increasing-strike put option that matures at $\min[\tau, T]$, where τ is the stochastic time of death and T is the maturity of the guarantee, if any. In other words, the put option guarantees at least $e^{gt}, g \geq 0$ at time T, or at death, whichever comes first. Of course, we can let $T = \infty$, which implies a guarantee at death only. In the most general formulation, the physical price process for the underlying asset obeys:

$$dU_t = (\mu_t - \alpha_t)U_t dt + \sigma(U_t, t)dB_t, \qquad U_0 = 1, \tag{1}$$

under the statistical (or actuarial measure), where B_t is a standard Brownian motion, μ_t is the drift rate, net of any non-insurance management fees, and α_t is the to-be-determined continuous insurance charge that 'pays for' the option. The variable α_t can be viewed as a dividend yield outflow. Of course, the dividend does not go to the fund holder, but to the insurance company. In fact, actual dividends are assumed to be completely re-invested in the fund, and are therefore absorbed in μ_t. Nevertheless, the continuous-time payment $\alpha_t U_t$ flows to the insurance company providing the maturity/death guarantee. In practice, the underlying fund consists of a collection of individual securities, each following its own diffusion process. Without any loss of generality, we will simply focus on the mutual fund value U_t, and assume that it can be treated as a single asset.[8]

The risk neutral process for the mutual fund, which we use for our computations, is:

$$dU_t = (r_t - \alpha_t)U_t dt + \sigma(U_t, t)d\widetilde{B}_t, \qquad U_0 = 1, \tag{2}$$

where r_t is the risk-free short rate, that replaces the drift, and $\widetilde{B}_t$, is the Brownian motion under the new (Girsanov-transformed) measure.

One can also think of the underlying process U_t as:

$$U_t = S_t A_t, \qquad \text{where} \qquad A_t = e^{-\int_0^t \alpha_s ds} \tag{3}$$

---

[8]Technically speaking, there is an element of a 'passport option' – see Shreve and Vecer (2000) for recent details on these option – since the holder can re-allocate the individual sub-accounts to increase the value of the guarantee. We ignore this problem at a first pass.

and S_t is the market value of the assets supporting the fund. In other words:

$$dS_t = r_t S_t dt + \sigma(S_t, t) d\widetilde{B}_t, \qquad S_0 = 1 \qquad (4)$$

Finally, we let:

$$R_t = e^{\int_0^t r_s ds}, \qquad (5)$$

denote the standard money market account, where r_s is defined as above.

2.2 Lapsation and Deferred surrender Charges.

The owner can lapse or surrender (read: sell) the contract at any time $t < \tau \leq T$, and immediately receive an amount $(1 - k_t)U_t$, where $0 \leq k_t < 1$. The *deterministic* function k_t represents the contingent deferred surrender charge (DSC), which also goes directly to the company insuring the contract. One can think of the deferred surrender charge k_t, as both an incentive to remain invested in the contract, and, more importantly, as a mechanism for funding the put option. Intuitively, one can think of k_t as back-up for α_t, in the event the owner lapses before the original option has been fully paid for. Clearly, at the extreme, if $k_t = 1$, the owner will never lapse and the 'full' α_t will be collected for the entire life of the product.

Practically speaking, our model will locate the minimal (suitably defined) DSC needed to cover a fixed (suitably defined) α_t continuous insurance charge, as well as the minimal α_t required to fund the guarantee, in the presence of a fixed DSC. Indeed, as we shall prove later, if we assume a $k_t = 0$, for all $t \geq 0$ – in other words, no deferred surrender charges – there is no viable continuous insurance charge α_t that will fund the put option. In other words, if $k_t = 0$, then the only α_t that will fund the guarantee is so high that the buyer's rational policy will be to lapse immediately. Such a product is not economically viable. Likewise, there is no deferred surrender charge schedule that can compensate for an $\alpha_t = 0$. A policy of not lapsing will simply leave no revenue with which to fund the guarantee. Both the deferred surrender and the continuous insurance charge are critical for completing the market and to properly hedge the contingent claim.

In particular, in the case of exogenously imposed constant values for the surrender and continuous insurance charge, our model will identify the (α, k) 'curve' that completely funds the guarantee. Any combination of parameters along this curve will result in a viable product. This curve will also induce lower and upper bounds for α, denoted by α_L and α_H respectively, outside which the product is unsustainable. More on this in section 4.

2.3 The Maturity and Death Guarantees

Let $G(U_t) = \max[e^{gt}, U_t], g \geq 0$ denote the guaranteed amount. The fund payoff can therefore be described as follows:

$$\text{Payoff} = \begin{cases} (1 - k_t)U_t & \text{if} \quad \text{Lapsed.} \\ \max[e^{g\tau}, U_\tau] & \text{at} \quad \text{Death } (t = \tau) \\ \max[e^{gT}, U_T] & \text{at} \quad \text{Maturity } (t = T) \end{cases} \tag{6}$$

The formulation in equation (6) is general enough to include cases with no maturity guarantee – such as with variable annuity policies in the U.S. – in which case we set $T = \infty$. Nevertheless, the following stylized facts should be evident from the structure of equation (6) First, our model assumes that all possible surrender charges are waived upon death. In other words, it is never optimal to lapse 'an instant' before death[9] since $\max[e^{g\tau}, U_\tau]$ is strictly greater than $(1 - k_\tau)U_\tau$, whenever $k_\tau > 0$. Second, in the event of a time$-T$ maturity guarantee, the money back guarantee implies that $k_T = 0$ and all surrender charges are eliminated. Note that we do not require k_t to be continuous at $t = T$.

Finally, it is very important to note that the actual word 'lapsation' can imply two very different activities. Lapsation can be *rational*, when it is immediately followed by a re-purchase and solely conducted to re-establish the basis of the guarantee. And lapsation can be *irrational*, when the deferred surrender charge 'penalty' exceeds the value of the new option. Regardless of whether lapsation was rational, or not, the payoff (or value) at the time of lapsation will always be $(1 - k_t)U_t$. Our intention is to locate the situations where it is rational to incur the $k_t U_t$ 'penalty' for the sole purpose of re-acquiring the contract and resetting the level of the guarantee. Our model, as it currently stands, does not account for consumption or other liquidity needs that would induce people to lapse for reasons other than swapping an old option, for a new one. As such, we stress for one final time that our optimal lapsation strategy means the optimal strategy for re-establishing the basis. Clearly, though, there is nothing irrational about withdrawing funds from an investment account, in order to fund general consumption needs.

---

[9]Note, however, that in practice, estate taxes might create an incentive to lapse the contract 'an instant' prior to death. Needless to say, we will ignore taxes and other market imperfections for the time being.

2.4 The Death Rate

Let $\lambda_t \geq 0$ be the (hazard) death rate for the insured population holding the variable annuities. In our current formulation of the problem, and throughout the paper, the hazard rate is assumed to be deterministic. Therefore, we let:

$$\beta_t = e^{-\int_0^t \lambda_s ds} \tag{7}$$

denote the probability of any individual within the group surviving to time t, conditional on being alive at time zero. This implies that $\beta_0 = 1$ and $\lim_{t \to \infty} \beta_t = 0$.

More importantly, and quite critical to our model, we assume a *very large* pool of insured fund owners – each of whom invests a relatively small amount in the protected mutual fund – so that a fraction $\beta_t \lambda_t dt$ of the population dies between time t and $t + dt$. This is another way of stating the classical (and simplifying) assumption that mortality risk is completely diversifiable, and therefore not priced by the market. As a result, the outflow (or payout) due only to death between time t and $t + dt$ is precisely:

$$c_t dt := v_0 G(U_t) \beta_t \lambda_t dt, \tag{8}$$

where v_0 is the (very large) originally invested capital of all the fund owners. We are tracking the dynamic evolution of the entire sum of money which was originally invested in the death-protected mutual fund at time $t = 0$. Therefore, to be absolutely precise, we will in fact be computing the optimal lapsation policy for the entire population, as opposed to any particular individual within that group. Rational behaviour will be for all individuals still alive at that time – who have the same investment guarantee – to lapse simultaneously.

It should be emphasized that we are not going to hedge any individual variable annuity account. Rather, we will hedge the seller's exposure to all such accounts in aggregate.

2.5 The Hedge Portfolio

The insurance company insuring the protected mutual fund hedges the guarantee by trading the underlying asset S_t (not U_t) and the money market account R_t, during the life of the product. Up to the optimal lapsation time, the hedge portfolio will be denoted by:

$$V_t = \varphi_t S_t + \psi_t R_t \tag{9}$$

11

where φ_t is the amount held in the underlying security, and ψ_t is the amount invested in the money market account. This formulation is intuitively consistent with the firm's commitments, since, if there were no guarantees, $V_t := v_0 U_t$, and $\varphi_t := v_0 A_t$, as per equation (3). The presence of the guarantee 'forces' a ψ_t term, and a more complex trading strategy φ_t.

The hedging portfolio will obey the following stochastic differential equation:

$$dV_t = \varphi_t dS_t + \psi_t dR_t - c_t dt, \tag{10}$$

The term $\varphi_t dS_t + \psi_t dR_t$ in equation (10) is a self-financing portion, while the quantity, $c_t dt$ represents the (consumption) outflow due to death. At this point we have not determined (optimized) α_t and k_t, and therefore V_0 does not necessarily equal v_0, the original amount invested. In other words, recall that our objective is to find the $\{\alpha_t, k_t\}$ 'pair' that makes the hedge self-financing. In our context, self-financing implies that the initial (aggregate) cost of the hedge, V_0, is exactly equal to the initial (aggregate) amount invested by the unit holders, denoted by v_0.

By the definition of a hedging strategy, for the portfolio V_t to cover the guarantee, we must have that:

$$V_t \geq L_t = \left\{ \begin{array}{ll} v_0(1 - k_t)\beta_t U_t & \text{for} \quad 0 \leq t \leq T \\ v_0 \beta_T G(U_T) & \text{for} \quad t = T \end{array} \right., \qquad a.s. \tag{11}$$

The value of the portfolio (assets) must always exceed the liability. The portion $v_0(1 - k_t)\beta_t U_t$ covers the 'lapsation value' of the fund, for the fraction (β_t) who are still alive, while $v_0 \beta_T G(U_T)$ covers the maturity guarantee.

2.6 Formulation as American Contingent Claim

Using martingale pricing methodology, extensively described in Karatzas and Shreve (1998), we now locate V_0, which is the initial cost of the hedge. First, notice that although V_t, from equation (10), is not a self-financing portfolio (SFP), as a result of $c_t dt$, it can obviously be converted to an SFP by adding an appropriate term. The portfolio $V_t + \eta_t R_t$, is an SFP if we construct $d\eta_t = (c_t/R_t)dt$, and $\eta_0 = 0$. This, in turn, implies that:

$$M_t := \frac{1}{R_t}(V_t + \eta_t R_t) = \frac{V_t}{R_t} + \eta_t, \tag{12}$$

is a Martingale. Therefore, since $R_0 = 1$, and $\eta_0 = 0$, we have that:

$$V_0 = M_0 = \widetilde{E}_0\left[M_s\right] = \widetilde{E}_0\left[\frac{V_s}{R_s} + \int_0^s \frac{c_t}{R_t}dt\right] = \widetilde{E}_0\left[\frac{V_s}{R_s} + \int_0^s \frac{v_0\beta_t\lambda_t G(U_t)}{R_t}dt\right], \tag{13}$$

for any $s > 0$, where $\widetilde{E}_t[.]$ denotes the risk-neutral conditional (on time$-t$) expectation. In words, the discounted value of the hedging portfolio *plus* the discounted sum of all payments made at death, is a Martingale. If, for example, the death rate and the continuous insurance charge are set to zero ($\alpha_t = 0$, $\lambda_t = 0$), while the deferred surrender charge is set at 100%, (i.e. $k_t = 1$), equation (13), with $s = T$, collapses to: $V_0 = \widetilde{E}_0\left[\max[1, S_T]/R_T\right]$. This is the Black-Scholes/Merton, risk-neutral expectation for an at-the-money European put option plus a position in the underlying security.

Finally, our paper's main theoretical result is as follows: The initial value (cost) of the hedging portfolio for the variable annuities must satisfy:

$$V_0 = v_0 \sup_{0 \leq s \leq T} \widetilde{E}_0\left[\frac{L_s}{R_s} + \int_0^s \frac{\beta_t\lambda_t G(U_t)}{R_t}dt\right], \tag{14}$$

where the supremum in equation (14) is taken over all possible stopping times, in the early exercise sense of McKean (1965). Most importantly, the value of $s^*(U.)$ which maximizes the risk-neutral expectation in equation (14), is the optimal lapsation time for the death-protected mutual fund.

At first glance, the expectation in equation (14), and the free boundary problem is creates, is more 'complicated' than a standard American option pricing situation. This is primarily due to the path-dependent nature of the integral term. And, in the most general case, all we can hope for are numerical approximation, *a la* techniques described in detail by Kim (1990) or Ju (1998), for example. Fortunately, as we shall see in the next section, when we impose a particular structure on the hazard rate λ_t (and maturity guarantee) and in addition we assume a fixed$\{\alpha, k\}$ 'pair', the problem can actually be solved in closed-form. This is quite similar to the valuation of the perpetual American put – which is available in closed form – because the relevant PDE collapses to an easily-solvable ODE.

To complete the argument, the expectation in equation (14) will provide us with V_0. We then set $V_0 = v_0$, to locate the $\{\alpha_t, k_t\}$ surface. To understand why this is the case, imagine the following. If we assume that k_t is small enough and the $\alpha_t = 0$, then it must be that for any fixed v_0, $V_0/v_0 > 1$, since the initial capital is not enough to fund the benefit. Also, if $\alpha_t = \alpha \to \infty$,

13

then $V_0/v_0 < 1$, since, $U_t \to 0$ and in this case the guarantee collapses to:

$$\frac{V_0}{v_0} = \max\left[1 - k_0, \int_0^T \frac{\beta_t \lambda_t e^{gt}}{R_t} dt + \frac{\beta_T e^{gT}}{R_T}\right].\qquad(15)$$

Now, $\int_0^T \beta_t \lambda_t dt = 1 - \beta_T$, so the second term in equation (15) is less than one, as long as $r_t > g \; \forall t$. This should be intuitive since otherwise the guarantee exceeds the risk free rate which can not be sustainable.[10] Likewise, the first term in equation (15) is less than one, provided that $k > 0$. And so, $V_0/v_0 < 1$, when $\alpha \to \infty$, provided that both conditions apply. Therefore, in either event, there must exist some intermediate value of $\widehat{\alpha}$, that results in $V_0/v_0 = 1$, and that exactly funds the maturity/death guarantee.

2.7 Alternative Risk-Neutral Representation.

Define the (future lifetime) random variable, τ, with density $\lambda_t \beta_t$, such that:

$$P(\tau > t) = \int_t^\infty \lambda_s \beta_s ds = \beta_t.\qquad(16)$$

We can now express equation (13) as:

$$V_0 = v_0 \sup_{0 \le s} E_0^* \left[\frac{G(U_{\min[\tau,s]})}{R_{\min[\tau,s]}}\right],\qquad(17)$$

which provides a risk-neutral pricing relationship. The sup in equation (17), is over all stopping times (unlike τ) with respect to the filtration of U_t. We should point out at this juncture, that equation (17), and equation (14), are general enough to include situations where the guarantee expires at a certain age. This is the exact opposite of a maturity guarantee, and is common in some of the variable annuity policies. This essentially implies that if death occurs *after* a certain age, (age 80 for example) the payoff from the fund is limited to U_τ at the time of death, and not $G(U_\tau)$.

2.8 The Discounted Value of the Continuous Insurance Charges

In some cases, it might be important to compute the discounted value of the continuous insurance charge. First, let I_t^o, denote the stochastic – discounted

---

[10]The few companies who do offer accumulation rates greater than the risk free rate are obviously undertaking a large amount of credit risk.

to time zero – value of fees collected until time t, on a simple account of initial investment \$1. By construction, we have that:

$$dI_t^o = \frac{\alpha_t U_t}{R_t} dt. \tag{18}$$

The quantity $\alpha_t U_t dt$ can be viewed as the instantaneous cash flow to the insurance company, while the R_t^{-1} factor, discounts the quantity to time zero. We are interested in both the dynamics of I_t^o and its risk neutral expectation $\widetilde{E}_0[I_\xi^0]$, where ξ is a general stopping time for the process U_t. First, by a simple chain rule, we have that:

$$
\begin{aligned}
d(\frac{U_t}{R_t}) &= \frac{-r_t U_t}{R_t} dt + \frac{1}{R_t} dU_t \\
&= \frac{-r_t U_t}{R_t} dt + \frac{(r_t - \alpha_t) U_t}{R_t} dt + \frac{\sigma(U_t, t) U_t}{R_t} d\widetilde{B}_t \\
&= \frac{-\alpha_t U_t}{R_t} dt + \frac{\sigma(U_t, t) U_t}{R_t} d\widetilde{B}_t \\
&= -dI_t^0 + \frac{\sigma(U_t, t) U_t}{R_t} d\widetilde{B}_t.
\end{aligned}
\tag{19}
$$

Therefore, by rearranging equation (19), and recalling that $R_0^{-1} U_0 = 1$, we have that:

$$
\begin{aligned}
I_\xi^o &= \int_0^\xi dI_t^o = -\int_0^\xi d(\frac{U_t}{R_t}) + \int_0^\xi \frac{\sigma(U_t, t) U_t}{R_t} d\widetilde{B}_t \\
&= 1 - \frac{U_\xi}{R_\xi} + \int_0^\xi \frac{\sigma(U_t, t) U_t}{R_t} d\widetilde{B}_t.
\end{aligned}
\tag{20}
$$

The discounted value of the insurance risk charge, up to a stopping time (subject to standard integrability conditions) ξ, is: $1 - R_\xi^{-1} U_\xi$ plus an Ito (martingale) integral term, whose expectation is clearly zero. This implies:

$$\widetilde{E}_0[I_\xi^o] = 1 - \widetilde{E}_0\left[\frac{U_\xi}{R_\xi}\right] = 1 - \widetilde{E}_0\left[S_\xi e^{-\int_0^\xi (r_t + \alpha_t) dt}\right] \tag{21}$$

where the second equality comes from equation (3). In specific cases, equation (21) can be solved to provide the entire distribution of the discounted value of fees.

3 Analytic Solution.

3.1 Constant DSC and Hazard Rate with Guarantee at Death only.

In this section we make some assumptions on the structure of λ_t, α_t, k_t in order to derive complete analytic solutions for the optimal lapsation time. Specifically, we start by assuming that $\lambda_t = \lambda$, which implies that $\beta_t = \exp\{-\lambda t\}$. This is an exponential assumption for future lifetime, which can be calibrated to any mortality table by fixing the same life expectancy. For illustrative purposed, Figure 1 displays the relationship between the exponential distribution and a proper mortality table, assuming they share the same life expectancy, or first moment. Specifically, the graph shows the Cumulative Density Function (CDF) of the age-at-death random variable, under both assumptions. In terms of 'goodness of fit' to population mortality, the exponential assumption kills too many people early on, and lets too many live later. As well, the exponential assumption allows for a finite probability of surviving to any age. This is clearly unrealistic, but, we claim it provides a reasonable first approximation for the optimal lapsation time.

Also, since most insurance companies charge the same insurance fee, α, regardless of the age of the account holder – which implies that the aggregate hazard rate is independent of age – we believe that an exponential assumption can be justified based on common practice in the industry.

In this section, we further assume that $T = \infty$, which means that the guarantee only applies at death. Likewise, we let $g = 0$, so that $G(U_t) = \max[1, U_t]$ and the deferred surrender charge is $k_t = k$ and the insurance charge is $\alpha_t = \alpha$. The constant assumption for both these variable is less problematic, since, in practice this is usually the case. Finally, we assume a simplified geometric Brownian motion economy with $\sigma(U_t, t) = \sigma U_t$, $r_t = r$, and $R_t = \exp\{rt\}$.

The main pricing equation, originally presented in equation (13), can be written as:

$$V_t = V(t, u) = v_0 \sup_{t \le s} \widetilde{E}_t \left[(1-k)e^{-\lambda s}e^{-r(s-t)}U_s + \int_t^s \lambda e^{-\lambda q}e^{-r(q-t)} \max[1, U_q]dq \right],$$
(22)

where the first term in the expectation captures the surrender/lapsation value, and the second (integral) term captures the actual death benefit. Now,

Figure 1: The graph displays the relationship between the probability of survival under an exponential distribution assumption (solid line) and a realistic Society of Actuaries population mortality table (dashed line), assuming the same life expectancy of 20 years, at age 60.
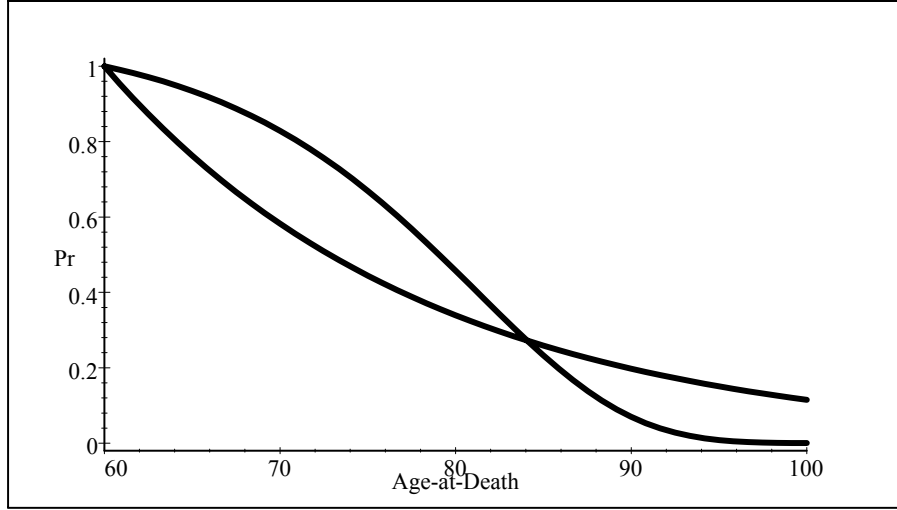
given the infinite maturity, we can re-write equation (22) as:

$$V(t, u) = e^{-\lambda t} v_0 W(u), \tag{23}$$

where

$$W(u) = \sup_{0 \leq s} \widetilde{E}_0 \left[(1 - k)e^{-(\lambda+r)s} U_s + \int_0^s \lambda e^{-(\lambda+r)q} \max[1, U_q] dq \right] \tag{24}$$

Using standard techniques obtained from the generator of the diffusion process – see Karatzas and Shreve (1998) for details – this leads to the following ordinary differential equation (ODE), satisfied by the valuation function $W(u)$;

$$(r - \alpha)uW'(u) + \frac{\sigma^2}{2}u^2 W''(u) - (\lambda + r)W(u) = -\lambda \max[1, u], \tag{25}$$

on the interval $(0, L)$, and $W(u) = (1 - k)u$, on the interval $[L, \infty)$, where L is the optimal lapsation boundary. Also, by construction, $W(1) = 1$.

17

The solution to this particular ODE, is:

$$W(u) = \begin{cases} \frac{\lambda}{\lambda+\alpha}u + b_1 u^{a_1} + b_2 u^{a_2} & 1 \le u \le L \\ \frac{\lambda + r u^{a_1}}{\lambda + r} & 0 \le u \le 1 \end{cases}, \tag{26}$$

where:

$$a_1 = \frac{-(r - \alpha - \frac{1}{2}\sigma^2) + \sqrt{(r - \alpha - \frac{1}{2}\sigma^2)^2 + 2\sigma^2(\lambda + r)}}{\sigma^2}, \tag{27}$$

$$a_2 = \frac{-(r - \alpha - \frac{1}{2}\sigma^2) - \sqrt{(r - \alpha - \frac{1}{2}\sigma^2)^2 + 2\sigma^2(\lambda + r)}}{\sigma^2} \tag{28}$$

and

$$b_1 = \frac{1}{a_1 - a_2}\left(\frac{a_1 r}{\lambda + r} - \frac{\lambda + \alpha a_2}{\lambda + \alpha}\right), \tag{29}$$

$$b_2 = \frac{1}{a_1 - a_2}\left(\frac{\lambda + \alpha a_1}{\lambda + \alpha} - \frac{a_1 r}{\lambda + r}\right) \tag{30}$$

The next step is to compute the optimal lapsation time, which is the level L, at which the fund should be lapsed, since the time dependency is irrelevant. This is similar to the process of computing the perpetual American put price by locating the lapsation level that maximizes the option value. In our case, the optimal lapsation boundary is at:

$$L = \left(\frac{b_2(1 - a_2)}{(a_1 - 1)b_1}\right)^{\frac{1}{a_1 - a_2}}, \tag{31}$$

and the corresponding[11] deferred surrender charge is:

$$k = 1 - \frac{\lambda}{\lambda + \alpha} - b_1 L^{(a_1-1)} - b_2 L^{(a_2-1)} \tag{32}$$

The feasible region for the continuous insurance is determined by the conditions $b_1 \ge 0$, and $L \ge 1$. This leads to a range of feasible values in between:

$$\alpha_L = \frac{\sigma^2}{2}\left[\frac{2r}{\sigma^2} - \frac{2(\lambda + r)}{a\sigma^2} - 1 + a\right], \tag{33}$$

---

[11]The details of this 'proof' are available from the authors upon request.

where a is the unique root of the cubic equation,

$$a^3 - a^2 + a\frac{2(\lambda + r)}{\sigma^2} - \frac{(\lambda + r)^2}{r\sigma^2} = 0 \tag{34}$$

and

$$\alpha_H = \frac{\sigma^2\lambda}{2r} \tag{35}$$

Naturally, a, the root of equation (34), must be obtained using (Newton) numerical methods. Note that when $\alpha = \alpha_L$, the DSC must be high enough to guarantee non-lapsation so that $L = \infty$. Likewise, when $\alpha = \alpha_H$, $L = 1$ and lapsation is instantaneous. The maximum contingent deferred surrender charge – that would make the product viable – is:

$$\overline{k} = 1 - \frac{\lambda}{\lambda + \alpha_L} = \frac{\alpha_L}{\lambda + \alpha_L}, \tag{36}$$

since $b_1 = 0$ and $L = \infty$. Stated differently, if the insurance company wants to levy the lowest possible fee on the variable annuity – denoted by α_L – they must charge $\overline{k}$ so they have enough to cover the hedging cost. *A Deferred Surrender Charge of $\overline{k}$ will complete the market.*

Finally, we conclude that the Total Expected (risk neutral) Discounted Fees consists of two portions The first portion is the present value of fees collected until *the earlier* of death or lapsation. The second portion is the present value of the DSC, *if* lapsation occurs prior to death. Consequently,

$$\text{TEDF} = \left[1 - \widetilde{E}_0\left[R^{-1}_{\min[\xi,\tau]}U_{\min[\xi,\tau]}\right]\right] + \widetilde{E}_0\left[R^{-1}_\xi kL, \xi < \tau\right] \tag{37}$$

$$= 1 - (1-k)L\widetilde{E}_0\left[R^{-1}_\xi, \xi < \tau\right] - \widetilde{E}_0\left[R^{-1}_\tau U_\tau, \xi \geq \tau\right] \tag{38}$$

$$= \frac{\alpha}{\lambda + \alpha}\left(1 - L^{1-a_1}\right) + kL^{1-a_1} \tag{39}$$

In particular, if $\alpha = \alpha_L$ (which implies that $L = \infty$), we have that $\text{TEDF} = \alpha/(\lambda + \alpha)$.

4   Numerical Examples.

In the following example we will assume an $r = 0.06$ interest rate environment, and a population of individuals, each with a life expectancy of exactly $1/\lambda = 20$ years. According to equations (35,33), a variable annuity with a

19

Figure 2: The graph displays the relationship between the continuous insurance charge (α), in basis points, and the optimal lapsation level (L). We assume an $r = 0.06$ interest rate, and a $1/\lambda = 20$ year life expectancy. The volatility is $\sigma_1 = 0.15$ (dashed line) and $\sigma_2 = 0.25$ (solid line).
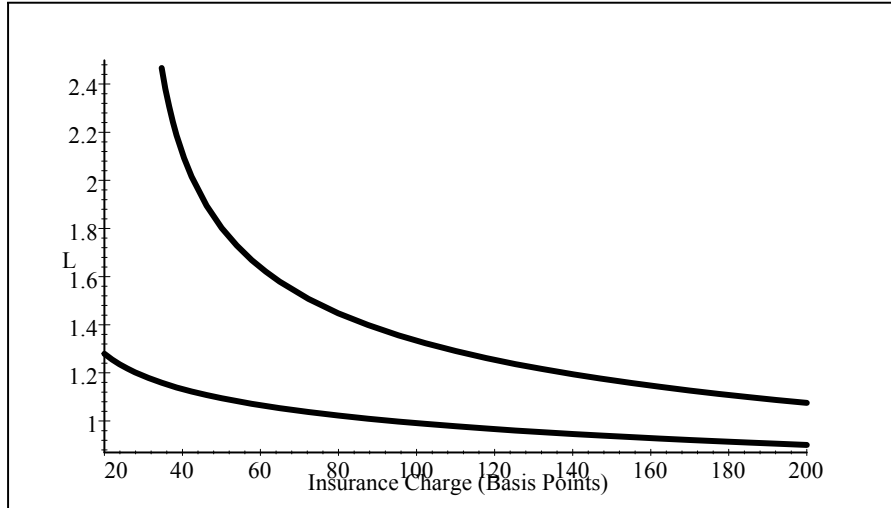
volatility of $\sigma = 15\%$, can 'afford' a continuous insurance expense charge of any number between $(10000\alpha_L =)$ 7.26 and $(10000\alpha_H =)$ 93.7 basis points. If, in fact, the lowest insurance charge, of 7.26 basis points is levied on the fund, then the company must charge a DSC of exactly $k = 0.0143$ ($\approx 1.4\%$), according to equation (36). The optimal lapsation level, L, will depend on the particular 'pair' (α, k) chosen by the company. For example if $\alpha = 10$ basis point, then $L = 1.564$, and the individual should lapse as soon as the fund appreciates by 56.4%.

Figure (2) displays the graphical relationship between the continuous insurance charge (in basis points) and the optimal lapsation level. Specifically, we assume a 6% interest rate and a 20 year life expectancy. The two curves represent a volatility of 15% and 25% respectively. Intuitively, a higher curve indicates a higher volatility because the value of the embedded option is higher and therefore one should wait longer – i.e. higher price appreciation – before discarding the old option. Also, *ceteris paribus*, as the insurance asset charge increases, the optimal lapsation level decreases. This is because the costs of holding the fund are higher relative to the value of the guarantee

Figure 3: The graph displays the relationship between the continuous insurance charge (α), in basis points, and the Deferred Sales Charge (k). We assume an interest rate of $r = 6\%$, and a life expectancy of $1/\lambda = 20$ years. The volatility is $\sigma_1 = 15\%$ (solid) and $\sigma_2 = 25\%$ (dashed).
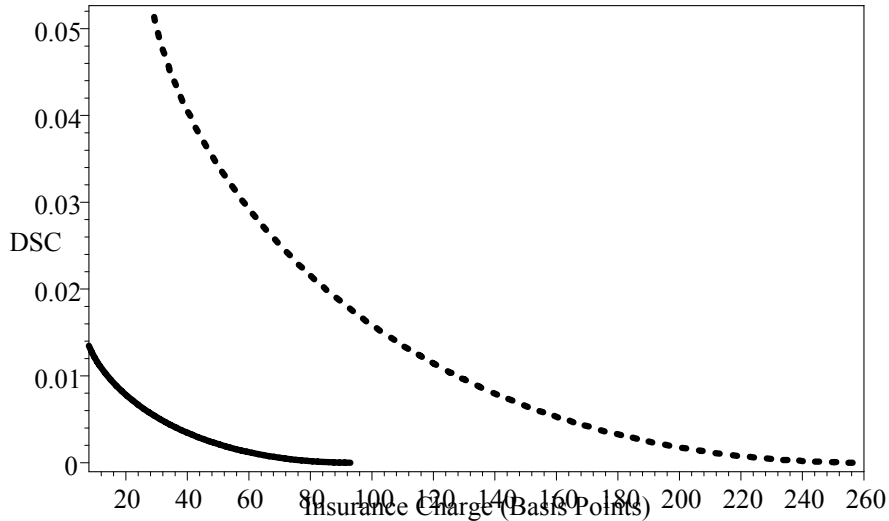
as well as the exit costs. In fact, although it is hard to see from the figure, if the insurance charge is exactly equal to α_H, the optimal lapsation level 'hits' a value of one. This means that one should lapse the contract instantaneously after purchasing it, or in other words, the contract should never be purchased. Stated differently, if $\alpha \geq \alpha_H$, the contract is not viable since no rational individual will ever hold it.

In the same vein, Figure (3) displays the relationship between the continuous insurance charge (in basis points), and the contingent deferred surrender charge. Once again, we assume a 6% interest rate and a 20 year life expectancy. The two curves represent a volatility of 15% and 25% respectively.

Perhaps at the risk of belaboring our main point, we emphasize that all combinations of (α, k) on the curve represent an appropriate charge for the embedded option. *In other words, there is no unique price for the contingent claim.* As the figure indicates – in the case of 25% volatility – the insurance company can pick a DSC of 5.3% and levy a continuous insurance charge of 28 basis points, or they can charge no DSC and impose a α of 260 basis

points. Either of these pairs – or any combination on the curve in between – will create enough money to construct a self financing portfolio to hedge the option. Of course, the optimal lapsation policy on the part of the individual will depend on which one of these pairs is chosen by the company, as evidenced by Figure (2). Nevertheless, there is no right or wrong continuous insurance charge, rather, an entire range of values can be justified, provided the contingent deferred surrender charge is properly selected.

Finally, as a summary, Table #1 displays the relationship between volatility σ and the feasible region for the continuous insurance charges. For example, a mutual fund with a volatility of 30%, should, at the very least, levy a continuous insurance expense charge of 42.3 basis points. This (low) amount would fund the death benefit, *only* if the company imposes a DSC of 7.8%. If the DSC is set lower than 7.8%, when $\alpha = 42.3$ basis point, the death benefit is un-hedgeable. Obviously, as intuition would dictate, if the company wants to charge a higher α than 42.3 basis points, they can charge a lower DSC.

A few points are in order. First, when the life expectancy is greater than 20 years, i.e. the hazard rate $\lambda < 1/20$, the feasible region (α_L, α_H) uniformly moves lower. In other words, the company can charge less for both (α, k). Intuitively, higher life expectancy is akin to lower volatility, which, as one can see in Table 1, has a dramatic effect on the values of α_L, α_H, and $\overline{k}$. For example, when $1/\lambda = 30$, which is roughly a 50 year old, we have that $\alpha_L = 22.1$ basis points, $\overline{k} = 6.2\%$ and $\alpha_H = 250$ basis points. In the same manner, a higher interest (discount) rate will also move the feasible region lower.

As a means of comparison, we contrast our numerical results to data supplied by Morningstar Inc., for the universe of 375 variable annuity policies (6,600 sub-accounts) sold in the United States. The median insurance charge levied on a simple *money back guarantee* variable annuity (fund) is 115 basis points. The median volatility for these funds is 18%, and the median surrender charge is 7%. According to equation (32), for a volatility of $\sigma = 0.18$, and $\alpha = 115$ basis points, with $r = 0.06$, and $1/\lambda = 20$, we get $k = 3.1 \times 10^{-4}$. If we take a more aggressive ten-year life expectancy $(1/\lambda = 10)$, we get $k = 9.2 \times 10^{-3}$. Both are nowhere near the median 7% reported by Morningstar. However, if we let $1/\lambda = 4.4$ years, we obtain 'fair' pricing.

Stated differently, according to our model, the average insurance charge on a variable annuity sold in the United States can be justified, if one assumes that all contract owners are completely rational and when the typical contract owner is expected to die in exactly 4.4 years. Furthermore, since

Assumptions: $r = 0.06, \lambda = 1/20$			
Volatility (σ)	$10000\alpha_L$	$100\overline{k}$	$10000\alpha_H$
10%	2.0	0.4	41.6
15%	7.2	1.4	93.7
20%	16.2	3.1	166.6
25%	28.2	5.3	260.4
30%	42.3	7.8	375.0
40%	74.2	12.9	666.6

Table 1: The Effect of Volatility on the Feasible Region

non-rational lapsation benefits the insurer, the reality of non-rational be-haviour implies that even a lower average life expectancy makes observed prices "fair". These results are consistent with the low numbers obtained in other estimates of the 'value' of the death benefit in variable annuities – see for example Milevsky and Posner (2000) and Windcliff, Forsyth and Vetzel (2000) – but fully accounts for lapsation in an economically parsimonious manner.

5   Conclusion.

In a recent issue (July 2000, page 48) of the *Dow Jones Investment Advisor*, a financial planner was quoted as saying:

> "...with the guarantee on a variable annuity providing a protected floor...then if the investments pan out, we can always 1035-exchange[12] the client into another contract that allows us to establish a new higher death benefit. A few months ago we had a 73-year old client in tech stocks...when the account appreciated by 40% we did a 1035-exchange. The old contract had the old floor...we got a new contract with a 40% higher floor....This strategy only works...with products that contain no surrender charges...We like to structure it so it does not cost the client anything..."

[12]A 1035-exchange is the practitioner terminology for the section of the United States income tax act (ITA) which allows this transaction without inducing a taxable event.

As the above article indicates, it appears that Real Option to lapse is valuable and is quite popular. Although we obviously disagree with the second part of the quote, namely that it only 'makes sense' when there are no surrender charges. In any event, our paper has examined a very basic question faced by all investors in variable annuity contracts. The issue at hand is: *At what point should I lapse my contract and re-establish the basis of the guarantee?* We view the personal ability to lapse the investment as a Real Option that is analogous to the abandonment option, or the option to shut down, in classical corporate finance. The symmetric opposite side to this question is: *What is the optimal asset-based charge that funds the guarantee, assuming investors will lapse rationally?* Indeed, the answer to both questions lie in the structure of the deferred surrender charge.

Our paper's main contribution lies in highlighting the critical importance of the deferred surrender charge (DSC) in *completing* the market and allowing the claim to be hedged. In some sense, one can say that transaction costs (frictions) complete the market and allow for the existence of a self-financing strategy. The self-financing strategy does not result in a unique price, per se, but rather a menu or schedule of charges that can support the claim. Technically, we formulated the problem of *when to lapse* as a free boundary problem and provided a closed-form analytic solution when hazard rates were constant.

Our model should enable users to answer any of the following questions:

1. By how much does the account value have to appreciate, before it is optimal to exercise the Real Option to lapse the contract, and re-establish a new (higher) basis?

2. Assuming a particular continuous insurance charge, what is the lowest contingent deferred surrender charge that will allow the company to recoup its hedging costs? Likewise, assuming a fixed DSC, what is the lowest continuous insurance charge that will allow the insurance company to recoup its hedging costs?

3. How high can the continuous insurance charge be set, while still maintaining a viable product? In other words, how expensive does the product have to be, for it to be optimal to *never purchase* the contract?

Further research will attempt to 'solve' the optimal lapsation policy for more general hazard rates, underlying processes and maturity guarantees.

24

Also, the authors will examine some of the issues pertaining to hedging these guarantees, especially as it relates to using exotic products, such as Barrier options, that take advantage of lapsation behavior. Finally, the authors will use some of the recent work on minimizing Shortfall Risk, which is in the spirit of an actuarial approach, as an alternative to No Arbitrage valuation.

References

[1] K. Aase and S. Persson. "Pricing of Unit-Linked Life Insurance Policies". *Scandinavian Actuarial Journal*, 1:26–52, 1994.

[2] M.-O. Albizzati and H. Geman. "Interest Rate Risk Managment and Valuation of the Surrender Option in Life Insurance Policies". *Journal of Risk and Insurance*, 61(4):616–637, 1994.

[3] M. Amram and N. Kulatilaka. *Real Options: Managing Strategic Investment in an Uncertain World*. Harvard Business School Press, 1999.

[4] A. Bacinello and F. Ortu. "Pricing of Equity-Linked Life Insurance Policies with Endogenous Minimum Guarantees". *Insurance: Mathematics and Economics*, 12:245–257, 1993.

[5] A. Bacinello and F. Ortu. "Fixed Income Linked Life Insurance Policies with Minimum Guarantees: Pricing Models and Numerical Results". *European Journal of Operational Research*, 91:235–249, 1996.

[6] J. B. Berk. "A Simple Approach for Deciding When to Invest". *American Economic Review*, 89(5):1319–1326, December 1999.

[7] F. Black and M. S. Scholes. "The Pricing of Options and Corporate Liabilities". *The Journal of Political Economy*, 81:637–653, 1973.

[8] P. Boyle, A. Kolkiewicz, and K. Tan. "Valuation of the Reset Option in Segregated Fund Contracts Using Quasi-Monte Carlo Methods". *University of Waterloo Institute of Insurance and Pension Research Report*, 99-10, 1999.

[9] P. Boyle and E. Schwartz. "Equilibrium Prices of Guarantees Under Equity-Linked Contracts". *Journal of Risk and Insurance*, 44:639–680, 1977.

[10] M. Brennan and E. Schwartz. "The Pricing of Equity-Linked Life Insurance Policies with an Asset Value Guarantee". *Journal of Financial Economics*, 3:195–213, 1976.

[11] P. Carr. "The Valuation of Sequential Exchange Opportunities". *The Journal of Finance*, 43(5):1235–1256, December 1988.

[12] P. Carr. "Randomization and the American Put". *The Review of Financial Studies*, 11(3):597–626, Fall 1998.

[13] R. M. Dammon and R. C. Green. "Tax Arbitrage and the Existence of Equilibrium Prices for Financial Assets". *The Journal of Finance*, 42(3):1143–1166, 1987.

[14] J. Dermody and R. Rockafellar. "Cash Stream Valuation in the Face of Transaction Costs". *Mathematical Finance*, 1(1):41–54, 1991.

[15] S. Ekern and S.-A. Persson. "Exotic Unit-Linked Life Insurance Contracts". *The Geneva Papers on Risk and Insurance Theory*, 21:35–63, 1996.

[16] M. Garman and J. Ohlson. "Valuation of Risky Assets in Arbitrage Free Economies with Transaction Costs". *Journal of Financial Economics*, 9(3):271–280, 1981.

[17] R. Geske and H. Johnson. "The American Put Value Analytically". *Journal of Finance*, 39:1511–1524, 1984.

[18] A. Grosen and P. L. Jorgensen. "Valuation of Early Exercisable Interest Rate Guarantees". *Journal of Risk and Insurance*, 64(3):481–503, 1997.

[19] J. Huang, M. Subrahmanyam, and G. Yu. "Pricing and Hedging American Options: A Recursive Integration Method'. *Review of Financial Studies*, 9:277–300, 1996.

[20] R. Hubbard. "Investment Under Uncertainty: Keeping One's Options Open". *Journal of Economic Literature*, 32(4):1816–1831, December 1994.

[21] J. E. Ingersoll and S. A. Ross. "Waiting to Invest: Investment and Uncertainty". *Journal of Business*, 65(1):1–29, 1992.

[22] S. Jacka. "Optimal Stopping and the American Put". *Mathematical Finance*, 1:1–14, 1991.

[23] H. Johnson and R. Stulz. "The Pricing of Options with Default Risk". *The Journal of Finance*, 42(2):267–280, 1987.

[24] I. Karatzas and S. Shreve. *Methods of Mathematical Finance*, volume 39 of *Applications of Mathematics Stochastic Modeling and Applied Probability*. Springer, New York, 1998.

[25] I. Kim. "The Analytical Valuation of American Options". *Review of Financial Studies*, 3:547–572, 1990.

[26] H. P. Mckean. "Appendix: A Free Boundary Problem for the Heat Equation Arising from a Problem in Mathematical Economics". *Industrial Managment Review*, 6:32–39, 1965.

[27] R. C. Merton. "Theory of Rational Option Pricing". *Bell Journal of Economics and Management Science*, 4:141–183, 1973.

[28] M. Milevsky and S. Posner. "The Titanic Option: Valuation of the Guaranteed Minimum Death Benefit in Variable Annuities and Mutual Funds". *The Journal of Risk and Insurance*, March, 2000.

[29] K. R. Miltersen and S.-A. Persson. "Pricing Rate of Return Guarantees in a HJM Framework". *Insurance: Mathematics and Economics*, 25:307–325, 1999.

[30] J. A. Nielsen and K. Sandman. "Equity-Linked Life Insurance: A Model with Stochastic Interest Rates". *Insurance: Mathematics and Economics*, 16:225–253, 1995.

[31] N.Ju. "Pricing an American Option by Approximating its Early Exercise Boundary as a Multipiece Exponential Function". *Review of Financial Studies*, 11:627–646, 1998.

[32] S.-A. Persson and K. Aase. "Valuation of the Minimum Guaranteed Return Embedded in Life Insurance Products". *Journal of Risk and Insurance*, 64:599–617, 1997.

[33] E. Z. Prisman. "Valuation of Risky Assets in Arbitrage Free Economies with Frictions". *The Journal of Finance*, 41(3):545–556, 1986.

[34] E. Ronn. " A New Linear Programming Approach to Bond Portfolio Management". *Journal of Financial and Quantitative Analysis*, 22(4):439–466, 1987.

[35] S. Ross. "Uses, Abuses and Alternatives to the Net Present Value Rule". *Financial Management*, 24(3):96–102, 1995.

[36] S. Shreve and J. Vecer. "Options on a Traded Account: Vacation Calls, Vacation Puts and Passport Options". *Finance and Stochastics*, 4(3):255–274, May 2000.

[37] L. Trigeorgis. *Real Options*. The MIT Press, Cambridge, Massachusetts, 1996.

[38] H. Windcliff, P. Forsyth, and K. Vetzel. "Valuation of Segregated Funds: Shout Options with Maturity Extensions". *University of Waterloo Working Paper*, 2000.

6 Appendix

Typos and minor changes in the body of the paper.

- Section 2.4: after "the hazard rate is assumed to be determinstic", add "and to be non-integrable".

- Formula (11): "$0 \leq t \leq T$" should be "$0 \leq t < T$".

- Section 2.6: delete the phrase "k_t is small enough and the".

- Formula (15): the final denominator should be R_T. And a few lines later, the appropriate condition should be that $k_0 > 1$.

- Section 2.6, end. Add: "The converse is true when the rates r_t and λ_t are constant. In that case the second term in the above expression equals

$$\frac{\lambda}{\lambda + r - g}\left(1 - e^{-(\lambda+r-g)T}\right) + e^{-(\lambda+r-g)T},$$

which exceeds one for every T, provided $g > r$. In other words, the guarantee is perfectly fundable for some α exactly when $k_0 > 0$ and $r > g$."

- Formula (17): E_0^* should be $\tilde{E}_0$

- Section 3, after (35): Change "$\alpha = \alpha_H, L = 1$" to "$\alpha = \alpha_H$, it follows that $k = 0, L = 1$,"

- Section 3, after (36). Add before the last sentence of this paragraph: "If $k \geq \bar{k}$ then L will be infinite, and the only α which perfectly funds the guarantee is $\alpha = \alpha_L$."

6.1 Generalizing the model

The model of the paper is one in which a closed formula solution can be obtained for the DSC k that perfectly hedges the contract for a given insurance charge α. Much of the same analysis applies to more general models, and for this reason, we will carry out the calculations for the more general model, as far as possible. At the end, the general model will produce k as the root

of an equation. In the context of the paper however, an explicit formula is possible, and we will point out this simplification.

We incorporate three effects into this model. The first already occurred in the early sections of the paper, namely a growth rate $g \geq 0$ for the guarantee (in section 3 and 4 of the paper we took $g = 0$ however).

The second effect is the inclusion of irrational lapsation at a fixed rate λ_2, typically around 10%. We now write λ_1 for the death rate, and set $\lambda = \lambda_1 + \lambda_2$. Of course, the original model hedged against all possible lapses, rational or irrational. Irrational lapsation leads to surplus revenue for the policy writer though, and including this effect into the model permits hedging fees to be lowered and made more competitive. We model such lapsation as occurring at a fixed rate, irrespective of the behaviour of the underlying mutual fund. The irrationality is only from the point of view of the writer of the contracts – many of the policyholders doubtless have good reasons for lapsing that simply have nothing to do with the economic value of the policies (eg. cash flow). Since poor judgement is not assumed to be the cause of lapsation, we do not assume any corresponding irrationality when it comes to failing to lapse. In other words, the worst-case scenario that we hedge against is still one in which there is a stopping time S such that the entire cohort of indiviuals who haven't previously died or lapsed (irrationally) will opt to lapse simultaneously at time S. More complicated models, which we have not considered, might try to stratify the initial population of policyholders into groups with different lapsation behaviours (eg. never lapse, only lapse irrationally, only lapse rationally, etc.)

The final effect we incorporate reflects the reality that most of the continuous insurance charges actually imposed do not get used for hedging purposes. Instead they get allocated to trailer fees, management expenses, etc. To model this, we break the total continuous insurance charge α into two components, $\alpha = \alpha_1 + \alpha_2$, where α_1 is the insurance charge reserved for hedging purposes, and the charge allocated to other purposes is α_2 (specified exogenously).

In other words, our model is as follows. The asset underlying the policy still obeys

$$dS_t = rS_t \, dt + \sigma S_t \, d\tilde{B}_t,$$

and the account value $U_t = e^{-\alpha t} S_t$ obeys

$$dU_t = (r - \alpha)U_t \, dt + \sigma U_t \, d\tilde{B}_t.$$

At the time ξ of lapsation we still recover kU_t for the hedge. Of course, one could also incorporate additional non-hedging DSCs. Unlike non-hedging insurance charges, these have no effect on how the contracts are hedged, since their only effect is that the client recovers less than $(1 - k)U_t$ upon lapsing. For this reason we ignore such a possibility.

The fraction of contracts still being held at time $t < \xi$ is now

$$\beta_t = e^{-\lambda t} = e^{-(\lambda_1 + \lambda_2)t},$$

and the book value of such contracts is $\beta_t v_0$. The outflow from the hedge is

$$c_t \, dt = v_0 \beta_t [\lambda_1 (U_t \vee e^{gt}) + [\lambda_2 (1 - k) + \alpha_2] U_t] \, dt$$

since funds leave the account steadily in three ways – death, irrational lapsation, and non-hedging insurance charges. For convenience, write

$$\gamma = \lambda_2 (1 - k) + \alpha_2. \tag{40}$$

To be hedged, the value of the hedging portfolio must satisfy the condition

$$V_t \geq L_t = \begin{cases} v_0 (1 - k) \beta_t U_t & , t < T \\ v_0 \beta_t (U_t \vee e^{gt}) & , t = T \end{cases}$$

for every t. As in Karatzas and Shreve (1998), the portfolio that hedges the aggregate accounts then satisfies

$$V_t = \sup_{T \geq S \geq t} \tilde{E} \left[e^{-r(S-t)} L_S \right.$$
$$\left. + \int_t^S v_0 e^{-r(q-t)} e^{-\lambda q} (\lambda_1 (U_q \vee e^{gq}) + \gamma U_q) \, dq \mid \mathcal{F}_t \right],$$

where the supremum is taken over stopping times $T \geq S \geq t$. As before, we wish to choose the free parameters g, α_1, and k to ensure that $V_0 = v_0$.

With a finite maturity T, it is hopeless to look for analytic solutions, and one has no recourse but to turn to numerical techniques, for example, solving a free boundary value problem numerically as in the corresponding problem for American options. Instead we work here with $T = \infty$, so that the only way to end a contract is to lapse or to die. Work in progress involves generalizations of the problem, including the case of a finite maturity. This is discussed briefly in section **??**.

32

To eliminate the t variable, we work with a discounted account value $Y_t = e^{-gt}U_t$. Then

$$V_t = v_0 e^{(g-\lambda)t} \sup_{S \geq t} \tilde{E}[(1-k)e^{-(\lambda+r-g)(S-t)}Y_S$$

$$+ \int_t^S e^{-(\lambda+r-g)(q-t)}(\lambda_1(Y_q \vee 1) + \gamma Y_q)\, dq],$$

and the last expression is independent of t by the Markov property. In other words,

$$V_t = v_0 e^{(g-\lambda)t} W(e^{-gt}U_t)$$

where

$$W(y) = \sup_{S \geq 0} \tilde{E}[(1-k)e^{-(\lambda+r-g)S}Y_S$$

$$+ \int_0^S e^{-(\lambda+r-g)q}(\lambda_1(Y_q \vee 1) + \gamma Y_q)\, dq],$$

$$dY_t = (r - \alpha - g)Y_t\, dt + \sigma Y_t\, d\tilde{B}_t,$$

$$Y_0 = y.$$

The question before us is that of finding parameters so that $W(1) = 1$. The approach we will take is the standard optimal stopping argument, in which it is assumed that the optimal S is the first time Y_t exceeds some level L, and a corresponding free boundary problem is solved. This gives a candidate for the function W. One then verifies that it is the solution using an argument via the Snell envelope. Existence and uniqueness is addressed separately, using monotonicity and the intermediate value theorem.

Having described the general model, we now proceed to work through the paper, providing technical details and justifications in places where they were originally omitted. Throughout, we will be careful to indicate places where our comments are meant to apply to the general model, rather than to the special cases considered in the paper ($\alpha_2 = 0$, $\lambda_2 = 0$, and in places, $g = 0$).

6.2 Commentary on section 2.4 of the paper

In the penultimate paragraph to section 2.4, it is stated that the worst-case scenario, the one that has to be hedged against, is that all remaining

individuals opt to lapse simultaneously at the worst possible time for the writer of the policies. This is clear from econimic considerations, and we have built the simultaneous lapsation principle into our model. A more general model could be constructed in which this is a consequence rather than a hypothesis, but we have not included this here. There are situations in which such an analysis would be useful however. For example, if the policy allows the writer to dynamically adjust the parameters within some given range, the writer might react to partial lapsation at some non-optimal time by using the surplus earnings to lower α and so delay future lapses. In such cases, the hedging strategy would have be adapted to react to partial lapsation. We hope to persue these issues in future work.

6.3 Commentary on section 2.6 of the paper

Consider the general model first. We need to address the following issue from the end of the section. Fix values of the exogenous variables $\lambda_1 \geq 0$, $\lambda_2 \geq 0$, $\alpha_2 \geq 0$, $r > 0$, and $\sigma > 0$. When do there exist values of the three free variables $\alpha_1 \geq 0$, $k \in (0,1]$, and $g \geq 0$ that make $V_0/v_0 = 1$, and to what extent are these values unique? In particular, given values for two of the free variables, does the above condition uniquely determine a value of the third variable? Note that we rule out the case $k = 0$, as in that case we expect the locking in of any gains to produce instantaneous lapsation.

Recalling that $U_t = e^{-\alpha_t} S_t$, we have that

$$\frac{V_0}{v_0} = \sup_{0 \leq s \leq T} \tilde{E}[(1-k)e^{-(r+\alpha+\lambda)s} S_s 1_{\{s<T\}} + e^{-(r+\lambda)s}(e^{-\alpha s} S_s \vee e^{gs}) 1_{\{s=T\}}$$

$$+ \int_0^s e^{-(\lambda+r)q}(\lambda_1(e^{-\alpha q} S_q \vee e^{gq}) + \gamma e^{-\alpha q} S_q) \, dq]. \quad (41)$$

Since $\gamma = \lambda_2(1-k) + \alpha_2$, (41) is easily seen to be continuous and monotone in the free variables $\alpha_1 \geq 0$, $k \in (0,1]$, and $g \geq 0$. It is decreasing in α_1 and k, and is increasing in g.

If $\lambda_2 > 0$ then (41) is strictly monotone in all three variables, at least while the optimal s satisfies $\tilde{P}(s = 0) < 1$. To see this observe that the integrand itself is strictly monotone (notice that $e^{-\alpha q} S_q - e^{gq}$ takes both positive and negative values for q arbitrarily close to 0). When $P(s = 0) = 1$ it follows that lapsation is instantaneous, and $V_0/v_0 = 1 - k < 1$. Thus, for any value of two of the three free variables, there is at most one value of the remaining variable that makes $V_0/v_0 = 1$.

If $\lambda_2 = 0$ then (41) is still strictly monotone in α_1 and g, but does not vary with k once the optimal s satisfies $P(s = T) = 1$. In this case, there will typically be a value $\bar{k}$ depending on α_1 and g such that V_0/v_0 is strictly monotone in k for $k < \bar{k}$, and is constant for $k \geq \bar{k}$.

We will analyze this expression further, in the special case that $T = \infty$. To ensure convergence, we must have $g < \lambda + r$. Taking limits as $\alpha_1 \to \infty$ gives

$$\frac{V_0}{v_0} = \int_0^\infty \lambda_1 e^{-(\lambda+r)q} e^{gq} \, dq = \frac{\lambda_1}{\lambda + r - g}.$$

If $g > \lambda_2 + r$ then this expression exceeds 1, so by monotonicity there will be no parameter values for which $V_0/v_0 = 1$, and the portfolio is unhedgeable. In other words, though values of g between $\lambda_2 + r$ and $\lambda + r$ could be hedged by charging an initial fee, they cannot be hedged just on the basis of DSC's or continuous insurance charges. And values of g exceeding $\lambda + r$ are simply unsupportable under any circumstances.

So in what follows, fix a value of g such that $g < \lambda_2 + r$. Then $V_0/v_0 < 1$ for α_1 sufficiently large, uniformly in k. At the other extreme, consider what happens when $\alpha_1 = k = 0$. In that case, since $e^{-rq} S_q$ is a martingale,

$$\begin{aligned}
\frac{V_0}{v_0} &= \sup_s \tilde{E}\Big[\int_s^\infty (\lambda + \alpha_2) e^{-(\lambda+\alpha_2)q} e^{-rq} S_q \, dq \\
&\quad + \int_0^s e^{-(\lambda+r)q} \big(\lambda_1 (e^{-\alpha_2 q} S_q \vee e^{gq}) + (\lambda_2 + \alpha_2) e^{-\alpha_2 q} S_q\big) \, dq\Big] \\
&> \sup_s \tilde{E}\Big[\int_s^\infty (\lambda + \alpha_2) e^{-(\lambda+\alpha_2)q} e^{-rq} S_q \, dq \\
&\quad + \int_0^s (\lambda + \alpha_2) e^{-(\lambda+\alpha_2)q} e^{-rq} S_q \, dq\Big] \\
&= \int_0^\infty (\lambda + \alpha_2) e^{-(\lambda+\alpha_2)q} \, dq = 1.
\end{aligned}$$

Appealing to monotonicity, we see that the values of α_1 such that there exists a $k \in [0,1]$ making $V_0/v_0 = 1$, form an interval $[\alpha_L, \alpha_H]$, where $0 \leq \alpha_L < \alpha_H < \infty$. This is the feaxible interval – the range of parameters for which there exists a perfect hedge. Moreover, when $\alpha_1 = \alpha_H$, the hedging value of k must be $k = 1$, otherwise α_1 could be increased still further.

Consider first the case $\lambda_2 > 0$. The parameter values are constrained by the inequalities $\alpha_L \geq 0$ and $k \leq 1$. By strict monotonicity, at least one of these inequalities must hold with equality at the lower endpoint the

feasible interval. In fact, substituting numerical values easily shows that either possibility can occur. For example, if λ_2 is large it may be possible to hedge the contract even without using continuous insurance charges for hedging, in effect using the DSC of lapsing customers to hedge the guarantee of the remaining customers. On the other hand, if λ_2 is small but non-zero it will turn out that this is impossible, and that $k = 1$ when $\alpha_1 = \alpha_L$.

If $\lambda_2 = 0$, on the other hand, then there is a further constraint to take account of, because at some point V_0/v_0 ceases to be strictly monotone in k. This will make it impossible to raise k to compensate for further lowering α_1. In fact, we will see later that in this case, this effect defines the value α_L, and that for $\alpha_1 = \alpha_L$, a range of values $k \in [\bar{k}, 1]$ will hedge the contract. The choice of k in this range is irrelevant, because even at $k = \bar{k}$, the optimal strategy is to never lapse.

In fact, there is a related effect even in the case $\lambda_2 > 0$. It will turn out that there is an intermediate value α_M of interest. Under the perfect hedge, the optimal lapsation time is a.s. infinite when $\alpha_1 \in [\alpha_L, \alpha_M]$, and is finite with positive probability when $\alpha_1 \in (\alpha_M, \alpha_H]$. If $\lambda_2 = 0$ then $\alpha_L = \alpha_M$, but otherwise we have $\alpha_L < \alpha_M$. When λ_2 is small but nonzero, the two values are very close of course, which is why $k = 1$ when $\alpha_1 = \alpha_L$ in this situation.

6.4   Commentary on Section 3: Snell's envelope

In the next section we are going to derive the function which is the conjectured solution to the optimal stopping problem discussed above. It is worth keeping in mind what properties this function must have, before we can conclude that it really does solve the problem. Recall that the solution to an optimal stopping problem

$$\sup_{s \geq 0} E[A_s]$$

(where s ranges over stopping times) is given by the Snell envelope. In other words, in its simplest form, by a process M_t such that

- M_t is a uniformly integrable continuous supermartingale

- $M_t \geq A_t \geq 0$ for every t

- $M_t^* = M_{t \wedge \xi}$ is a martingale, where $\xi = \inf\{t > 0 \mid M_t = A_t\}$.

- $\lim_{t \to \infty} M_t = \lim_{t \to \infty} A_t$ a.s. on $\{\xi = \infty\}$

36

Then for any stopping time s we have that

$$M_0 \geq E[M_s] \geq E[A_s],$$

and

$$M_0 = M_0^* = E[M_\xi^*] = E[A_\xi].$$

We take

$$M_t = e^{-(\lambda+r-g)t}W(Y_t) + \int_0^t e^{-(\lambda+r-g)q}(\lambda_1(Y_q \vee 1) + \gamma Y_q)\, dq$$

$$A_t = (1-k)e^{-(\lambda+r-g)t}Y_t + \int_0^t e^{-(\lambda+r-g)q}(\lambda_1(Y_q \vee 1) + \gamma Y_q)\, dq$$

Assume that

$$g < \lambda + r. \tag{42}$$

In fact, to actually find a suitable function W satisfying $W(1) = 1$, we'll end up needing the stronger condition $g < \lambda_2 + r$. But the latter is not required for this part of the argument.

What will be required of $W(y)$ is as follows. There is a value $L \in (1, \infty]$ such that:

- $W(y) \geq (1-k)y$ for every $y > 0$.

- $W(1) = 1$.

- $W(y)$ is bounded near $y = 0$.

- $W(y)$ is C^2 on $(0, L)$.

- $\mathcal{L}W(y) - (\lambda + r - g)W(y) = -(\lambda_1(y \vee 1) + \gamma y)$ on $(0, L)$, where $\mathcal{L}$ is the generator of Y_t.

If $L = \infty$ then there is a $k' \leq k$ such that

- $W(y)/y \to (1 - k')$ as $y \to \infty$.

If $L < \infty$ then there is a $k' \leq k$ such that

- $W(y) > (1 - k')y$ for $y < L$ and $W(y) = (1 - k')y$ for $y > L$.

37

- $W(y)$ is convex and C^1 in some neighbourhood of $y = L$.

Note that when $L < \infty$, the only situation in which it will turn out that $k' \neq k$ is that $\lambda_2 = 0$ and $\alpha_1 = \alpha_L$. But when $L = \infty$ it will turn out that $k' < k$ for $\alpha_1 \in [\alpha_L, \alpha_M)$.

Given the above conditions, the necessary verification goes as follows. First of all, $e^{-(\lambda+r-g)t}Y_t$ is a geometric Brownian motion with negative drift, hence is uniformly integrable and converges to 0 as $t \to \infty$. By local boundedness and linear growth of W, as well as the fact that $g < \lambda + r$, the same is true for $e^{-(\lambda+r-g)t}W(Y_t)$. Using the exponential decay of the integrand, in fact M_t is uniformly integrable as well. Moreover, $\lim_{t\to\infty} M_t = \lim_{t\to\infty} A_t$. Since W is C^2 away from L we can apply Ito's lemma up to ξ, the first time Y hits L. Because of the ODE satisfied by W, it follows that M^* is a local martingale. By uniform integrability, in fact it is a martingale. The supermartingale property follows similarly, the only potential problem being near times t when $Y_t = L < \infty$. Those get handled using the smooth pasting condition, namely that W is C^1 and convex near L.

6.5 Finite Horizon

Before proceeding to solve the optimal stopping, it is worth recording the form the problem takes if some of our simplifying assumptions (constant k, infinite time horizon, exponential mortality) are relaxed. For example, in this section we'll exogenously specify a DSC schedule k_t. Our goal will be to determine the insurance charge α_1 that allows the product to be hedged.

We will use the notation $\tilde{E}_{t,u}$ to denote expectations for the process $(U_s)_{s \geq t}$ under the risk-neutral probability measure making $U_t = u$ a.s. Write λ_t^1 for the hazard rate at time t for the population who bought into the contract at time 0. We still assume that this population is large enough that the hazard rate can be taken to be deterministic, but we no longer assume that the rate is constant. A commonly specification is that of Gompertz mortality, which can often be adjusted to provide a good fit to empirical mortality rates.

Likewise we assume that the rate of irrational lapsation λ_t^2 (of investors still invested at time t) is deterministic, and not dependent, say, on the performance of the market. Deciding how to specify λ_t^2 is in fact one of the most proprietary aspects of corporate modelling of variable annuities. As before, the net hazard rate is $\lambda_t = \lambda_t^1 + \lambda_t^2$.

As before, the value of the hedging portfolio at time t, as a function of the value u of the underlying account, is

$$\sup_{T \geq S \geq t} \tilde{E}_{t,u} \left[e^{-r(S-t)} L(S, U_S) + \int_t^S v_0 e^{-r(q-t)} \beta_q G(q, U_q) \, dq \right], \qquad (43)$$

where $\gamma_t = \lambda_t^2 (1 - k_t) + \alpha_2$, $\beta_t = e^{-\int_0^t \lambda_q \, dq}$,

$$L(t, u) = \begin{cases} v_0 (1 - k_t) \beta_t u & , t < T \\ v_0 \beta_t (u \vee e^{gt}) & , t = T, \end{cases}$$
$$G(t, u) = \lambda_q^1 (U_q \vee e^{gq}) + \gamma_q U_q.$$

To calculate this supremum we set

$$M_t = e^{-rt} V(t, U_t) + \int_0^t v_0 e^{-rq} \beta_q G(q, U_q) \, dq,$$

$$A_t = e^{-rt} L(t, U_t) + \int_0^t v_0 e^{-rq} \beta_q G(q, U_q) \, dq,$$

where $V(t, u)$ is chosen to make the conditions on M_t from the previous section hold. This being so, the same argument as given earlier shows that

$$M_t = \sup_{T \geq S \geq t} \tilde{E}[A_S \mid t],$$

in which case simple algebra shows that

$$V(t, U_t) = \sup_{T \geq S \geq t} \tilde{E} \left[e^{-r(S-t)} L(S, U_S) + \int_t^S v_0 e^{-r(q-t)} \beta_q G(q, U_q) \, dq \mid \mathcal{F}_t \right],$$

so (43) equals $V(t, u)$.

As before, $V(t, u)$ must satisfy a variety of conditions in order to carry through the argument, of which we record simply the three most important. The first comes from the requirement that the contract be hedgeable on the basis of the initial account deposits. In other words, that

$$V(0, 1) = v_0. \qquad (44)$$

The second expresses the fact that the contract is hedged against lapsation, namely that

$$V(t, u) \geq L(t, u) \qquad (45)$$

39

for every t, u. Finally, the requirement that M_t be a supermartingale (and a martingale when stopped at the time of optimal lapsation) implies that $V(t, u)$ satisfies a PDE for values of (t, u) making the inequality (45) is strict. To read off this PDE, note that

$$dM_t = e^{-rt}\Big[- rV(t, U_t) + \frac{\partial V}{\partial t}(t, U_t) + \mathcal{L}V(t, U_t)$$
$$+ v_0 \beta_t G(t, U_t)\Big] dt + d_t$$

where $_t$ is a local martingale, and

- $\tilde{V}(0, 1) = 1$.

- $\tilde{V}(t, u) \geq \tilde{L}(t, u) = \begin{cases} (1 - k_t)u, & t < T \\ u \vee e^{gt}, & t = T. \end{cases}$

- $\frac{\partial \tilde{V}}{\partial t} + (r - \alpha)u\frac{\partial \tilde{V}}{\partial u} + \frac{\sigma^2}{2}u^2\frac{\partial^2 \tilde{V}}{\partial u^2} - (r - \lambda_t)\tilde{V} = -G$.

Note, as a check of consistency, that these conditions are indeed satisfied in the case
$$\tilde{V}(t, u) = e^{gt}W(e^{-gt}u)$$
of the preceding sections.

Work currently in progress will examine the numerical solution of such problems. Note that the condition $\tilde{V}(0, 1) = 1$ means that an iterative scheme is called for. In particular, the algorithm takes a preliminary value of α_1, solves the PDE using a coarse grid, updates the value of α_1 depending on whether the derived value of $\tilde{V}(0, 1)$ overshoots or undershoots, then re-solves the PDE with a finer grid, etc. The adjustment made to α_1 is made using an empirical Newton scheme, based on the change observed over the previous iteration. An alternative numerical procedure familiar from the American option literature, which is also suitable for use as part of an iterative solution scheme, would first translate the PDE into an integral equation.

6.6   Commentary on Section 3: the free BVP

We carry out the argument, as far as possible, with the generalized model described above. Recall that we are searching for a $W(y)$ satisfying the list

of conditions given in section **??**. In particular, W solves the inhomogeneous ODE

$$(r - \alpha - g)yW'(y) + \frac{\sigma^2}{2}y^2W''(y) - (\lambda + r - g)W(y)$$
$$= -[\lambda_1(1 \vee y) + \gamma y]. \quad (46)$$

The general solution to the corresponding homogeneous ODE is

$$A_1 y^{a_1} + A_2 y^{a_2},$$

where $a_1 > a_2$ satisfy

$$\frac{\sigma^2}{2}a(a - 1) + (r - \alpha - g)a - (\lambda + r - g) = 0. \quad (47)$$

The value of this expression when $a = 0$ and $a = 1$ are $-(\lambda + r - g)$ and $-(\lambda + \alpha)$ respectively. By (42), both values are strictly negative. As a result, we have that $a_1 > 1$ and $a_2 < 0$. Since $W(y)$ is bounded near $y = 0$, it follows that $A_2 = 0$. Of course, we have explicitly that

$$a_1 = \frac{-(r - \alpha - g - \frac{1}{2}\sigma^2) + \sqrt{(r - \alpha - g - \frac{1}{2}\sigma^2)^2 + 2\sigma^2(\lambda + r - g)}}{\sigma^2} \quad (48)$$

$$a_2 = \frac{-(r - \alpha - g - \frac{1}{2}\sigma^2) - \sqrt{(r - \alpha - g - \frac{1}{2}\sigma^2)^2 + 2\sigma^2(\lambda + r - g)}}{\sigma^2}. \quad (49)$$

Note that neither a_1 nor a_2 depend on k. As well, from the tilting of (47) we have that a_1 and a_2 are increasing in α, with $a_1 \to \infty$ and $a_2 \to 0$ as $\alpha \to \infty$.

By direct substitution, a particular solution of (46) on $(0, 1)$ is

$$\frac{\lambda_1}{\lambda + r - g} + \frac{\gamma}{\lambda + \alpha}y.$$

Since $W(1) = 1$, it follows that

$$W(y) = \frac{\lambda_1}{\lambda + r - g}(1 - y^{a_1}) + \frac{\gamma}{\lambda + \alpha}(y - y^{a_1}) + y^{a_1}. \quad (50)$$

Similarly, a particular solution of (46) on $(1, \infty)$ is

$$\frac{\gamma + \lambda_1}{\lambda + \alpha}y.$$

41

Thus on $(1, L)$ our function has the form

$$W(y) = \frac{\gamma + \lambda_1}{\lambda + \alpha} y + b_1 y^{a_1} + b_2 y^{a_2}. \tag{51}$$

We require that W be C^2 on $(0, L)$, which will be the case provided the values of $W(1)$ and $W'(1)$ agree from the left and right. In other words, we need the following conditions:

$$\frac{\gamma + \lambda_1}{\lambda + \alpha} + b_1 a_1 + b_2 a_2 = -a_1 \frac{\lambda_1}{\lambda + r - g} + (1 - a_1) \frac{\gamma}{\lambda + \alpha} + a_1 \tag{52}$$

$$\frac{\gamma + \lambda_1}{\lambda + \alpha} + b_1 + b_2 = 1. \tag{53}$$

Multiplying (53) by a_1 and subtracting (52), we get that

$$(a_1 - a_2) b_2 = a_1 \lambda_1 \left[\frac{1}{\lambda + r - g} - \frac{1}{\lambda + \alpha} \right] + \frac{\lambda_1}{\lambda + \alpha} \tag{54}$$

$$= a_1 \left[\frac{(\lambda + \alpha) - \lambda_1}{\lambda + \alpha} - \frac{(\lambda + r - g) - \lambda_1}{\lambda + r - g} \right] + \frac{\lambda_1}{\lambda + \alpha}.$$

In other words,

$$b_2 = \frac{1}{a_1 - a_2} \left[\frac{\lambda_1 + (\alpha + \lambda_2) a_1}{\lambda + \alpha} - \frac{a_1 (\lambda_2 + r - g)}{\lambda + r - g} \right], \tag{55}$$

$$b_1 = 1 - b_2 - \frac{\gamma + \lambda_1}{\lambda + \alpha}. \tag{56}$$

Note that the expression for b_2 is independent of k and generalizes formula (30) of the paper. Further, (54) can be written in the form

$$(a_1 - a_2) b_2 = \frac{\lambda_1}{\lambda + \alpha} \left[a_1 \frac{\alpha + g - r}{\lambda + r - g} + 1 \right]. \tag{57}$$

But by (47),

$$a_1 \frac{\alpha + g - r}{\lambda + r - g} + 1 = \frac{\sigma^2 a_1 (a_1 - 1)}{2(\lambda + r - g)}.$$

Thus

$$b_2 = \frac{\lambda_1 \sigma^2 a_1 (a_1 - 1)}{2(\lambda + \alpha)(a_1 - a_2)(\lambda + r - g)} > 0. \tag{58}$$

42

6.7 CASE 1:

$L < \infty$.

We first treat the case that $L < \infty$. In that case, $W(y)/y$ and $W'(y)$ must both equal $1 - k$ when $y = L$. In other words,

$$1 - k = b_2 a_2 L^{a_2 - 1} + b_1 a_1 L^{a_1 - 1} + \frac{\gamma + \lambda_1}{\lambda + \alpha} \tag{59}$$

$$= b_2 L^{a_2 - 1} + b_1 L^{a_1 - 1} + \frac{\gamma + \lambda_1}{\lambda + \alpha}. \tag{60}$$

Therefore

$$b_1(a_1 - 1)L^{a_1 - 1} = b_2(1 - a_2)L^{a_2 - 1}, \tag{61}$$

and so

$$L = \left[\frac{b_2(1 - a_2)}{b_1(a_1 - 1)} \right]^{\frac{1}{a_1 - a_2}} \tag{62}$$

which gives formula (31) of the paper. Note as well that $b_1 \geq 0$. Moreover, if $\lambda_2 = 0$ then not only b_2, but b_1 as well, will be independent of k. Thus (62) determines L and so (60) determines k as in (32) of the paper.

If $\lambda_2 \neq 0$, we must solve for L numerically instead. Subtracting (59) from a_1 times (60) gives the equation

$$(1 - k)(a_1 - 1) = b_2(a_1 - a_2)L^{a_2 - 1} + (a_1 - 1)\frac{\gamma + \lambda_1}{\lambda + \alpha}.$$

Substituting (40) and solving, it follows that

$$1 - k = \frac{b_2(a_1 - a_2)L^{a_2 - 1}(\lambda + \alpha) + (a_1 - 1)(\alpha_2 + \lambda_1)}{(a_1 - 1)(\lambda_1 + \alpha)}. \tag{63}$$

Using (56) and (40), equation (60) becomes that

$$(1 - k)\left[1 - \frac{\lambda_2}{\lambda + \alpha}(1 - L^{a_1 - 1}) \right]$$
$$= b_2 L^{a_2 - 1} + (1 - b_2)L^{a_1 - 1} + \frac{\lambda_1 + \alpha_2}{\lambda + \alpha}(1 - L^{a_1 - 1}).$$

In other words,

$$(1 - k)[\lambda_1 + \alpha + \lambda_2 L^{a_1 - 1}]$$
$$= b_2(\lambda + \alpha)L^{a_2 - 1} + (1 - b_2)(\lambda + \alpha)L^{a_1 - 1} + (\lambda_1 + \alpha_2)(1 - L^{a_1 - 1}).$$

43

Substituting (63) gives that

$$\lambda_1 + \alpha_2 + \frac{b_2(\lambda + \alpha)(a_1 - a_2)}{a_1 - 1} L^{a_2 - 1}$$
$$+ \frac{\lambda_2(\lambda_1 + \alpha_2)}{\lambda_1 + \alpha} L^{a_1 - 1} + \frac{\lambda_2 b_2(\lambda + \alpha)(a_1 - a_2)}{(a_1 - 1)(\lambda_1 + \alpha)} L^{a_1 + a_2 - 2}$$
$$= b_2(\lambda + \alpha) L^{a_2 - 1} + (1 - b_2)(\lambda + \alpha) L^{a_1 - 1} + (\lambda_1 + \alpha_2)(1 - L^{a_1 - 1}).$$

Collecting terms,

$$b_2(\lambda + \alpha)\frac{1 - a_2}{a_1 - 1} L^{a_2 - 1} + (\lambda + \alpha)\left[b_2 - \frac{\alpha_1}{\lambda_1 + \alpha}\right] L^{a_1 - 1}$$
$$+ \frac{\lambda_2 b_2(\lambda + \alpha)(a_1 - a_2)}{(a_1 - 1)(\lambda_1 + \alpha)} L^{a_1 + a_2 - 2} = 0.$$

That is,

$$\frac{b_2(1 - a_2)}{a_1 - 1} + \left[b_2 - \frac{\alpha_1}{\lambda_1 + \alpha}\right] L^{a_1 - a_2} + \frac{\lambda_2 b_2(a_1 - a_2)}{(a_1 - 1)(\lambda_1 + \alpha)} L^{a_1 - 1} = 0. \qquad (64)$$

For there to be a positive root, the middle coefficient must be negative, that is, $b_2 < \alpha_1/(\lambda_1 + \alpha)$. In this case it is easily seen that there is exactly one root. It can be found simply using Newton$'$smethod, whereuponkisfoundusing(63).

Having worked our way down to an equation for L, let us turn the argument around, and verify that we really have obtained a solution to the desired problem, and under what conditions we have done so.

The parameters $g \geq 0$, $r \geq 0$, $\lambda_1 \geq 0$, $\lambda_2 \geq 0$, and $\alpha_2 \geq 0$ are given, and (42) is assumed. Fix an $\alpha_1 > 0$. Formulae (48), (49), and (58) then define $a_1 > 1$, $a_2 < 0$, and $b_2 > 0$. Assume that

$$b_2 < \frac{\alpha_1}{\lambda_1 + \alpha}, \qquad (65)$$

in which case (64) has a unique positive solution L. Assume further that

$$L > 1. \qquad (66)$$

Define $1 - k \geq 0$ by (63), and then b_1 by (56). Equations (59) and (60) hold, from which (61) follows, so that $b_1 \geq 0$. Defining the function $W(y)$ by (50)

and (51) gives us a solution to (46) on $(0, L)$, which is tangent to the line $(1 - k)y$ at $y = L$. Take $W(y) = (1 - k)y$ for $y \geq L$.

For $y \in (1, L)$,

$$W''(y) = a_1(a_1 - 1)b_1 y^{a_1 - 2} + a_2(a_2 - 1)b_2 y^{a_2 - 2} > 0.$$

For $y \in (0, 1)$,

$$W''(y) = a_1(a_1 - 1)y^{a_1 - 2} \left[1 - \frac{\lambda_1}{\lambda + r - g} - \frac{\gamma}{\lambda + \alpha} \right], \qquad (67)$$

which does not change sign and is positive at $y = 1$ (by continuity of W''). Thus W is convex on $[0, L]$. Since $(1 - k)y$ is the tangent line at $y = L$, we actually have that $W(y) \geq (1 - k)y$ for $y \in [0, L]$, and that W is convex on $[0, \infty)$. Since $W(1) = 1$, also $1 - k \leq 1$. It follows that W satisfies all the properties demanded in section **??**, so that W really does give the solution to the optimal stopping problem. Note also that by (67) we have that

$$1 - \frac{\lambda_1}{\lambda + r - g} > 0,$$

or in other words, we recover the condition

$$g < \lambda_2 + r. \qquad (68)$$

which is stronger than (42). Where we stand is that, assuming (68), the conditions required to make the above argument work are precisely (65), and (66). In sections **??** and **??** we identify the range of parameters for which these conditions hold as an interval (α_M, α_H).

6.8 CASE 2:

$L = \infty$.

The main argument in this section will only be relevant when $\lambda_2 > 0$. We reserve discussion of the case $\lambda_2 = 0$ for section **??**.

Since $L = \infty$, we are searching for solutions to the PDE on $(0, \infty)$. In view of the linear growth condition, we consider solutions with $b_1 = 0$. But by (56) this means that

$$1 - b_2 = \frac{\gamma + \lambda_1}{\lambda + \alpha}.$$

In other words,

$$1 - k = -\frac{1}{\lambda_2} \left[(\lambda + \alpha) b_2 - \alpha_1 - \lambda_2. \right] \tag{69}$$

Set $k' = b_2 > 0$. Then

$$W(y) = (1 - k')(y - y^{a_2}) + y^{a_2}$$

for $y > 1$, so that $W(y)/y \to 1 - k'$ as $y \to \infty$. Assume that the k determined by (69) satisfies

$$1 - k \geq 0. \tag{70}$$

Then

$$1 - k' = \frac{(1 - k)\lambda_2 + \alpha_2 + \lambda_1}{\lambda + \alpha} \geq 0$$

too, and as before, it is easy to see that W is convex. Thus $W(y) \geq 1 - k'$ for every $y > 0$. We have almost shown that W satisfies the conditions of section **??**. The only additional constraint, other than (70) is that

$$k' \leq k.$$

In section **??** we will identify an $\alpha_M > 0$ such that (65) implies that $\alpha_1 > \alpha_M$. So assume that

$$0 \leq \alpha_1 \leq \alpha_M. \tag{71}$$

Then $b_2 \geq \alpha_1/(\lambda_1 + \alpha)$. In other words,

$$1 - k' = 1 - b_2 \leq \frac{\lambda_1 + \alpha_2}{\lambda_1 + \alpha}.$$

Therefore

$$k - k' = (1 - k') - \frac{(\lambda + \alpha)(1 - k') - \alpha_2 - \lambda_1}{\lambda_2}$$

$$= \frac{\alpha_2 + \lambda_1 - (\lambda_1 + \alpha)(1 - k')}{\lambda_2} \geq 0,$$

as required. The upshot is that, subject to the conditions (70) and (71), we have found a range of solutions to the optimal stopping problem, with $L = \infty$.

6.9 Commentary on Section 3: α_L and α_H

In this section, we will derive the statements made concerning α_L and α_H, namely that (34) of the paper has a unique solution, and that formulas (33) and (35) of the paper hold. Because it simplifies the arguments, we assume in this section that $\alpha_2 = \lambda_2 = 0$. We will take up the general case in section ??. In the current section we must also take up the issue, left unaddressed in subsection ??, of solutions to the problem with $\lambda_2 = 0$ and $L = \infty$.

Before starting, observe that the constraint that defines α_L in the paper is that $b_1 = 0$, and the constraint defining α_H is that $L = 1$. Observe also that by (68), we are assuming that $g < r$.

To obtain α_H, note that $W'(L) = 1 - k$. When $L = 1$ we must have $k = 0$, so $W'(1) = 1$. In other words,

$$a_1 = \frac{\lambda + r - g}{r - g} = 1 + \frac{\lambda}{r - g}. \tag{72}$$

Substituting this into equation (47) shows that

$$
\begin{aligned}
0 &= (r - g - \alpha)\frac{\lambda + r - g}{r - g} + \frac{\sigma^2}{2} \cdot \frac{\lambda + r - g}{r - g} \cdot \frac{\lambda}{r - g} - (\lambda + r - g) \\
&= \frac{\lambda + r - g}{r - g}\left(\frac{\sigma^2 \lambda}{2(r - g)} - \alpha\right),
\end{aligned}
$$

from which equation (35) of the paper follows.

Turning to α_L, note that both a_1 and a_2 are solutions to the equation

$$\frac{\sigma^2}{2}a(a - 1) + (r - g - \alpha)a - (\lambda + r - g) = 0.$$

In particular, both satisfy

$$r - g - \alpha = \frac{\lambda + r - g}{a} + \frac{\sigma^2}{2}(1 - a), \tag{73}$$

and so also

$$\alpha = r - g - \frac{\lambda + r - g}{a} - \frac{\sigma^2}{2}(1 - a). \tag{74}$$

Note that the latter expression is exactly formula (33) of the paper. In fact, what we will argue is that when $b_1 = 0$ it follows that a_1 satisfies equation (34) of the paper.

First note that by (73),

$$\frac{\lambda + r - g}{a_1} - \frac{\sigma^2 a_1}{2} = \frac{\lambda + r - g}{a_2} - \frac{\sigma^2 a_2}{2}.$$

Thus

$$(\lambda + r - g) \left(\frac{a_2 - a_1}{a_1 a_2} \right) = (\lambda + r - g) \left(\frac{1}{a_1} - \frac{1}{a_2} \right) = \frac{\sigma^2}{2} (a_1 - a_2).$$

In other words,

$$a_2 = -\frac{2(\lambda + r - g)}{\sigma^2 a_1}. \tag{75}$$

Now assume that $b_1 = 0$, or in other words that $b_2 = \alpha/(\lambda + \alpha)$. By (58) we have that

$$\frac{\sigma^2 a_1 (a_1 - 1)}{2} = \alpha (a_1 - a_2) \frac{\lambda + r - g}{\lambda}$$

and thus (47) becomes that

$$\alpha (a_1 - a_2) \frac{\lambda + r - g}{\lambda} + (r - g - \alpha) a_1 - (\lambda + r - g) = 0.$$

Substitute (74) and (75) and simplify, to get the equation

$$0 = a_1^4 + a_1^3 \left[\frac{2(\lambda + r - g)}{\sigma^2} - 1 \right]$$

$$+ a_1 \left[2 \left(\frac{\lambda + r - g}{\sigma^2} \right)^2 \left(2 - \frac{\sigma^2}{r - g} \right) \right] - 4 \left(\frac{\lambda + r - g}{\sigma^2} \right)^3 \cdot \frac{\sigma^2}{r - g}.$$

Or, writing

$$\eta = \frac{2(\lambda + r - g)}{\sigma^2}, \qquad \delta = \frac{\sigma^2}{2(r - g)}, \qquad a = a_1$$

we have

$$a^4 + a^3 (\eta - 1) + a \eta^2 (1 - \delta) - \eta^3 \delta = 0. \tag{76}$$

But $a = -\eta$ solves this equation, and factoring this out, we see that

$$a^3 - a^2 + a\eta - \eta^2 \delta = 0, \tag{77}$$

48

which is exactly equation (34) of the paper.

By definition, we have that

$$\eta > 0, \qquad \eta\delta > 1 \tag{78}$$

So what remains to be shown is that (77) has a unique root, under the conditions (78).

But

$$a^3 - a^2 + a\eta - \eta^2\delta = (a^2 + \eta)(a - 1) - \eta(\delta\eta - 1).$$

The first term has a unique root $a = 1$, and is strictly increasing on $(1, \infty)$. Since $\eta(\delta\eta - 1) > 0$ by (78), it follows that the equation (77) has a unique root a as well. Of course this root also satisfies $a > 1$. It can be found numerically very efficiently using Newton's method. Note that it is also strictly less than the value of a_1 corresponding to $\alpha = \alpha_H$. This can be seen directly from the above equations: by (72), this a_1 equals $\eta\delta$, and substituting $\eta\delta$ into the polynomial yields a strictly positive outcome.

To summarize, we have obtained equations that determine values $0 < \alpha_L < \alpha_H < \infty$ such that we can exhibit a solution to the optimal stopping problem whenever $\alpha \in (\alpha_L, \alpha_H]$. But we have actually done more than this. Taking $\alpha = \alpha_L$, and W as in section **??**, we have a solution to the PDE, with $L = \infty$ and $b_1 = 0$. Let $1 - k' \in (0, 1)$ be the value determined by equation (63). Then this W and α are easily seen to satisfy the conditions of section **??** for every $k \in [k', 1]$. Thus we have actually exhibited a solution to the equation $V_0/v_0 = 1$ for every choice of $k \in [0, 1]$. By the uniqueness argument of section **??**, we have found the entire class of solutions.

6.10 $\quad \alpha_L$, α_M, and α_H in general

In this section, we extend the previous arguments to the case $\lambda_2 \neq 0$, $\alpha_2 \neq 0$. We obtain computational formulae satisfied by α_M and α_L, and discuss the complete solution to the optimal stopping problem. Finally we look at the behaviour of these solutions in the limiting case, that $g \uparrow \lambda_2 + r$.

First observe that the general constraint giving rise to α_L was (65). When $\lambda_2 = 0$ this was equivalent to having $b_1 > 0$, which was the condition of the paper, and of the preceding section.

Consider α_H. As before, when $L = 1$ we have $k = 0$. To ensure that $L \geq 1$, it must be the case that substituting $L = 1$ in (64) gives a result that

is ≥ 0. In other words, we must have that

$$b_2 \frac{a_1 - a_2}{a_1 - 1} \left[1 + \frac{\lambda_2}{\lambda_1 + \alpha} \right] \geq \frac{\alpha_1}{\lambda_1 + \alpha}.$$

Thus by (58),

$$\alpha_1 \leq b_2 \frac{(a_1 - a_2)(\lambda + \alpha)}{a_1 - 1} = \frac{\lambda_1 \sigma^2 a_1}{2(\lambda + r - g)}. \tag{79}$$

The expression of (47) is ≤ 0 for $a \in [0, a_1]$, so (79) is equivalent to

$$\alpha_1 \frac{\lambda + r - g}{\lambda_1} \left[\frac{2(\lambda + r - g)}{\lambda_1 \sigma^2} \alpha_1 - 1 \right]$$
$$+ (r - g - \alpha_1 - \alpha_2) 2\alpha_1 \frac{\lambda + r - g}{\lambda_1 \sigma^2} - (\lambda + r - g) \leq 0.$$

Or in other words,

$$2\alpha_1^2 \left[\frac{\lambda + r - g}{\lambda_1} - 1 \right] + \alpha_1 [-\sigma^2 + 2(r - g - \alpha_2)] - \lambda_1 \sigma^2 \leq 0.$$

Since $\lambda_2 + r - g > 0$, this is equivalent to $\alpha_1 \leq \alpha_H$, where

$$\alpha_H = \frac{\lambda_1}{4(\lambda_2 + r - g)} \left[\sigma^2 - 2(r - g - \alpha_2) \right.$$
$$\left. + \sqrt{(\sigma^2 - 2(r - g - \alpha_2))^2 + 8\sigma^2 (\lambda_2 + r - g)} \right]. \tag{80}$$

Turning to α_M, the argument proceeds as before, except that now that the constraint is that $b_2 < \alpha_1/(\lambda_1 + \alpha)$, or in other words, that

$$\frac{\sigma^2 a_1 (a_1 - 1)}{2} < \alpha_1 (a_1 - a_2) \left(\frac{\lambda + r - g}{\lambda_1} \right) \left(\frac{\lambda + \alpha}{\lambda_1 + \alpha} \right)$$
$$= (\alpha - \alpha_2)(a_1 - a_2) \left(\frac{\lambda + r - g}{\lambda_1} \right) \left(\frac{\lambda + \alpha}{\lambda_1 + \alpha} \right),$$

and thus by (47), that

$$(\alpha - \alpha_2)(a_1 - a_2) \left(\frac{\lambda + r - g}{\lambda_1} \right) \left(\frac{\lambda + \alpha}{\lambda_1 + \alpha} \right)$$
$$+ (r - g - \alpha)a_1 - (\lambda + r - g) > 0.$$

50

By (47),

$$\frac{\sigma^2}{2}a(a-1) + (\lambda_1 + r - g)a - (\lambda + r - g) = (\lambda_1 + \alpha)a > 0.$$

Multiplying these two inequalities, and substituting (74) and (75) leads to a more complicated relation than before, but it may still be solved numerically to give a value for a_1 as a root of a polynomial $p(a)$, after which (79) determines α and hence $\alpha_1 = \alpha_M$.

Carrying out this algebra using Maple yields the following form of the constraint:

$$p(a)\frac{(a-1)\sigma^6}{8a^2\lambda_1} > 0,$$

where

$$p(a) = \theta a^5 + a^4\left[\frac{4\lambda\lambda_1}{\sigma^4} - \theta(1+\theta) + \eta(3\theta - \rho)\right] + a^3\eta(\eta(\eta - \rho) - \theta)$$
$$+ a^2\eta^2(\eta - \rho - 1) + a\eta^3(\eta - \rho - 2) - \eta^4. \quad (81)$$

Here

$$\eta = \frac{2(\lambda + r - g)}{\sigma^2}, \quad \theta = \frac{2(\lambda_2 + r - g)}{\sigma^2}, \quad \rho = \frac{2(\lambda + \alpha_2)}{\sigma^2} \quad a = a_1$$

In order for (81) to be really useful, we must verify that the corresponding equation has a unique root $a > 1$.

Simplifying $p(a)$, we can write it in the form

$$p(a) = \frac{4\lambda_1\lambda_2}{\sigma^4}a^4 + (a + \eta)\left[(a-1)(a^3\theta + \eta^2(a^2 + a + \eta)) - \eta\rho a(a^2 + \eta)\right].$$

In consequence,

$$p(1) = \frac{4\lambda_1\lambda_2}{\sigma^4} - \eta\rho(1+\eta)^2 < 0.$$

Since $\theta > 0$, there will always be at least one root $a > 1$.

To address uniqueness, observe that

$$\frac{p(a)}{a(a+\eta)(a^2+\eta)}$$
$$= \frac{4a^3\lambda_1\lambda_2}{\sigma^4(a+\eta)(a^2+\eta)} - \eta\rho + \frac{(a-1)(a^3\theta + \eta^2(a^2 + a + \eta))}{a(a^2+\eta)}.$$

Taking derivatives in Maple gives the following expression:

$$\frac{4a^2\lambda_1\lambda_2\eta(2a + a^2 + 3\eta)}{\sigma^4(a+\eta)^2(a^2+\eta)^2} + \frac{a\theta(a^3 + a\eta + 2\eta(a-1))}{(a^2+\eta)^2}$$
$$+ \frac{\eta^2(2a^3 + 3a^2\eta + \eta^2)}{a^2(a^2+\eta)^2},$$

each term of which is seen to be non-negative when $a > 1$. Thus $p(a)$ has exactly one such root. We denote the corresponding value of α_1 by α_M. In combination with the results of section **??**, we have therefore obtained a solution to the optimal stopping problem with $1 \leq L < \infty$, for every $\alpha_1 \in (\alpha_M, \alpha_H]$.

If $\lambda_2 > 0$, the results of section **??** provide a value $\alpha_L < \alpha_M$ and exhibit solutions to the optimal stopping problems with $L = \infty$, for every $\alpha_i \in [\alpha_L, \alpha_M]$. Thus solutions have actually been found for $\alpha_1 \in [\alpha_L, \alpha_H]$. If (69) is nonegative when $\alpha_1 = 0$, then $\alpha_L = 0$. Otherwise, α_L is the value of α_1 making (69) vanish. Either way, the uniqueness argument of section **??** shows that we have found the entire class of solutions.

Similarly, if $\lambda_2 = 0$, we set $\alpha_L = \alpha_M$ and apply the results at the very end of (**??**) (which were actually derived with $\alpha_2 = 0$, but the same arguments work when $\alpha_2 > 0$). Again, a range of solutions was obtained for $\alpha_1 = \alpha_L$, which completes the description of the full class of solutions.

To close this section, we consider what happens in the limiting case, that $g \uparrow r + \lambda_2$. First of all, from (80), it is clear that $\alpha_H \to \infty$, as $g \uparrow r + \lambda_2$. Similarly, as $\theta \downarrow 0$, $p(a)$ converges to

$$a^4[\frac{4\lambda\lambda_1}{\sigma^4} + -3\eta\rho] + a^3\eta^2(\eta - \rho) + a^2\eta^2(\eta - \rho - 1) + a\eta^3(\eta - \rho - 2) - \eta^4,$$

where $\eta = 2\lambda_1/\sigma^2 < \rho$. Since every term of the latter is negative, the positive root of $p(a)$ must be diverging to ∞ in the limit, and thus so does α_M. Finally, consider α_L. We may as well assume that $\lambda_2 > 0$. Consider the limiting form of (69). By (57) and then (47), it is

$$1 - k = \frac{-1}{\lambda_2(a_1 - a_2)}[a_1(\alpha + \lambda_2) + \lambda_1 - (\alpha_1 + \lambda_2)(a_1 - a_2)]$$
$$= \frac{-1}{\lambda_2(a_1 - a_2)}[\alpha_2(a_1 - a_2) + \lambda_1 + a_2(\alpha + \lambda_2)]$$
$$= \frac{-1}{\lambda_2(a_1 - a_2)}\left[\alpha_2(a_1 - a_2) + \frac{\sigma^2}{2}a_2(a_2 - 1)\right] < 0$$

In consequence, every α_1 is eventually $< \alpha_L$, so that $\alpha_L \to \infty$. Of course, this asymptotic implies those for α_M and α_H, but it is also useful to see these directly.

6.11 Commentary on Section 3: TEDF

In the calculation of the total expected discounted fees, at the end of Section 3, we have τ (the time of death) an exponential time with mean $1/\lambda$, independent of the price process. The lapsation time is denoted ξ, which is a stopping time with respect to the filtration of U_t, namely the first hitting time of L by Y_t. In other words, the first time U_t equals Le^{gt}. We will follow the calculation under the assumption that $\lambda_2 = \alpha_2 = 0$. As in the paper,

$$
\begin{aligned}
\text{TEDF} &= \left[1 - \tilde{E}_0\left[R_{\xi \wedge \tau}^{-1} U_{\xi \wedge \tau}\right]\right] + \tilde{E}_0\left[R_\xi^{-1} k U_\xi, \xi < \tau\right] \\
&= 1 - \tilde{E}_0\left[R_\xi^{-1} Le^{g\xi}, \xi < \tau\right] \\
&\quad - \tilde{E}_0\left[R_\tau^{-1} U_\tau, \xi \geq \tau\right] + \tilde{E}_0\left[R_\xi^{-1} k Le^{g\xi}, \xi < \tau\right] \\
&= 1 - (1-k)L\tilde{E}_0\left[R_\xi^{-1} e^{g\xi}, \xi < \tau\right] - \tilde{E}_0\left[R_\tau^{-1} U_\tau, \xi \geq \tau\right]. \quad (82)
\end{aligned}
$$

Consider the second term in (82). Write $_t$ for the filtration of U_t. By independence,

$$
\begin{aligned}
\tilde{E}_0\left[R_\xi^{-1} e^{g\xi}, \xi < \tau\right] &= \tilde{E}_0\left[e^{-(r-g)\xi} \tilde{P}_0(\xi < \tau \mid \xi)\right] \\
&= \tilde{E}_0\left[e^{-(r-g+\lambda)\xi}\right].
\end{aligned}
$$

The latter can be read off as a Laplace transform of a passage time for a Brownian motion with drift. Alternatively, since

$$
\frac{a_1^2 \sigma^2}{2} + a_1(r - g - \alpha - \frac{\sigma^2}{2}) - (r - g + \lambda) = 0,
$$

we have that

$$
\begin{aligned}
\tilde{E}_0\left[e^{-(r-g+\lambda)\xi}\right] &= \tilde{E}_0\left[e^{-a_1(r-g-\alpha-\frac{\sigma^2}{2})\xi - \frac{a_1^2\sigma^2}{2}\xi}\right] \\
&= \tilde{E}_0\left[(e^{-g\xi} U_\xi)^{-a_1} e^{-\frac{a_1^2\sigma^2}{2}\xi + a_1\sigma \tilde{B}_\xi}\right] \\
&= L^{-a_1}.
\end{aligned}
$$

53

Thus the second term (82) equals

$$-(1-k)L^{1-a_1}.$$

Turning to the third term in (82), we have by the martingale property that

$$\tilde{E}_0\left[R_\tau^{-1}U_\tau, \xi \geq \tau\right] = \tilde{E}_0\left[e^{\sigma \tilde{B}_\tau - \frac{\sigma^2}{2}\tau}e^{-\alpha\tau}, \xi \geq \tau\right]$$
$$= \tilde{E}_0\left[e^{\sigma \tilde{B}_\xi - \frac{\sigma^2}{2}\xi}e^{-\alpha\tau}, \xi \geq \tau\right]$$
$$= \tilde{E}_0\left[e^{\sigma \tilde{B}_\xi - \frac{\sigma^2}{2}\xi}\tilde{E}_0\left[e^{-\alpha\tau}, \xi \geq \tau \,|_\xi\right]\right].$$

Since τ has an exponential distribution, a simple integration shows that this equals

$$\tilde{E}_0\left[e^{\sigma \tilde{B}_\xi - \frac{\sigma^2}{2}\xi} \cdot \frac{\lambda}{\lambda + \alpha}(1 - e^{-(\lambda+\alpha)\xi})\right]$$
$$= \frac{\lambda}{\lambda + \alpha}\left(E_0\left[e^{\sigma \tilde{B}_\xi - \frac{\sigma^2}{2}\xi}\right] - E_0\left[e^{\sigma \tilde{B}_\xi - (\alpha+\lambda+\frac{\sigma^2}{2})\xi}\right]\right)$$
$$= \frac{\lambda}{\lambda + \alpha}\left(1 - E_0\left[U_\xi e^{-(r+\lambda)\xi}\right]\right)$$
$$= \frac{\lambda}{\lambda + \alpha}\left(1 - E_0\left[Le^{-(r-g+\lambda)\xi}\right]\right)$$
$$= \frac{\lambda}{\lambda + \alpha}\left(1 - L^{1-a_1}\right).$$

Finally, we combine the above expressions, to get that

$$\text{TEDF} = 1 - (1-k)L^{1-a_1} - \frac{\lambda}{\lambda + \alpha}\left(1 - L^{1-a_1}\right)$$
$$= \frac{\alpha}{\lambda + \alpha}\left(1 - L^{1-a_1}\right) + kL^{1-a_1},$$

which is the formula given in the paper.
End of Appendix.