Analysis of optimal dynamic withdrawal policies in withdrawal guarantees products

Yao Tung Huang and Yue Kuen Kwok*

Department of Mathematics
Hong Kong University of Science and Technology
Hong Kong, China

Abstract

The guaranteed minimum withdrawal benefits (GMWB) are popular riders in variable annuities with withdrawal guarantees. With withdrawals spread over the life of the annuities contract, the benefit promises to return the entire initial annuitization amount irrespective of the market performance of the underlying fund portfolio. Treating the dynamic withdrawal rate as the control variable, the earlier works have considered the construction of a continuous singular stochastic control model and the numerical solution of the resulting pricing model. This paper presents a more detailed characterization of the behavior of the GMWB price function and performs a full mathematical analysis of the optimal dynamic withdrawal policies under the competing forces of time value of fund and penalty charge on excessive withdrawal. When proportional penalty charge is applied on any withdrawal amount, we can reduce the pricing formulation to an obstacle problem with lower and upper obstacles. We then derive the integral equations for the determination of a pair of optimal withdrawal boundaries. When proportional penalty charge is applied only on the amount that is above the contractual withdrawal rate, we manage to characterize the behavior of the optimal withdrawal boundaries that separate the domain of the pricing models into three regions: no withdrawal, continuous withdrawal at the contractual rate and immediate withdrawal of finite amount. Under certain limiting conditions, we manage to obtain analytical approximate solution to the singular stochastic control model of dynamic withdrawal.

Keywords: singular stochastic control model, guaranteed minimum withdrawal benefit, optimal withdrawal policies, penalty charge

*Correspondence author, E-mail: maykwok@ust.hk

1 Introduction

The withdrawal guarantee benefits have emerged in the last two decades as popular riders in variable annuities. In the so-called guaranteed minimum withdrawal benefit (GMWB) in annuities policies, the policyholder pays an initial lump sum to an insurance company (issuer). This initial annuitization amount is then invested in a portfolio of risky assets (typically in the form of mutual funds). The policyholder is then entitled to withdraw cash amount periodically (annually or semi-annually) and the withdrawal payments are deducted from the policyholder's fund account. In our pricing model of GMWB, continuous dynamic withdrawal is assumed (though withdrawals occur at discrete time instants in actual contracts). We treat the dynamic withdrawal rate as the control variable and formulate the model as a continuous singular stochastic control problem. Under the continuous dynamic withdrawal framework, the policyholder adopts an optimal policy to withdraw at some chosen continuous rate (the limiting cases of zero rate and infinite rate are inclusive) from his policy fund account so as to maximize the policy value. More specifically, there is a contractual withdrawal rate such that the policyholder is allowed to withdraw at or below this contractual rate with no penalty, otherwise a proportional penalty charge is applied to the withdrawal amount above the contractual amount. When the underlying asset portfolio is not producing sufficiently strong returns, it is plausible that the policyholder's fund account may become depleted before the maturity date of the policy contract. Under the provision of the GMWB, the issuer guarantees to finance the remaining withdrawal payments even the fund becomes depleted before maturity. However, if the policyholder's fund account stays positive at maturity, the policyholder is entitled to receive at maturity the remaining balance in either the fund account or guarantee account, whichever is higher, but the amount in the guarantee account is subject to proportional penalty charge.

The earliest version of the Hamilton-Jacobi-Bellman (HJB) variational inequalities formulation of the pricing model of GMWB under dynamic withdrawal policies is presented by Milevsky and Salisbury (2006). Using the penalty approximation approach, Dai *et al.* (2008) derive a more general singular stochastic control formulation of the GMWB and construct effective numerical finite difference schemes for solving the pricing models. Other enhanced versions of the singular stochastic control models and the construction of the associated numerical schemes can be found in Bauer *et al.* (2008), Chen *et al.* (2008), Chen and Forsyth (2008) and Huang *et al.* (2013). The numerical schemes based on the penalty approximation approach are seen to exhibit distinctive advantages over other numerical methods for solving the singular stochastic control GMWB models. Huang and Forsyth (2012) present the rigorous convergence proof of the penalty approximation schemes for solving the GMWB pricing models.

In these earlier papers, they have not presented the detailed characterization of the separating boundaries of various withdrawal regions and the financial interpretation of the optimal withdrawal policies. Some of the numerical papers illustrate interesting plots of the different regions of optimal withdrawal in the solution domain of the pricing model. The optimal dynamic withdrawal policies can be shown to be limited to three decision choices: zero withdrawal, withdraw at the contractual rate (the highest withdrawal rate without penalty charge) and withdraw at the infinite rate (finite amount). In this paper, we present the detailed studies on the optimal dynamic withdrawal policies and derive analytical approximate solution to the pricing model under various limiting conditions. Our analysis uses financial intuition to understand the competing forces between time value of cash, optionality provided by the guarantee and proportional penalty charge on excessive withdrawal.

This paper is organized as follows. In the next section, we briefly review the formulation of the GMWB pricing model as a continuous singular stochastic control model. Section 3 is devoted to the analysis of the pricing model under the special case where proportional penalty charge is applied on any withdrawal amount. Under this penalty charge policy, we manage to deduce the corresponding optimal dynamic withdrawal policy: no withdrawal or withdrawal of finite amount until depletion of the guarantee amount. The resulting GMWB pricing model achieves dimension reduction. As a result, it can be simplified to become an optimal stopping problem with upper and lower obstacles, similar to a real investment option model with investment and abandonment rights. To solve for the value function, it is necessary to determine a pair of time dependent withdrawal boundaries as part of the solution. We derive the corresponding integral equations for the determination of the optimal withdrawal boundaries. The GMWB value function is then expressed as a sum of European type option price and an integral that represents the withdrawal premium. We also analyze the behavior of the withdrawal boundaries at infinite time to maturity and time close to maturity. In Section 4, we present the asymptotic analysis of the separating boundaries of the pricing model under various limiting conditions subject to usual penalty charge policy (no penalty on withdrawal at or below the contractual rate). The limiting cases considered include perpetuality of policy life, infinitely large value of the policy fund value and time close to expiry. In Section 5, we present the numerical studies that were performed to verify the theoretical results on the separating boundaries with respect to different optimal dynamic withdrawal policies and asymptotic price formulas for the GMWB value function under various limiting conditions. Financial interpretation of the behavior of the separating boundaries are presented. We also show the evolution of the withdrawal regions with respect to varying calendar times. Conclusive remarks are presented in the last section.

2 Linear complementarity formulation of continuous dynamic withdrawal model

Let S_t denote the value of the reference portfolio of assets underlying the variable annuity policy before the deduction of any proportional fees charged on the policy fund account by the issuer for the provision of the guaranteed withdrawal benefit. Taking the usual geometric Brownian distribution assumption on the price dynamics of equity, the dynamics of S_t under the risk neutral measure Q is governed by

$$dS_t = rS_t\ dt + \sigma S_t\ dB_t,$$

where B_t represents the standard Brownian motion under Q, σ is the volatility and r is the riskfree interest rate. Let $\mathcal{F}_t$ be the natural filtration generated by the Brownian motion B_t.

We let A_t denote the account balance of the guarantee, where A_t is non-negative and non-increasing $\{\mathcal{F}_t\}_{t\geq 0}$-adaptive left continuous process. At initiation, A_0 equals the initial policy premium paid upfront. The withdrawal guarantee becomes insignificant when A_t hits 0. As the withdrawal process continues, A_t decreases over the life of the policy until it hits the zero value. To derive the continuous time pricing model, we first consider a restricted class of withdrawal policies in which A_t is constrained to be absolutely continuous with bounded derivatives, where

$$A_t = A_0 - \int_0^t \gamma_u\ du,\ 0 \leq \gamma_u \leq \lambda. \tag{2.1}$$

Here, we follow the penalty approximation approach by assuming the continuous withdrawal rate γ to have the upper bound λ. Later, we take the limit $\lambda \to \infty$ to obtain the singular stochastic control formulation (Dai *et al.*, 2008). The limiting case of infinite withdrawal rate corresponds to immediate withdrawal of finite amount. The maximum finite amount of withdrawal prior to maturity is capped by the outstanding balance in the guarantee account.

We let W_t denote the policy fund value and η be the proportional fee charged on the policy fund value paid by the policyholder for the provision of the guaranteed withdrawal benefit. The dynamics of W_t then follows

$$dW_t = (r-\eta)W_t\ dt + \sigma W_t\ dB_t + dA_t, \tag{2.2}$$

for $W_t > 0$ and $\eta > 0$. Once W_t hits the value 0, it stays at this value thereafter. That is, $W = 0$ is an absorbing barrier. Let w_0 be the initial account value of the policy, which is simply equal to the policy premium paid upfront. We then have $W_0 = A_0 = w_0$. At maturity of the policy, the policyholder is entitled to receive the larger amount chosen between the remaining balance of the policy fund account W_T if $W_T > 0$ and the remaining balance of the guarantee account A_T subject to proportional penalty charge.

Let $f(\gamma)$ denote the rate of cash flow received by the policyholder from the continuous withdrawal process. Let k denote the proportional penalty charge rate applied to excessive withdrawal beyond the contractual withdrawal rate G. We then have

$$f(\gamma) = \begin{cases} \gamma & \text{if } 0 \le \gamma \le G \\ G + (1-k)(\gamma - G) & \text{if } \gamma > G \end{cases}. \tag{2.3}$$

When proportional penalty charge is applied on any withdrawal amount, we then have $f(\gamma) = (1-k)\gamma$. This is equivalent to set $G = 0$.

The no-arbitrage value $\overline{V}$ of the policy with upper cap λ on γ is given by

$$\overline{V}(W, A, t) = \sup_{\gamma} E_t \left[e^{-r(T-t)} \max(W_T, (1-k)A_T) + \int_t^T e^{-r(u-t)} f(\gamma_u)\ du \right], \tag{2.4}$$

where T is the maturity date of the policy and expectation E_t is taken under the risk neutral measure Q conditional on $W_t = W$ and $A_t = A$. Here, γ is the stochastic control variable that is chosen to maximize the expected value of discounted cash flows. Using the standard procedure of deriving the Hamilton-Jacobi-Bellman (HJB) equation in stochastic control problems (Yong and Zhou, 1999), the governing equation for $\overline{V}$ is found to be

$$\frac{\partial \overline{V}}{\partial t} + L\overline{V} + \max_{\gamma} h(\gamma) = 0$$

where

$$L\overline{V} = \frac{\sigma^2}{2} W^2 \frac{\partial^2 \overline{V}}{\partial W^2} + (r-\eta) W \frac{\partial \overline{V}}{\partial W} - r\overline{V}$$

and

$$\begin{aligned} h(\gamma) &= f(\gamma) - \gamma \frac{\partial \overline{V}}{\partial W} - \gamma \frac{\partial \overline{V}}{\partial A} \\ &= \begin{cases} (1 - \frac{\partial \overline{V}}{\partial W} - \frac{\partial \overline{V}}{\partial A})\gamma & \text{if } 0 \le \gamma < G \\ kG + (1 - k - \frac{\partial \overline{V}}{\partial W} - \frac{\partial \overline{V}}{\partial A})\gamma & \text{if } \gamma \ge G \end{cases}, \end{aligned}$$

with terminal payoff: $\overline{V}(W, A, T) = \max(W_T, (1-k)A_T)$. By taking the limit $\lambda \to \infty$, Dai *et al.* (2008) obtain the following linear complementarity formulation for the GMWB value function $V(W, A, t)$:

$$\min\left[-\frac{\partial V}{\partial t} - LV - G\max\left(1 - \frac{\partial V}{\partial W} - \frac{\partial V}{\partial A}, 0\right), \frac{\partial V}{\partial W} + \frac{\partial V}{\partial A} - (1-k) \right] = 0,$$
$$W > 0, 0 < A < w_0, 0 < t < T. \tag{2.5}$$

The discussion of the boundary conditions at $A = 0$, $W = 0$ and $W \to \infty$ in the pricing model is relegated to Section 4.

3 Simplified pricing model under penalty charge on any withdrawal

In this section, we limit our discussion to the case where proportional penalty charge is applied on any withdrawal amount (equivalent to set $G = 0$). There are two reasons for analyzing the simplified pricing model under this penalty charge policy. Firstly, an analytic representation of the solution to the GMWB value function (up to an integral representation of the withdrawal premium in terms of the withdrawal boundaries) can be obtained under $G = 0$. Secondly, the asymptotic behavior of optimal dynamic withdrawal policies under the usual penalty policy with $G > 0$ can be inferred from those under the penalty charge policy with $G = 0$. This is similar to the studies of optimal consumption and portfolio investment under zero transaction costs (Merton, 1971) and finite transaction costs (Davis and Norman, 1990), where the analysis of the zero transaction costs model provides insight for the solution of the finite transaction costs model.

When $G = 0$, the function $h(\gamma)$ reduces to

$$h(\gamma) = \left(1 - k - \frac{\partial \overline{V}}{\partial W} - \frac{\partial \overline{V}}{\partial A}\right)\gamma, \quad \gamma \geq 0.$$

It is seen that the maximum value of $h(\gamma)$ is achieved at $\gamma = 0$ or $\gamma = \lambda$, depending on the sign of $1-k-\frac{\partial \overline{V}}{\partial W}-\frac{\partial \overline{V}}{\partial A}$. Taking $G = 0$ in eq. (2.5), we obtain the following reduced form of the linear complementarity formulation of the value function $V(W, A, t)$ under the continuous stochastic control framework (Dai *et al.*, 2008):

$$\min\left(-\frac{\partial V}{\partial t} - LV, \frac{\partial V}{\partial W} + \frac{\partial V}{\partial A} - (1-k)\right) = 0,\ W > 0,\ 0 < A < w_0,\ 0 < t < T, \tag{3.1}$$

with terminal condition: $V(W, A, T) = \max(W, (1-k)A)$. To complete the model formulation, it is necessary to prescribe the full set of boundary conditions. When $A = 0$, the policy contract reduces to the usual asset portfolio; so the value function becomes $V(W, 0, t) = e^{-\eta(T-t)}W$. Here, the exponential time decay factor $e^{-\eta(T-t)}$ arises due to the proportional fee paid at the rate η throughout the remaining life of the contract. For the far field boundary condition, the value function becomes linear in W as $W \to \infty$. We then have $\lim_{W\to\infty} \frac{\partial V}{\partial W}(W, A, t) = e^{-\eta(T-t)}$. When $W = 0$, since proportional penalty charge is applied on any withdrawal, it is optimal for the policyholder to withdraw the remaining guarantee account immediately; so $V(0, A, t) = (1-k)A$.

There are two regions in the solution domain of the pricing model that lies in the first quadrant of the W-A plane, namely, $\mathcal{D}_\infty$ where $\gamma = \infty$ (withdrawal of finite amount) and $\mathcal{D}_0$ where $\gamma = 0$ (zero withdrawal). Intuitively, we should withdraw finite amount when either A is sufficiently above W or W achieves value that is sufficiently above A; and no withdrawal when the values of A and W are relatively close to each other. The later case of zero withdrawal arises from the optimal decision to capture the option value of waiting. The value function V inside $\mathcal{D}_0$ satisfies

$$\frac{\partial V}{\partial t} + LV = 0 \quad \text{and} \quad \frac{\partial V}{\partial W} + \frac{\partial V}{\partial A} > 1 - k. \tag{3.2a}$$

On the other hand, the value function V inside $\mathcal{D}_\infty$ satisfies

$$\frac{\partial V}{\partial t} + LV < 0 \quad \text{and} \quad \frac{\partial V}{\partial W} + \frac{\partial V}{\partial A} = 1 - k. \tag{3.2b}$$

There are two separate withdrawal regions D_∞, each on either side of the line: $A = W$. It is relatively straightforward to deduce the solution to the above linear complementarity formulation (3.2b) in the withdrawal region $\mathcal{D}_\infty$. We consider the two separate cases:

(i) The solution in $\mathcal{D}_\infty$ that lies on the side $A > W$ is given by

$$V(W, A, t) = (1 - k)A, \qquad A > W. \tag{3.3a}$$

The resulting value function indicates an immediate withdrawal of the full guarantee amount A subject to proportional penalty charge k. It is seen that the above solution satisfies eq. (3.2b) together with the terminal condition: $V(W, A, t) = \max((1 - k)A, W) = (1 - k)A$ and boundary condition: $V(0, A, t) = (1 - k)A$, $A > 0$.

(ii) The solution in $\mathcal{D}_\infty$ that lies on the side $A < W$ is given by

$$V(W, A, t) = (1 - k)A + e^{-\eta(T-t)}(W - A), \qquad A < W. \tag{3.3b}$$

The value function reveals a similar optimal policy of immediate withdrawal of the full guarantee amount A and the residual policy fund account after withdrawal of finite amount A has an expected value equals $e^{-\eta(T-t)}(W - A)$. Again, the solution satisfies eq. (3.2b), the terminal condition: $V(W, A, T) = \max((1 - k)A, W) = W$, and other boundary conditions.

Homogeneity property of the value function

With $G = 0$, the value function $V(W, A, t)$ becomes homogeneous in A and W. The dimension of the pricing model under this penalty charge policy can be reduced to one by normalizing $V(W, A, t)$ by A and defining the similarity variable $Y = W/A$. Let $P(Y, t) = V(W, A, t)/A$, the linear complementarity formulation (3.1) can be expressed in terms of $P(Y, t)$ as

$$\min\left(-\frac{\partial P}{\partial t} - \frac{\sigma^2}{2}Y^2\frac{\partial^2 P}{\partial Y^2} - (r - \eta)Y\frac{\partial P}{\partial Y} + rP, (1 - Y)\frac{\partial P}{\partial Y} + P - (1 - k)\right) = 0, \tag{3.4}$$

together with terminal condition: $P(Y, T) = \max(Y, 1 - k)$ and boundary conditions: (i)$\frac{\partial P}{\partial Y}(\infty, t) = e^{-\eta(T-t)}$, (ii)$P(0, t) = 1 - k$.

As a remark, the special case of zero penalty charge $k = 0$ also achieves the same simplified form of homogeneity in the pricing formulation. This is revealed by setting the parameter k to be zero in the above linear complementarity formulation. This result is easily observed since G disappears in the pricing formulation (2.5) when k becomes zero [note that k and G appear together in the product kG in $h(\gamma)$]. As another remark, we would like to quote a mathematical property of $P(Y, t)$ that is frequently used in later proofs of optimal withdrawal policies: $P(Y, t)$ is a convex function in Y, where $\frac{\partial^2 P}{\partial Y^2} > 0$ [Flemming and Soner (1993), Lemma VIII. 3.2].

Optimal dynamic withdrawal policies

Similar to an optimal stopping problem with two-sided free boundaries, the regions of D_∞ and D_0 are separated by $Y^*_{\text{low}}(t)$ and $Y^*_{\text{up}}(t)$, where $Y^*_{\text{low}}(t)$ and $Y^*_{\text{up}}(t)$ are the lower and upper withdrawal boundary, respectively. When the value function V and policy fund value W are normalized by A, the two solutions in $\mathcal{D}_\infty$ as given by eqs. (3.3a,b) can be expressed as $P(Y, t) = 1 - k$ when $Y < Y^*_{\text{low}}(t)$; and $P(Y, t) = 1 - k + e^{-\eta(T-t)}(Y - 1)$ when $Y > Y^*_{\text{up}}(t)$. Here, $Y^*_{\text{low}}(t)$ and $Y^*_{\text{up}}(t)$ are the time dependent threshold values to be determined in a separate solution procedure that solves the value function in the continuation (no withdrawal) region $\mathcal{D}_0$ (to be discussed later).

The above results can be summarized by the various regions in the domain of the pricing model in the W-A plane as depicted in Figure 1. By virtue of homogeneity of the value function, the separating boundaries are a pair of straight lines. The optimal choice of an immediate depletion of the guarantee amount occurs when (W, A) falls within either one of the "$\gamma = \infty$" regions. This phenomenon is illustrated by the two paths (shown in arrows) in Figure 1. The slope of the upper left (lower right) withdrawal boundary has slope $\frac{1}{Y^*_{\text{low}}(t)}$ $\left(\frac{1}{Y^*_{\text{up}}(t)}\right)$ that is greater (smaller) than one. For any point (W, A) that falls within either one of the infinite withdrawal regions, the withdrawal of a finite amount δ (less than or equal to A) moves the point (W, A) to the new point $((W-\delta)^+, A-\delta)$, where x^+ denotes $\max(x, 0)$. Interestingly, $((W-\delta)^+, A-\delta)$ remains in the same withdrawal region (see Figure 1). Therefore, the policyholder should continue to withdraw until the complete depletion of the guarantee account is resulted. These observations are consistent with the analytic solutions presented in eqs. (3.3a,b).

By using convexity property of $P(Y, t)$, one can show mathematically that once it is optimal to withdraw under $G = 0$, then the whole guarantee account will be withdrawn to complete depletion. From the linear complementarity formulation (3.4), we deduce that it is optimal to withdraw finite amount if and only if

$$H(Y, t) = (Y-1)\frac{\partial P(Y, t)}{\partial Y} - P(Y, t) + (1-k) = 0. \tag{3.5}$$

Let the finite withdrawal amount be δ_0, some positive value that is less than A. After the withdrawal of δ_0, Y becomes $\widetilde{Y} = \frac{W - \delta_0}{A - \delta_0}$. To prove the property of complete depletion once withdrawal of finite amount is initiated, it suffices to show that

$$H(\widetilde{Y}, t) = (\widetilde{Y}-1)\frac{\partial P(\widetilde{Y}, t)}{\partial Y} - P(\widetilde{Y}, t) + (1-k) = 0.$$

In other words, it remains to be optimal to continue to withdraw finite amount instantaneously. We consider

$$\frac{\partial H(Y, t)}{\partial Y} = (Y-1)\frac{\partial^2 P(Y, t)}{\partial Y^2} + \frac{\partial P(Y, t)}{\partial Y} - \frac{\partial P(Y, t)}{\partial Y} = (Y-1)\frac{\partial^2 P(Y, t)}{\partial Y^2}.$$

By virtue of convexity of $P(Y, t)$, where $\frac{\partial^2 P}{\partial Y^2} > 0$ for all values of Y, $H(Y, t)$ is increasing (decreasing) when $Y > 1$ $(Y < 1)$. For $Y \leq 1$, we can deduce that $\widetilde{Y} \leq Y \leq 1$, so $H(\widetilde{Y}, t) \geq H(Y, t) = 0$. Similarly, for $Y > 1$, we have $\widetilde{Y} > Y > 1$, and $H(\widetilde{Y}, t) \geq 0$ is again resulted. On the other hand, the linear complementarity formulation implies $H(Y, t) \leq 0$ for any value of Y. Combining the results, we obtain the desired result: $H(\widetilde{Y}, t) = 0$.

Determination of $P(Y, t)$ in the continuation region

The next procedure is to solve for the value function $P(Y, t)$ in the continuation (no withdrawal) region $\mathcal{D}_0$, where $P(Y, t)$ is governed by

$$\frac{\partial P}{\partial t} + \frac{\sigma^2}{2}Y^2\frac{\partial^2 P}{\partial Y^2} + (r-\eta)Y\frac{\partial P}{\partial Y} - rP = 0, \quad Y^*_{\text{low}}(t) < Y < Y^*_{\text{up}}(t), \quad 0 < t < T. \tag{3.6}$$

In order to complete the formulation of the obstacle problem, it is necessary to prescribe the value matching and smooth pasting conditions at the two ends $Y^*_{\text{low}}(t)$ and $Y^*_{\text{up}}(t)$. We have

(i) value matching conditions:

$$P\left(Y^*_{\text{low}}(t), t\right) = 1-k \quad \text{and} \quad P\left(Y^*_{\text{up}}(t), t\right) = 1-k+e^{-\eta(T-t)}\left[Y^*_{\text{up}}(t)-1\right]. \tag{3.7a}$$

(ii) smooth pasting conditions:

$$\frac{\partial P}{\partial Y}\left(Y^*_{\text{low}}(t),t\right)=0 \quad \text{and} \quad \frac{\partial P}{\partial Y}\left(Y^*_{\text{up}}(t),t\right)=e^{-\eta(T-t)}. \tag{3.7b}$$

The corresponding obstacle constraint is given by

$$P(Y,t)\geq 1-k+\max\left(e^{-\eta(T-t)}(Y-1),0\right), \qquad t<T. \tag{3.7c}$$

As a standard procedure for solving an obstacle problem, the two time dependent thresholds $Y^*_{\text{low}}(t)$ and $Y^*_{\text{up}}(t)$ have to be determined by solving a pair of integral equations as part of the solution procedure.

In Figure 2, we show the schematic plot of the value function $P(Y,t)$ against Y and the time dependent obstacle function as specified in eq. (3.7c). The value function $P(Y,t)$ stays above the obstacle function $1-k+\max\left(e^{-\eta(T-t)}(Y-1),0\right)$ in the continuation region: $Y^*_{\text{low}}(t)<Y<Y^*_{\text{up}}(t)$ and $P(Y,t)$ equals the obstacle function in the withdrawal regions: $Y\leq Y^*_{\text{low}}(t)$ and $Y\geq Y^*_{\text{up}}(t)$. For the perpetual case where $T\to\infty$, the obstacle function becomes the constant value of $1-k$. One then deduces that the continuation region vanishes under perpetuality when $G=0$. This is consistent with the result: $\lim_{\tau\to\infty} Y^*_{\text{low}}(\tau)=\lim_{\tau\to\infty} Y^*_{\text{up}}(\tau)=1$ (see a more detailed discussion in Section 4.1).

As a remark, the challenge in solving the linear complementarity formulation (3.1) arises from the constraint condition that is expressed in terms of the gradient of the value function. However, once we have obtained the solution to the value function $P(Y,t)$ in the withdrawal regions, we can rewrite the pricing formulation as an obstacle problem as shown in eq. (3.6) and eqs. (3.7a,b,c). With the explicit form of the obstacle function known [see eq. (3.7c)], the pricing model resembles that of a real investment option model with investment and abandonment rights. Since the terminal payoff is given by $\max(Y,1-k)=1-k+\max(Y-(1-k),0)$, one can express the value function in the form of a discount bond of par value $1-k$ plus a European call price function with strike $1-k$, together with an integral that represents the withdrawal premium (Detemple, 2005; Kwok, 2008). Here, the proportional fee η plays the same role as dividend yield q and the payoff at $Y\geq Y^*_{\text{up}}(t)$ and $Y\leq Y^*_{\text{low}}(t)$ resemble the respective exercise payoff of the real investment option with investment and abandonment rights. The value function can be expressed as

$$P(Y,t)=(1-k)e^{-r(T-t)}+c(Y,t;1-k)+M(Y,t), \tag{3.8}$$

where $M(Y,t)$ represents the withdrawal premium and $c(Y,t;1-k)$ is the time-t price of the European call option with strike $1-k$. Before we derive the analytical integral representation of the withdrawal premium, it is necessary to find the corresponding time intervals where the optimal withdrawal boundaries $Y^*_{\text{up}}(t)$ and $Y^*_{\text{low}}(t)$ are defined.

Recall that $H(Y,t)$ is increasing with respect to Y for $Y>1$; and $H(1,t)=1-k-P(1,t)<0$ since $P(Y,t)$ is above the intrinsic value $1-k$ at $Y=1$ (see Figure 2). Also, we can deduce that

$$\begin{aligned}\lim_{Y\to\infty} H(Y,t) &= 1-\lim_{Y\to\infty}\frac{\partial P}{\partial Y}(Y,t)-\lim_{Y\to\infty}P(Y,t)+\lim_{Y\to\infty}Y\frac{\partial P}{\partial Y}(Y,t)-k\\ &= 1-e^{-\eta(T-t)}-Ye^{-\eta(T-t)}+Ye^{-\eta(T-t)}-k\\ &= 1-k-e^{-\eta(T-t)}.\end{aligned}$$

Putting all the results together, we deduce that

(i) when $1-k-e^{-\eta(T-t)}\leq 0$, we have $H(Y,t)<0$ for $Y>1$;

(ii) when $1-k-e^{-\eta(T-t)}>0$, $H(Y,t)$ as a function of Y has a unique root that lies in $(1,\infty)$.

In other words, for $Y>1$, it is never optimal to withdraw finite amount immediately if $1-k-e^{-\eta(T-t)}\leq 0$ since $H(Y,t)$ stays negative. Equivalently, $Y^*_{\text{up}}(t)$ is not defined within the time interval $(T+\frac{\ln(1-k)}{\eta},T)$; that is, when the time to expiry is less than $-\frac{\ln(1-k)}{\eta}$.

By following a similar argument, we can show that $Y^*_{\text{low}}(t)$ is defined for all times, $t<T$.

Withdrawal premium

Let $\tau^*=-\frac{\ln(1-k)}{\eta}$ and recall that $Y^*_{\text{up}}(t)$ is not defined for $t\geq T-\tau^*$. Once we know the characterization of $Y^*_{\text{up}}(t)$ and $Y^*_{\text{low}}(t)$, by following a similar formulation of the delayed exercise premium in an American option (Detemple, 2005), the withdrawal premium is given by (see Appendix A)

$$\begin{aligned} M(Y,t) &= (1-k)r\int_t^{T-\widehat{\tau}^*} e^{-r(u-t)}N(d_{12}(Y,u-t;Y^*_{\text{up}}(u)))\,du \\ &\quad -(r-\eta)\int_t^{T-\widehat{\tau}^*} e^{-r(u-t)}e^{-\eta(T-u)}N(d_{12}(Y,u-t;Y^*_{\text{up}}(u)))\,du \\ &\quad +(1-k)r\int_t^{T} e^{-r(u-t)}N(-d_{22}(Y,u-t;Y^*_{\text{low}}(u)))\,du, \end{aligned} \tag{3.9}$$

where

$$\widehat{\tau}^* = \min\left(T-t,\tau^*\right),$$

$$d_{12}\left(Y,u-t;Y^*_{\text{up}}(u)\right) = \frac{\ln\frac{Y}{Y^*_{\text{up}}(u)}+\left(r-\eta-\frac{\sigma^2}{2}\right)(u-t)}{\sigma\sqrt{u-t}},$$

$$d_{22}\left(Y,u-t;Y^*_{\text{low}}(u)\right) = \frac{\ln\frac{Y}{Y^*_{\text{low}}(u)}+\left(r-\eta-\frac{\sigma^2}{2}\right)(u-t)}{\sigma\sqrt{u-t}}.$$

Integral equations for the determination of the optimal withdrawal boundaries

The withdrawal boundaries $Y^*_{\text{up}}(t)$ and $Y^*_{\text{low}}(t)$ can be determined by applying the value matching conditions (3.7a), which lead to the following pair of integral equations for solving the withdrawal boundaries $Y^*_{\text{up}}(t)$ and $Y^*_{\text{low}}(t)$. Recall that $Y^*_{\text{up}}(t)$ is defined for $t<T-\tau^*$ while $Y^*_{\text{low}}(t)$ is defined for $t<T$. For $T-t\leq\tau^*$, the integral equation for $Y^*_{\text{low}}(t)$ is given by

$$\begin{aligned} 1-k &= (1-k)e^{-r(T-t)}+c\left(Y^*_{\text{low}}(t),t;1-k\right) \\ &\quad +(1-k)r\int_t^{T} e^{-r(u-t)}N\left(-d_{22}\left(Y^*_{\text{low}}(t),u-t;Y^*_{\text{low}}(u)\right)\right)\,du. \end{aligned} \tag{3.10}$$

For $T-t>\tau^*$, the pair of integral equations for $Y^*_{\text{up}}(t)$ and $Y^*_{\text{low}}(t)$ are given by

$$\begin{aligned} 1-k &= (1-k)e^{-r(T-t)}+c\left(Y^*_{\text{low}}(t),t;1-k\right) \\ &\quad +(1-k)r\int_t^{T-\tau^*} e^{-r(u-t)}N\left(d_{12}\left(Y^*_{\text{low}}(t),u-t;Y^*_{\text{up}}(u)\right)\right)\,du \\ &\quad -(r-\eta)\int_t^{T-\tau^*} e^{-r(u-t)}e^{-\eta(T-u)}N\left(d_{12}\left(Y^*_{\text{low}}(t),u-t;Y^*_{\text{up}}(u)\right)\right)\,du \\ &\quad +(1-k)r\int_t^{T} e^{-r(u-t)}N\left(-d_{22}\left(Y^*_{\text{low}}(t),u-t;Y^*_{\text{low}}(u)\right)\right)\,du; \end{aligned} \tag{3.11a}$$

and

$$
\begin{aligned}
1-k+e^{-\eta(T-t)}\left[Y_{\text{up}}^{*}(t)-1\right]=&(1-k)e^{-r(T-t)}+c\left(Y_{\text{up}}^{*}(t),t;1-k\right)\\
&+(1-k)r\int_{t}^{T-\tau^{*}}e^{-r(u-t)}N\left(d_{12}\left(Y_{\text{up}}^{*}(t),u-t;Y_{\text{up}}^{*}(u)\right)\right)\,du\\
&-(r-\eta)\int_{t}^{T-\tau^{*}}e^{-r(u-t)}e^{-\eta(T-u)}N\left(d_{12}\left(Y_{\text{up}}^{*}(t),u-t;Y_{\text{up}}^{*}(u)\right)\right)\,du\\
&+(1-k)r\int_{t}^{T}e^{-r(u-t)}N\left(-d_{22}\left(Y_{\text{up}}^{*}(t),u-t;Y_{\text{low}}^{*}(u)\right)\right)\,du. \qquad (3.11b)
\end{aligned}
$$

The numerical solution of the above pair of integral equations can be performed easily using the renowned recursive integration method (Kwok, 2008). More discussion on the behavior of $Y_{\text{up}}^{*}(t)$ and $Y_{\text{low}}^{*}(t)$ at time close to expiry and under perpetuality can be found in Section 4. Numerical studies on the various properties of $Y_{\text{up}}^{*}(t)$ and $Y_{\text{low}}^{*}(t)$ are presented in Section 5.

4 Optimal dynamic withdrawal policies under various limiting conditions

Under the usual proportional penalty charge policy with $G>0$, the linear complementarity formulation for the value function $V(W,A,t)$ becomes

$$
\min\left[-\frac{\partial V}{\partial t}-LV-G\max\left(1-\frac{\partial V}{\partial W}-\frac{\partial V}{\partial A},0\right),\frac{\partial V}{\partial W}+\frac{\partial V}{\partial A}-(1-k)\right]=0,
$$
$$
W>0, 0<A<w_0, 0<t<T, \qquad (4.1)
$$

with terminal payoff: $V(W,A,T)=\max(W,(1-k)A)$ and boundary conditions: $V(W,0,t)=e^{-\eta(T-t)}W$ and $\frac{\partial V}{\partial W}(W,A,t)=e^{-\eta(T-t)}$ as $W\to\infty$. Dai *et al.* (2008) have shown that the boundary condition at $W=0$ is given by

$$
V(0,A,t)=(1-k)\max\left(A-G\overline{\tau}^{*},0\right)+\frac{G}{r}[1-e^{-r\min\left(\frac{A}{G},\overline{\tau}^{*}\right)}], \qquad (4.2)
$$

where

$$
\overline{\tau}^{*}=\min\left(-\frac{\ln(1-k)}{r},T-t\right).
$$

In Section 4.2, we will present more detailed discussion of the far field boundary condition as $W\to\infty$.

It will be shown later in this section that dimension reduction of the pricing formulation can be achieved under various limiting conditions: (i) perpetuality of the policy life, (ii) infinitely large value of the policy fund value and (iii) time close to expiry. Thanks to the use of dimension reduction under these limiting cases, it is possible to derive the corresponding analytical representation of the value function corresponding to these cases.

4.1 Perpetuality of policy life

Similar to the analysis of the optimal stopping rules in American options, it is relatively straightforward to analyze the optimal withdrawal policies under perpetuality, $T\to\infty$. Indeed, we manage to obtain closed form solution to the linear complementarity formulation when the value function has no dependence on time. The nice analytical tractability stems from the property that the number of state variables in the linear complementarity formulation is reduced by one under perpetuality.

Let $V_\infty(W,A)$ denote the value function of the perpetual policy, where $T \to \infty$. With absence of time dependency, the corresponding linear complementarity formulation reduces to

$$\min\left[-LV_\infty - G\max\left(1-\frac{\partial V_\infty}{\partial W}-\frac{\partial V_\infty}{\partial A},0\right), \frac{\partial V_\infty}{\partial W}+\frac{\partial V_\infty}{\partial A}-(1-k)\right]=0,$$
$$W>0, 0<A<w_0, 0<t<T, \tag{4.3}$$

with boundary conditions: $V_\infty(W,0)=0$ and

$$V_\infty(0,A)=(1-k)\max\left(A-A^*,0\right)+\frac{G}{r}\left[1-e^{-\frac{r}{G}\min(A,A^*)}\right], \tag{4.4}$$

where $A^*=-\frac{G}{r}\ln(1-k)$. For notational convenience, we write $\widehat{V}_\infty(A)=V_\infty(0,A)$ as defined in eq. (4.4). One can check that $\widehat{V}_\infty(A)$ satisfies

$$\begin{cases} r\widehat{V}_\infty+\left(\frac{d\widehat{V}_\infty}{dA}-1\right)=0 & \text{when } 0<1-\frac{d\widehat{V}_\infty}{dA}<k \\ r\widehat{V}_\infty-kG>0 & \text{when } \frac{d\widehat{V}_\infty}{dA}=1-k \end{cases}. \tag{4.5}$$

Interestingly, with no dependence on W in $\widehat{V}_\infty(A)$, it is observed that $\widehat{V}_\infty(A)$ also satisfies the linear complementarity formulation (4.3). We then deduce that the solution to the value function of the perpetual policy is given by

$$V_\infty(W,A)=\widehat{V}_\infty(A)=\begin{cases} (1-k)\left(A-A^*\right)+\frac{G}{r}k, & \text{when } A>A^* \\ \frac{G}{r}\left(1-e^{-r\frac{A}{G}}\right), & \text{when } A\le A^* \end{cases}. \tag{4.6}$$

Since the policy fund value W is received at maturity, so its time value is zero under perpetuality. Therefore, it is not surprising that the value function under perpetuality is independent of W and the boundary value function $\widehat{V}_\infty(A)$ is the solution to the linear complementarity formulation. The mathematical justification and financial intuition behind the above result are presented in Appendix B.

In Figure 3, we illustrate the separation of the domain $\{(W,A): W\ge 0 \,\text{and}\, 0\le A\le w_0\}$ into the infinite withdrawal region and the region of withdrawal at the contractual rate G under perpetuality. When $A>A^*$, it is optimal to withdraw the amount $A-A^*$ immediately, then followed by withdrawal at the rate $\gamma=G$.

Optimal withdrawal policies under perpetuality and $G=0$

For the degenerate case where $G=0$, we have $A^*=0$ so that the whole solution domain corresponds to the infinite withdrawal region ($\gamma=\infty$). That is, $D_0=\emptyset$ under $T\to\infty$ and $G=0$. Referring to the free boundaries defined in Section 3, the continuation region lies inside $\left(Y^*_{\text{low}}(t), Y^*_{\text{up}}(t)\right)$, where $Y^*_{\text{up}}(t)\ge 1$ and $Y^*_{\text{low}}(t)\le 1$ for all t. Since the continuation region vanishes under perpetuality, we deduce that $Y^*_{\text{up}}(t)$ and $Y^*_{\text{low}}(t)$ tend to one from above and below, respectively, when time to expiry is infinite.

4.2 Far field boundary condition at infinitely large policy fund value

Though the far field boundary condition in differential form: $\lim_{W\to\infty}\frac{\partial V}{\partial W}(W,A,t)=e^{-\eta(T-t)}$ is essentially correct, we would like to derive the far field boundary condition of the value function $V(W,A,t)$ at $W\to\infty$. We would like to derive the more precise analytical representation of the far field boundary condition by identifying the optimal withdrawal policies through a combination of intuition and analytical analysis, similar to the approach presented in Appendix B. The asymptotic analytic formula of the value function at $W\to\infty$ provides valuable insight in the characterization of the horizontal asymptote of the separating withdrawal boundary at $W \to \infty$ (see later discussion).

Writing in full, the linear complementarity formulation (4.1) can be expressed as follows:

(i) When $\frac{\partial V}{\partial W} + \frac{\partial V}{\partial A} > 1$, which corresponds to zero withdrawal, $V(W, A, t)$ in the continuation region is governed by

$$-\frac{\partial V}{\partial t} - (r - \eta)W\frac{\partial V}{\partial W} - \frac{\sigma^2}{2}W^2\frac{\partial^2 V}{\partial W^2} + rV = 0. \tag{4.7a}$$

(ii) When $1 \geq \frac{\partial V}{\partial W} + \frac{\partial V}{\partial A} > 1 - k$, which corresponds to continuous withdrawal at the rate G, $V(W, A, t)$ is governed by

$$-\frac{\partial V}{\partial t} - (r - \eta)W\frac{\partial V}{\partial W} - \frac{\sigma^2}{2}W^2\frac{\partial^2 V}{\partial W^2} + rV - G\left(1 - \frac{\partial V}{\partial W} - \frac{\partial V}{\partial A}\right) = 0. \tag{4.7b}$$

(iii) In the region that corresponds to withdrawal at the infinite rate (withdrawal of finite amount), $V(W, A, t)$ observes

$$\frac{\partial V}{\partial W} + \frac{\partial V}{\partial A} = 1 - k. \tag{4.7c}$$

Optimal withdrawal policies

Note that when $W \gg A$, the terminal payoff $V(W, A, T)$ is almost surely to be W. The value of optionality in the terminal payoff becomes vanishingly small. In the far field, the solution in the region of zero withdrawal would be $V = e^{-\eta(T-t)}W$ so that it satisfies eq. (4.7a) and observes the terminal condition. However, such solution cannot exist since it violates the gradient constraint: $\frac{\partial V}{\partial W} + \frac{\partial V}{\partial A} > 1$. We then deduce that zero withdrawal is ruled out in the far field, $W \to \infty$. This is consistent with the financial intuition that it is non-optimal not to withdraw if there is no optionality in the terminal payoff. The remaining choices in the withdrawal policies are continuous withdrawal at the rate G and immediate withdrawal of finite amount.

An alternative argument that justifies the non-optimal choice of zero withdrawal ($\gamma = 0$) at $W \to \infty$ can be revealed from Figure 1 that shows the configuration of the continuation region ($\gamma = 0$) under $G = 0$. Firstly, we observe that the continuation region shrinks with an increasing value of G since the policyholder has a wider strategic choice of continuous withdrawal under a larger value of G. On the other hand, when $G = 0$, the continuation region can only extend to a finite value of W. Since the continuation region shrinks with $G > 0$, the optimal policy of zero withdrawal should be ruled out at some sufficiently large value of W.

We observe that the two separate limiting cases of perpetuality and infinitely large policy fund value (far field) are similar, where the value of optionality in the terminal payoff is zero and the value function is dependent on single state variable. More precisely, the value function under perpetuality has no dependence on W. In the far field case, the value function is a function of W while A only serves as a parameter. In addition, the guarantee account is depreciated at the riskfree interest rate r under perpetuality while the policy fund value is depreciated at the proportional fee η in the far field case. Therefore, we expect to have a similar set of optimal withdrawal policies in both limiting cases: either (i) continuous withdrawal at the contractual rate G until the earlier time chosen among maturity date and the date of complete depletion of the guarantee account or (ii) immediate withdrawal of a finite amount followed by continuous withdrawal at the rate G.

Analytical representation of the far field boundary condition

The analytical representation of the far field boundary condition takes different forms, depending on the relative magnitude of the proportional penalty charge upon excessive withdrawal and proportional fees paid within the remaining life of the policy, and also the level of account balance of the guarantee. Let A^{**} denote the unique root to the following equation

$$1-k-e^{-\eta(T-t)}-e^{-r\frac{A}{G}}\left\{1-e^{-\eta\left[(T-t)-\frac{A}{G}\right]}\right\}=0. \tag{4.8}$$

At $W\to\infty$, the asymptotic solution to the value function $V(W,A,t)$ is given by (see Appendix C for the detailed derivation):

(i) $e^{-\eta(T-t)}<1-k$ and $A>A^{**}$

$$V(W,A,t)\approx e^{-\eta(T-t)}W+\left[1-k-e^{-\eta(T-t)}\right](A-A^{**})+\frac{G}{r}\left(1-e^{-r\frac{A^{**}}{G}}\right) \\ -\frac{Ge^{-\eta(T-t)}}{r-\eta}\left[1-e^{-(r-\eta)\frac{A^{**}}{G}}\right],\quad W\to\infty. \tag{4.9a}$$

The optimal withdrawal policy is to withdraw the finite amount $A-A^{**}$ immediately, then followed by continuous withdrawal at the rate G.

(ii) $e^{-\eta(T-t)}<1-k$ and $A\le A^{**}$

$$V(W,A,t)\approx e^{-\eta(T-t)}W+\frac{G}{r}\left(1-e^{-r\frac{A}{G}}\right)-\frac{Ge^{-\eta(T-t)}}{r-\eta}\left[1-e^{-(r-\eta)\frac{A}{G}}\right],\quad W\to\infty. \tag{4.9b}$$

The optimal withdrawal policy is to withdraw at the rate G.

(iii) $e^{-\eta(T-t)}\ge 1-k$

$$V(W,A,t)\approx e^{-\eta(T-t)}W+\frac{G}{r}\left[1-e^{-r\min\left(\frac{A}{G},T-t\right)}\right]-\frac{Ge^{-\eta(T-t)}}{r-\eta}\left[1-e^{-(r-\eta)\min\left(\frac{A}{G},T-t\right)}\right],\quad W\to\infty. \tag{4.9c}$$

The optimal withdrawal policy is to withdraw at the rate G.

In summary, the region of $\gamma=\infty$ (immediate withdrawal of finite amount) exists in the far field only when both A is above some threshold level A^{**} and time to expiry is longer than $-\dfrac{\ln(1-k)}{\eta}$. Otherwise, the optimal withdrawal policy is continuous withdrawal at the contractual rate G until the time of complete depletion of the guarantee account or maturity date, whichever comes earlier.

4.3 At time close to expiry

At time close to expiry, $t\to T^-$, the value of optionality associated with the terminal payoff almost vanishes. We would expect that the optimal strategy of zero withdrawal is almost ruled out, except under the unlikely event of $A\approx W$ (see the plot for $t=9.91667$ in Figure 5). To show the claim, we consider the value function at time close to expiry $V(W,A,T^-)$. By continuity of the value function, we have

$$V(W,A,T^-)=\begin{cases}(1-k)A & \text{if } (1-k)A>W\\ W & \text{if } (1-k)A<W\end{cases}.$$

For either payoff of $(1-k)A$ or W, we observe that the gradient constraint: $\frac{\partial V}{\partial W}+\frac{\partial V}{\partial A}>1$ is violated. Hence, the region of zero withdrawal ($\gamma=0$) almost vanishes as $t\to T^-$, except possibly in an asymptotically narrow strip along the separating boundary line $A=W$.

We would like to determine the asymptotic behavior of the value function at $t\to T^-$ under the following two separate cases:

(i) $W > (1-k)A$

Given that $t \to T^-$, the terminal payoff is almost surely to be W_T. As $\gamma = 0$ is ruled out when $t \to T^-$, the choice of taking either $\gamma = G$ or $\gamma = \infty$ depends on the relative magnitude of various depreciation factors; namely, $e^{-\eta(T-t)}$ due to proportional fee η and $1-k$ due to proportional penalty charge. When $T-t$ is small, $e^{-\eta(T-t)}$ is almost surely smaller than $1-k$. As a result, it is optimal to choose $\gamma = G$. The asymptotic value function is given by

$$
\begin{aligned}
V(W,A,t) &\approx \int_t^T Ge^{-ru}\,du + e^{-r(T-t)}E_t\left[W_T\right] \\
&= \frac{G}{r}\left[1-e^{-r(T-t)}\right] + e^{-\eta(T-t)}\left\{W - \frac{G}{r-\eta}\left[1-e^{-(r-\eta)(T-t)}\right]\right\}, \quad t \to T^-. \qquad (4.10a)
\end{aligned}
$$

To observe consistency with the earlier result, note that the solution (4.10a) is identical to eq. (4.9c) under the assumption: $T-t < \frac{A}{G}$.

(ii) $W < (1-k)A$

In this case, the terminal payoff is almost surely to be $(1-k)A$. In order to minimize loss of time value of the cash amount received, the optimal strategy is to withdraw the finite amount $A - G(T-t)$ immediately, followed by continuous withdrawal at the rate G in the remaining time until maturity date T. The asymptotic value function is given by

$$
\begin{aligned}
V(W,A,t) &\approx \int_t^T Ge^{-ru}\,du + (1-k)\left[A - G\left(T-t\right)\right] \\
&= \frac{G}{r}\left[1-e^{-r(T-t)}\right] + (1-k)\left[A - G(T-t)\right], \quad t \to T^-. \qquad (4.10b)
\end{aligned}
$$

To check for consistency again, the solution (4.10b) is seen to be identical to eq. (4.2) under the assumption: $T-t < -\frac{\ln(1-k)}{r}$ (this assumption holds when $t \to T^-$).

In summary, the value function at time close to expiry tends to the far field solution when $W > (1-k)A$ and the value function at $W=0$ when $W < (1-k)A$.

Optimal withdrawal policies under $G = 0$ at time close to expiry

We extend the above deduced optimal withdrawal policies at $t \to T^-$ to the special case $G = 0$. Recall that the value function under G is dependent on $Y = W/A$. We deduce that it is optimal to withdraw finite amount immediately when $Y < 1-k$ so that $Y^*_{\text{low}}(t)$ tends to $1-k$ as $t \to T^-$, $k \geq 0$. However, when $Y > 1-k$, eq. (4.10a) reveals the optimal policy of continuous withdrawal at the rate G. Given $G = 0$, this is equivalent to choose zero withdrawal optimally for all values of Y. Hence, we deduce that $Y^*_{\text{up}}(t) \to \infty$ as $t \to T^-$, $k > 0$. For the special case $k = 0$, it is always optimal to withdraw finite amount immediately for any value of Y since there is no penalty charge. Therefore, when $G = k = 0$, we have

$$
\lim_{t \to T^-} Y^*_{\text{low}}(t) = \lim_{t \to T^-} Y^*_{\text{up}}(t) = 1.
$$

Furthermore, it has been discussed in Section 4.2 that a necessary condition for adopting the optimal withdrawal policy of finite amount is $1 - k - e^{-\eta(T-t)} > 0$. For $k > 0$, this necessary condition is equivalent to $T - t > -\frac{\ln(1-k)}{\eta}$. In other words, $Y^*_{\text{up}}(t)$ is not defined when time to expiry is less or equal to $-\frac{\ln(1-k)}{\eta}$. Interestingly, this necessary condition becomes redundant when $k = 0$. As a result, $Y^*_{\text{up}}(t)$ can be defined for all times when $k = 0$.

5 Numerical studies on optimal dynamic withdrawal policies

In this section, we would like to present our numerical studies that were performed to verify the theoretical results on the separating boundaries and optimal dynamic withdrawal policies under various cases. Firstly, we present the recursive integration schemes that solve the integral equations for the determination of $Y^*_{\text{low}}(t)$ and $Y^*_{\text{up}}(t)$ under $G = 0$. We provide the financial interpretation of the optimal withdrawal policies as revealed by the numerical plots of the time dependent behavior of the upper and lower free boundaries of the corresponding simplified version of the pricing model of GMWB. Next, we present the numerical plots that show the time dependence of the separating boundaries under $G > 0$ in the W-A plane that divide the domain of the pricing model into "$\gamma = 0$" region, "$\gamma = G$" region and "$\gamma = \infty$" region. The separating boundaries under $G = 0$ form the oblique asymptotes of the separating boundaries of the pricing model of GMWB under the general case $G > 0$. We examine the pattern of the withdrawal regions at different calendar times and under varying values of the proportional penalty charge parameter k. As predicted by eqs. (4.9a,b,c), the pattern of the withdrawal regions does demonstrate a drastic change when time of expiry falls below a threshold value or the penalty charge parameter increases beyond a threshold value. For the limiting cases of time close to expiry and small value of A, we show the plots of the withdrawal regions that verify the corresponding theoretical asymptotic results derived in Section 4. Lastly, we present the comparison of numerical values that verify the different forms of the analytical approximation formula of the value function under the limiting case of large value of W [far field boundary condition as shown in eqs. (4.9a,b,c)].

5.1 Numerical studies on the free boundaries: $Y^*_{\text{low}}(t)$ and $Y^*_{\text{up}}(t)$

We construct the recursive integration schemes for the solution of the integral equations that solve for the free boundaries: $Y^*_{\text{low}}(t)$ and $Y^*_{\text{up}}(t)$ [see eqs. (3.10) and (3.11a,b)]. As usual, we compute the free boundaries backward in time, starting with $Y^*_{\text{low}}(T^-) = 1 - k$ [recall that $Y^*_{\text{up}}(t)$ is not defined for $t \geq T - \tau^*$, where $\tau^* = -\frac{\ln(1-k)}{\eta}$]. In our later exposition, it is more convenient to use the time to expiry $\tau = T - t$ instead of the calendar time t as the temporal variable. We divide the overall policy life by n equally spaced sub-intervals $[\tau_{i-1}, \tau_i]$, where $\Delta\tau = \tau_i - \tau_{i-1}$, $i = 1, 2, \ldots, n$, so $\tau_i = i\Delta\tau$, $i = 0, 1, \ldots, n$. In our numerical calculations, we set i^* to be the largest integer that observes $\tau_{i^*} \leq \tau^*$ and compute numerical values for $Y^*_{\text{up}}(\tau_i)$ for $i = i^* + 1, i^* + 2, \ldots, n$. We define

$$
\begin{aligned}
f_{\text{low}}(x, y; \tau, u) &= (1-k) r e^{-ru} N(-d_{22}(x, u; y)), \\
f_{\text{up}}(x, y; \tau, u) &= r e^{-ru} N(d_{12}(x, u; y)) - (r - \eta) e^{-\eta\tau} e^{-(r-\eta)u} N(d_{12}(x, u; y)), \\
g(x; \tau) &= (1-k) e^{-r\tau} + c\,(x, \tau; 1-k)\,.
\end{aligned}
$$

The numerical procedure is split into two sequential steps: (i) determination of $Y^*_{\text{low}}(\tau_k)$, $k = 1, 2, \ldots, i^*$, starting with $Y^*_{\text{low}}(\tau_0) = 1-k$, (ii) simultaneous calculations of $Y^*_{\text{low}}(\tau_k)$ and $Y^*_{\text{up}}(\tau_k)$, $k = i^*+1, i^*+2, \ldots, n$. By approximating the integral representation of the withdrawal premium using the trapezoidal rule in numerical integration, the recursive integration scheme for the determination of $Y^*_{\text{low}}(\tau_k)$, $k = 1, 2, \ldots, i^*$, is depicted as follows:

$$
\begin{aligned}
1 - k = \frac{\Delta\tau}{2} \Big[& f_{\text{low}}\left(Y^*_{\text{low}}(\tau_k), Y^*_{\text{low}}(\tau_k); \tau_k, \tau_0\right) + f_1\left(Y^*_{\text{low}}(\tau_k), Y^*_{\text{low}}(\tau_0); \tau_k, \tau_k\right) \\
& + 2\sum_{i=1}^{k-1} f_{\text{low}}\left(Y^*_{\text{low}}(\tau_k), Y^*_{\text{low}}(\tau_k - \tau_i); \tau_k, \tau_i\right) \Big] + g\left(Y^*_{\text{low}}(\tau_k); \tau_k\right),
\end{aligned} \tag{5.1}
$$

where $Y^*_{\text{low}}(\tau_k)$, $k = 1, 2, \ldots, i^*$, are solved sequentially using a root finding method. It is necessary to solve the following coupled system of non-linear algebraic equations for the simultaneous calculations of $Y^*_{\text{low}}(\tau_k)$ and $Y^*_{\text{up}}(\tau_k)$, $k = i^* + 1, i^* + 2, \ldots, n$, derived from the numerical approximation of the pair of integral equations (3.11a,b):

$$\begin{aligned} 1 - k = \frac{\Delta\tau}{2}&[f_{\text{low}}(Y^*_{\text{low}}(\tau_k), Y^*_{\text{low}}(\tau_k); \tau_k, \tau_0) + f_{\text{low}}(Y^*_{\text{low}}(\tau_k), Y^*_{\text{low}}(\tau_0); \tau_k, \tau_k) \\ &+ 2\sum_{i=1}^{k-1} f_{\text{low}}(Y^*_{\text{low}}(\tau_k), Y^*_{\text{low}}(\tau_k - \tau_i); \tau_k, \tau_i)] \\ &+ \left(\frac{\Delta\tau}{2} + \frac{\tau_{i^*+1} - \tau^*}{2}\right) f_{\text{up}}\left(Y^*_{\text{low}}(\tau_k), Y^*_{\text{up}}\left(\tau_{i^*+1}\right); \tau_k, \tau_k - \tau_{i^*+1}\right) \\ &+ \Delta\tau \sum_{i=1}^{k-i^*-2} f_{\text{up}}(Y^*_{\text{low}}(\tau_k), Y^*_{\text{up}}(\tau_k - \tau_i); \tau_k, \tau_i) + g\left(Y^*_{\text{low}}(\tau_k); \tau_k\right); \end{aligned} \tag{5.2a}$$

$$\begin{aligned} 1 - k + e^{-\eta\tau_k}\left(Y^*_{\text{up}}(\tau_k) - 1\right) = \frac{\Delta\tau}{2}&[f_{\text{low}}(Y^*_{\text{up}}(\tau_k), Y^*_{\text{low}}(\tau_k); \tau_k, \tau_0) + f_{\text{low}}(Y^*_{\text{up}}(\tau_k), Y^*_{\text{low}}(\tau_0); \tau_k, \tau_k) \\ &+ 2\sum_{i=1}^{k-1} f_{\text{low}}(Y^*_{\text{up}}(\tau_k), Y^*_{\text{low}}(\tau_k - \tau_i); \tau_k, \tau_i)] \\ &+ \left(\frac{\Delta\tau}{2} + \frac{\tau_{i^*+1} - \tau^*}{2}\right) f_{\text{up}}\left(Y^*_{\text{up}}(\tau_k), Y^*_{\text{up}}\left(\tau_{i^*+1}\right); \tau_k, \tau_k - \tau_{i^*+1}\right) \\ &+ \Delta\tau \sum_{i=1}^{k-i^*-2} f_{\text{up}}(Y^*_{\text{up}}(\tau_k), Y^*_{\text{up}}(\tau_k - \tau_i); \tau_k, \tau_i) + g\left(Y^*_{\text{up}}(\tau_k); \tau_k\right). \end{aligned} \tag{5.2b}$$

As a remark, it is more convenient to choose the spacing $\Delta\tau$ such that τ^* falls within $(\tau_{i^*}, \tau_{i^*+1})$ in order to avoid the ambiguous approximation of the infinite value of $Y^*_{\text{up}}(\tau)$ at $\tau = \tau^*$.

Numerical tests

We performed various numerical tests to reveal the properties of the free boundaries: $Y^*_{\text{low}}(\tau)$ and $Y^*_{\text{up}}(\tau)$ and demonstrate the effectiveness of the numerical recursive integration schemes. To assess the numerical accuracy of our proposed numerical schemes, we use the same set of GMWB parameter values as adopted by Huang and Forsyth (2012) (see Table 1 below) while G is set to be zero.

Parameter	Value
Interest rate r	0.05
Maximum no penalty withdrawal rate G	0/year
Volatility σ	0.3
Insurance fee η	0.0312856
Initial lump-sum premium w_0	100
Initial guarantee account balance A_0	100
Initial personal annuity account balance W_0	100

Table 1: The GMWB contract parameter values used in the numerical calculation of the pair of free boundaries: $Y^*_{\text{low}}(t)$ and $Y^*_{\text{up}}(t)$.

In Figure 4, we show the plots of $Y^*_{\text{low}}(\tau)$ and $Y^*_{\text{up}}(\tau)$ against time to expiry τ for varying values of proportional penalty charge k. The numerical plots reveal good agreement with the theoretical results in our earlier discussions. Firstly, we observe the trend that the free boundaries $Y^*_{\text{low}}(\tau)$ and $Y^*_{\text{up}}(\tau)$ tend to one from below and above, respectively, as time to expiry lengthens. Recall that $Y^*_{\text{up}}(\tau)$ is not defined for $\tau \le \tau^*$, where $\tau^* = -\frac{\ln(1-k)}{\eta}$. Taking $\eta = 0.0312856$, for $k = 0.1$ and $k = 0.05$, these threshold values are found to be 3.3677 and 1.6395, respectively; These numerical values confirm well with the plots of $Y^*_{\text{up}}(\tau)$ shown in Figure 4. For $k = 0$, $Y^*_{\text{up}}(\tau)$ is well defined for all values of τ with $Y^*_{\text{up}}(\tau_0) = 1$. The plots of the lower free boundaries $Y^*_{\text{low}}(\tau)$ for varying values of k in Figure 4 are seen to be quite insensitive to change in value of k; and $Y^*_{\text{low}}(\tau_0) = 1-k$. The plots of the free boundaries reveal that the continuation region widens as the value of k increases. This is consistent with financial intuition since the holder should wait for a higher or lower value of W before adopting optimal withdrawal of the whole guarantee account A.

In Table 2, we list the numerical values of $Y^*_{\text{low}}(\tau)$ and $Y^*_{\text{up}}(\tau)$ at varying values of τ using our proposed numerical recursive integration schemes with varying number of sub-intervals n and compare with those reported in Huang and Forsyth (2012). We observe good agreement with Huang-Forsyth's results and fast convergence of the numerical results is observed even with relatively low values of n (say, $n = 40$).

		$\tau = 5$				$\tau = 10$			
		Recursive scheme			Huang-Forsyth	Recursive scheme			Huang-Forsyth
	n	40	80	120		40	80	120	
$k = 0.05$	$Y^*_{\text{up}}(\tau)$	1.80919	1.81057	1.81101	1.80998	1.62899	1.62868	1.62937	1.62172
	$Y^*_{\text{low}}(\tau)$	0.64776	0.64781	0.64782	0.65014	0.68767	0.68765	0.68764	0.69027

Table 2: The numerical values of $Y^*_{\text{low}}(\tau)$ and $Y^*_{\text{up}}(\tau)$ at varying values of τ and $k = 0.05$ are shown. Here, n is the total number of sub-intervals used in the recursive integration scheme. We observe good agreement with the numerical results reported in Huang and Forsyth (2012).

5.2 Optimal withdrawal boundaries under $G > 0$

We would like to verify the theoretical results (see Section 4) on the optimal withdrawal strategies under the general case $G > 0$ by performing similar calculations shown in Huang and Forsyth (2012). In Figure 5, we show the plots of the various withdrawal regions ($\gamma = 0$, $\gamma = G$ and $\gamma = \infty$) in the W-A plane at varying values of the calendar time t and $k = 0.1$. The parameter values of the GMWB contract used in the calculations are the same as those in Table 1; in addition, T is taken to be 10 and $G = 10$.

When $t = 0$ and $t = 5.0$, which are sufficiently far from expiry, we observe the existence of the horizontal asymptote: $A = A^{**}$, where A^{**} is the unique solution to eq. (4.8). We obtain $A^{**} = 30.2118$ in our sample calculations. At large value of W, we deduce from the plots for $t = 0$ and $t = 5.0$ that it is optimal to withdraw immediate finite amount when $A > A^{**}$ ($\gamma = \infty$) and withdraw continuously at $\gamma = G$ when $A \le A^{**}$. It is seen that the horizontal line $A = A^{**}$ is the horizontal asymptote for the "$\gamma = \infty$" region in the far field. In these two plots, we also add the oblique dashed lines that are the optimally separating boundaries between the "$\gamma = 0$" and "$\gamma = \infty$" regions corresponding to the scenario $G = 0$. These two dashed lines are seen to be the oblique asymptotes for the "$\gamma = \infty$" regions. Interestingly, they are also asymptotes to the small island of "$\gamma = 0$" region close to $A \to 0$ and $W \to 0$. These plots are consistent with the asymptotic behavior of the value function at large value of A and small value of A, the studies of which are relegated to a later work. When $t = 7.0$ and $t = 9.91667$, the corresponding "$\gamma = \infty$" region in the far field vanishes. This is consistent with the theoretical prediction in Section 4.2 that the horizontal asymptote does not exist when time to expiry is shorter than $-\frac{\ln(1-k)}{\eta}$ (equals 3.3677 in this set of sample calculations). As expected, for $t = 7.0$ and $t = 9.91667$ (too close to expiry), the optimal withdrawal policy is to withdraw at the rate G [see eq. (4.9c)] and the "$\gamma = G$" region prevails at large value of W. For all values of t, we observe from the plots that it is always optimal to choose $\gamma = \infty$ when W is sufficiently small and A is above the threshold value $-\frac{\ln(1-k)}{r}$ (equals 21.0721 in this set of sample calculations). Also, when the ratio of A and W are within certain range of values, it is optimal not to withdraw at all. It is interesting to observe the evolution of the "$\gamma = 0$" region, from a small island at $t = 0$ to a strip at later times. The strip becomes narrower as t tends to maturity date T, consistent with $Y^*_{\text{up}}(t) \to 1^+$ and $Y^*_{\text{low}}(t) \to 1^-$ as $t \to T$. It is observed in the plot for $t = 9.91667$ that the "$\gamma = 0$" region becomes a narrow strip lying within the "$\gamma = G$" region. The strip is around the line: $W = A$ (see Section 4.3).

5.3 Numerical accuracy of the analytical approximate formula at $W \to \infty$

We would like to examine numerical accuracy of the asymptotic formulas (4.9a,b,c) for the value function at $W \to \infty$. Huang and Forsyth (2012) performed their calculations of the value function using the penalty approximation scheme and their numerical results at varying values of A and W are shown in the first column in Table 3. We performed numerical valuation of the asymptotic formulas and the corresponding numerical results are shown in the second column in the same table for comparison. Very good agreement between the two sets of numerical values is observed even at moderate values of W. Using Huang-Forsyth's penalty approximation scheme calculations as the benchmark, the percentage differences in the numerical values between Huang-Forsyth's results and those computed using asymptotic formulas (4.9a,b,c) are truly very small (see the last column in Table 3).

	Huang-Forsyth	asymptotic formulas (4.9a,b,c)	percentage difference
$A = 10$, $W = 80$	61.017327	61.016989	-0.00055%
$A = 10$, $W = 100$	75.644804	75.644094	-0.00094%
$A = 20$, $W = 80$	63.18349	63.184184	0.00110%
$A = 20$, $W = 100$	77.810965	77.811287	0.00041%
$A = 30$, $W = 80$	65.035297	65.030709	-0.00705%
$A = 30$, $W = 100$	79.657330	79.657813	0.00061%
$A = 40$, $W = 80$	66.763615	66.717224	-0.06949%
$A = 40$, $W = 100$	81.345396	81.344328	-0.00131%
$A = 50$, $W = 80$	68.821701	68.403672	-0.60741%
$A = 50$, $W = 100$	83.038705	83.030776	-0.00955%

Table 3: Comparison of the numerical values for the value function at varying values of A and W that are obtained from Huang-Forsyth's (2012) numerical calculations and asymptotic formulas (4.9a,b,c) at large value of W. Very good agreement between the two sets of numerical values is observed even at moderate values of W.

Remark

When $t < T + \frac{\ln(1-k)}{\eta}$, solution to A^{**} in eq. (4.8) always exists. The value of A^{**} is seen to be dependent on G. We do expect A^{**} to approach zero as G goes to zero, a result that is consistent with absence of "$\gamma = G$" region when $G = 0$. As a check, we obtain $A^{**} = 0.0604236$ at $G = 0.02$, $A^{**} = 0.0302118$ at $G = 0.01$ and $A^{**} = 0.00302118$ at $G = 0.001$. The convergence trend of A^{**} approaching to zero value as $G \to 0$ is apparent.

6 Conclusion

We have performed a complete characterization of the optimal dynamic withdrawal policies in GMWB contracts. The optimal withdrawal policies of zero withdrawal, withdrawal at the contractual rate and withdrawal of an immediate amount are determined by the competing forces between time value of cash, proportional penalty charge and optionality associated with the terminal payoff. We have derived asymptotic formulas for the value function and optimal withdrawal boundaries at various limiting conditions: (i) zero value of the contractual withdrawal rate, (ii) perpetuality, (iii) time close to expiry, (iv) infinitely large value of the underlying fund value. Under perpetuality, we show that the GMWB value function is independent of the policy fund value. As the value of optionality derived from the terminal payoff vanishes, it is then always non-optimal to adopt the policy of zero withdrawal. There remain two choices for the optimal withdrawal policy. When the guarantee account value is above certain threshold value, the optimal withdrawal policy is to withdraw finite amount immediately, then followed by withdrawal at the contractual rate of the remaining guarantee account. Otherwise, it is optimal to withdraw at the contractual rate until complete depletion of the guarantee account. Similar optimal withdrawal policies are adopted for the other two limiting cases, where the policy fund value is much larger than the guarantee account and at time close to expiry.

In summary, we manage to deduce various asymptotes for the free boundaries that separate different withdrawal regions in the domain of the pricing model. When the underlying fund value is large, it is optimal to withdraw immediate amount provided that the guarantee account value is sufficiently high and the current time is sufficiently far from expiry. On the other hand, when the underlying fund value is sufficiently small, it is always optimal to withdraw immediate amount provided that the guarantee account value is not too low. When the ratio of the underlying fund value to the guarantee account value falls within certain range, it may become optimal to adopt the optimal policy of zero withdrawal to take advantage of optionality associated with the terminal payoff.

Reference

Bauer, D., Kling, A., Russ, J. (2008) A universal pricing framework for guaranteed minimum benefits in variable annuities. *ASTIN Bulletin*, 38: 621-651.

Chen, Z., Forsyth, P.A. (2008) A numerical scheme for the impulse control formulation for pricing variable annuities with a guaranteed minimum withdrawal benefit (GMWB). *Numerische Mathematik*, 109: 535-569.

Chen, Z., Vetzal, K., Forsyth, P.A. (2008) The effect of modelling parameters on the value of GMWB guarantees. *Insurance: Mathematics and Economics*, 43: 165-173.

Dai, M., Kwok, Y.K., Zong, J. (2008) Guaranteed minimum withdrawal benefit in variable annuities. *Mathematical Finance*, 8(6): 561-569.

Davis, M.H.A., Norman, A.R. (1990) Portfolio selection with transaction costs. *Mathematics of Operations Research*, 15: 676-713.

Detemple, J. (2005) *American-Style Derivatives: Valuation and Computation*, Chapman & Hall/CRC, New York.

Flemming, W.H., Soner, H.M. (1993) *Controlled Markov Processes and Viscosity Solutions*, Springer-Verlag, New York.

Huang, Y., Forsyth, P.A. (2012) Analysis of a penalty method for pricing a guaranteed minimum withdrawal benefit (GMWB). *IMA Journal of Numerical Analysis*, 32(1): 320-351.

Huang, Y., Forsyth, P.A., Labahn, G. (2013) Iterative methods for the solution of a singular control formulation of a GMWB pricing problem. To appear in *Numerische Mathematik*.

Kwok, Y.K. (2008) *Mathematical Models of Financial Derivatives*, second edition, Springer, Berlin.

Merton, R.C. (1971) Optimal consumption and portfolio rules in a continuous time model. *Journal of Economic Theory*, 3: 373-413.

Milevsky, M.A., Salisbury, T.S. (2006) Financial valuation of guaranteed minimum withdrawal benefits. *Insurance: Mathematics and Economics*, 38: 21-38.

Protter, P. (1990) *Stochastic Integration and Differential Equation*, Springer, New York.

Yong, J., Zhou, X. (1999) *Stochastic controls: Hamiltonian systems and HJB equations*, Springer, New York.

Acknowledgements

This work was supported by the Hong Kong Research Grants Council under the General Research Funds. The authors benefit from the discussion with Min Dai, Marcus Lo and Dennis Chan.

Appendix A - Proof of withdrawal premium formula (3.9)

We may adopt the notion of delayed exercise premium in an American option to derive the delayed withdrawal premium $M(Y,t)$. Similar to an American put option, the payoff of strike price of $1-k$ dollars at $Y=Y^*_{\text{low}}(t)$ contributes the following amount to the withdrawal premium:

$$\begin{aligned} &E_t\left[\int_t^T e^{-r(u-t)}(1-k)r\mathbf{1}_{\{Y_u\le Y^*_{\text{low}}(u)\}}\ du\right] \\ &= (1-k)r\int_t^T e^{-r(u-t)}N\left(-\frac{\ln\frac{Y}{Y^*_{\text{low}}(u)}+\left(r-\eta-\frac{\sigma^2}{2}\right)(u-t)}{\sigma\sqrt{u-t}}\right)\ du. \end{aligned} \tag{A.1}$$

In the above expectation calculation, we have made use of the observation that the dynamics of Y_u for $u>t$ is the same as that of W_u since A_u stays at the same value as A_t when Y_u remains in the continuation region.

Define $\tau^*=-\frac{\ln(1-k)}{\eta}$ and recall that $Y^*_{\text{up}}(t)$ is not defined for $t\ge T+\frac{\ln(1-k)}{\eta}=T-\tau^*$, so there is no contribution to the withdrawal premium over this time period close to expiry. Consider the payoff at the upper free boundary, the term $e^{-\eta(T-t)}Y$ does not contribute to the withdrawal premium since $P=e^{-\eta(T-t)}Y$ satisfies the governing equation. For the remaining terms: $1-k-e^{-\eta(T-u)}$, suppose Y_u falls within the upper withdrawal region: $Y_u>Y^*_{\text{up}}(u)$, the amount that has to be paid to the holder over $(u,u+du)$ as dollar compensation is

$$\left\{r\left[1-k-e^{-\eta(T-u)}\right]+\eta e^{-\eta(T-u)}\right\}\ du=\left[(1-k)r-(r-\eta)e^{-\eta(T-u)}\right]\ du,$$

if the holder agrees not to withdraw even when it is optimal to do so. The first term is the interest earned from holding $1-k-e^{-\eta(T-u)}$ dollars over du while the second term arises from the adjustment of the notional amount of the money market account over $(u,u+du)$. For $t<T-\tau^*$, the withdrawal premium corresponding to the withdrawal region beyond the upper free boundary over the period $(t,T-\tau^*)$ is then given by

$$\begin{aligned} &E_t\left[\int_t^{T-\tau^*} e^{-r(u-t)}\left[(1-k)r-(r-\eta)e^{-\eta(T-u)}\right]\mathbf{1}_{\{Y_u\ge Y^*_{\text{up}}(u)\}}\ du\right] \\ &= (1-k)r\int_t^{T-\tau^*} e^{-r(u-t)}N\left(\frac{\ln\frac{Y}{Y^*_{\text{up}}(u)}+\left(r-\eta-\frac{\sigma^2}{2}\right)(u-t)}{\sigma\sqrt{u-t}}\right)\ du \\ &\quad -(r-\eta)\int_t^{T-\tau^*} e^{-r(u-t)}e^{-\eta(T-u)}N\left(\frac{\ln\frac{Y}{Y^*_{\text{up}}(u)}+\left(r-\eta-\frac{\sigma^2}{2}\right)(u-t)}{\sigma\sqrt{u-t}}\right)\ du. \end{aligned} \tag{A.2}$$

In summary, when $t<T-\tau^*$, the withdrawal premium $M(Y,t)$ is given by the sum of the terms in eqs. (A.1) and (A.2). When $t\ge T-\tau^*$, $M(Y,t)$ is given by the single term in eq. (A.1). Combining these results, we obtain the withdrawal premium formula in eq. (3.9).

Appendix B - Proof of eq. (4.6)

The value of optionality in the terminal payoff tends to zero under perpetuality. As a result, the optimal strategy is to withdraw from the guarantee account as soon as possible in order to avoid loss of time value; while at the same time, one takes into account the tradeoff between proportional penalty charge on excessive withdrawal and loss of time value of cash. The optimal withdrawal policy is either (i) immediate withdrawal of a finite amount, then followed by continuous withdrawal at the contractual rate G until depletion of the guarantee account, or (ii) continuous withdrawal at G until depletion of the guarantee account. As a remark, one may show that once continuous withdrawal has started, it is always non-optimal to withdraw finite amount later due to convexity property of the value function [see a similar result in Section 3].

With no dependence on W, the price function $V_\infty(W,A)$ is determined by the optimal strategy on the choice of the amount of immediate withdrawal, then followed by continuous withdrawal at the contractual rate G. Let δ be the finite amount of immediate withdrawal. Including the special case of zero withdrawal where $\delta = 0$, the value function $V_\infty(W,A)$ is determined by finding δ such that

$$\begin{aligned} V_\infty(W,A) &= \sup_{0\le\delta\le A}\left[(1-k)\delta + \int_0^{\frac{A-\delta}{G}} Ge^{-ru}\,du\right] \\ &= \sup_{0\le\delta\le A}\left[(1-k)\delta + \frac{G}{r}\left(1-e^{-r\frac{A-\delta}{G}}\right)\right]. \end{aligned}$$

Let $L(\delta) = (1-k)\delta + \frac{G}{r}\left(1-e^{-r\frac{A-\delta}{G}}\right)$, we have

$$\frac{dL(\delta)}{d\delta} = 1-k-e^{-r\frac{A-\delta}{G}},$$

which is seen to be decreasing over $[0,A]$ and $\left.\frac{dL(\delta)}{d\delta}\right|_{\delta=A} = -k < 0$. Therefore, $\frac{dL(\delta)}{d\delta} = 0$ has a unique solution in $[0,A]$ if and only if

$$\left.\frac{dL(\delta)}{d\delta}\right|_{\delta=0} = 1-k-e^{-r\frac{A}{G}} \ge 0.$$

Let $A^* = -\frac{G}{r}\ln(1-k)$, the above inequality is equivalent to $A \ge A^*$. The inequality indicates the tradeoff between the net proportional amount $1-k$ received after paying the proportional penalty charge and the discount factor $e^{-r\frac{A}{G}}$ over the time period $\frac{A}{G}$. We consider the following two separate cases:

1. When $A \ge A^*$, the unique solution to $\frac{dL}{d\delta} = 0$ is $\delta = A - A^*$. The value function is then given by

$$V_\infty(W,A) = (1-k)(A-A^*) + \frac{G}{r}\left(1-e^{-\frac{r}{G}A^*}\right) = (1-k)(A-A^*) + \frac{G}{r}k. \tag{B.1}$$

2. When $A < A^*$, we should take $\delta = 0$ as the optimal choice. The resulting value function is given by

$$V_\infty(W,A) = \frac{G}{r}(1-e^{-r\frac{A}{G}}). \tag{B.2}$$

It is relatively straightforward to check that the above solution for $V_\infty(W,A)$ satisfies the linear complementarity formulation (4.3) and the corresponding auxiliary conditions.

Appendix C - Proof of eqs. (4.9a,b,c)

Similar to the case of perpetuality presented in Appendix B, we determine the amount of immediate withdrawal δ such that the value function $V(W,A,t)$ is maximized. After an immediate withdrawal of δ (note that δ may be zero), the maximum length of the period of continuous withdrawal at the rate G is given by $\frac{A-\delta}{G}$. With reference to finite maturity date T, the period of continuous withdrawal lasts until T^*, where

$$T^* = \min\left(T, t+\frac{A-\delta}{G}\right).$$

The policy fund W_t is depleted by continuous withdrawal, so its dynamics under the risk neutral measure Q is governed by

$$dW_u = (r-\eta)W_u\,du + \sigma W_u\,dB_u - G\,du, \quad t<u<T^*. \tag{C.1}$$

The value function at the far field, $W \to \infty$, is determined by finding δ such that

$$V(W,A,t) = \sup_{0\le\delta\le A}\left\{(1-k)\,\delta + \int_t^{T^*} Ge^{-ru}\,du + e^{-r(T-t)}E_t\left[W_T\right]\right\}.$$

To compute $E_t\left[W_T\right]$, we use the tower law of conditional expectation:

$$E_t\left[W_T\right] = E\left[E\left[W_T\,|\mathcal{F}_{T^*}\right]|\mathcal{F}_t\right] = E_t\left[W_{T^*}e^{(r-\eta)(T-T^*)}\right].$$

By solving the dynamics equation (C.1), we obtain

$$W_{T^*} = e^{\left(r-\eta-\frac{\sigma^2}{2}\right)(T^*-t)+\sigma(B_{T^*}-B_t)}\left[W_{t^+} - G\int_t^{T^*} e^{-\left(r-\eta-\frac{\sigma^2}{2}\right)(u-t)-\sigma(B_u-B_t)}\,du\right],$$

where $W_{t^+} = W - \delta$. By taking the expectation conditional on $\mathcal{F}_t$, we obtain

$$\begin{aligned} e^{-r(T-t)}E_t\left[W_T\right] &= e^{-r(T-t)}e^{(r-\eta)(T-T^*)}e^{(r-\eta)(T^*-t)}\left[(W-\delta) - G\int_t^{T^*} e^{-(r-\eta)(u-t)}\,du\right] \\ &= e^{-\eta(T-t)}\left\{W-\delta-\frac{G}{r-\eta}\left[1-e^{-(r-\eta)(T^*-t)}\right]\right\}. \end{aligned}$$

Putting the results together, we have

$$V(W,A,t) = \sup_{0\le\delta\le A} L(\delta),$$

where

$$L(\delta) = \left[1-k-e^{-\eta(T-t)}\right]\delta + e^{-\eta(T-t)}W + \frac{G}{r}\left[1-e^{-r(T^*-t)}\right] - \frac{Ge^{-\eta(T-t)}}{r-\eta}\left[1-e^{-(r-\eta)(T^*-t)}\right]. \tag{C.2}$$

Note that T^* has an implicit dependence on δ, where

$$T^*-t = \begin{cases} T-t & \text{when } A \ge G(T-t) \\ \frac{A-\delta}{G} & \text{when } A < G(T-t) \end{cases}.$$

We determine the optimal choice of δ by invoking the standard constrained convex optimization procedure, where the optimal solution δ^* must satisfy

$$(\delta-\delta^*)\left.\frac{dL}{d\delta}\right|_{\delta=\delta^*} \le 0 \quad \text{for } \delta\in[0,A].$$

The first and second order derivatives of $L\left(\delta\right)$ are found to be

$$\begin{aligned} \frac{dL}{d\delta} &= \left[1-k-e^{-\eta(T-t)}\right] - \left[e^{-r\left(\frac{A-\delta}{G}\right)} - e^{-\eta(T-t)}e^{-(r-\eta)\left(\frac{A-\delta}{G}\right)}\right]\mathbf{1}_{\{A<G(T-t)\}}, \\ \frac{d^2L}{d\delta^2} &= \left[e^{-\eta(T-t)}\frac{r-\eta}{G}e^{-(r-\eta)\left(\frac{A-\delta}{G}\right)} - \frac{r}{G}e^{-r\left(\frac{A-\delta}{G}\right)}\right]\mathbf{1}_{\{A<G(T-t)\}}. \end{aligned}$$

It is easily seen that $\frac{d^2L}{d\delta^2} < 0$ when $A < G(T-t)$ and $\frac{d^2L}{d\delta^2} = 0$ when $A \ge G(T-t)$. Therefore, $\frac{dL}{d\delta}$ is monotonically decreasing when $A < G(T-t)$. Furthermore, we observe $\left.\frac{dL}{d\delta}\right|_{\delta=A} = -k < 0$ for $k>0$. Hence, when $\left.\frac{dL}{d\delta}\right|_{\delta=0} \le 0$, we have $\delta^*=0$; while when $\left.\frac{dL}{d\delta}\right|_{\delta=0} > 0$, δ^* lies within $(0,A)$ and it is the unique root of the equation derived from the first order condition: $\left.\frac{dL}{d\delta}\right|_{\delta=\delta^*} = 0$. Since $\left.\frac{dL}{d\delta}\right|_{\delta=0}$ has dependence on A, we write

$$h(A) = \left.\frac{dL}{d\delta}\right|_{\delta=0} = \left[1-k-e^{-\eta(T-t)}\right] - \left[e^{-r\frac{A}{G}} - e^{-\eta(T-t)-(r-\eta)\frac{A}{G}}\right]\mathbf{1}_{\{A<G(T-t)\}}.$$

It is straightforward to show that $h(A)$ is monotonically increasing for $A \in [0, G(T-t))$ and stays at the constant value $1-k-e^{-\eta(T-t)}$ for $A \geq G(T-t)$. Therefore, when $1-k-e^{-\eta(T-t)} \leq 0$, we have $h(A) \leq 0$ for all values of A. On the other hand, when $1-k-e^{-\eta(T-t)} > 0$, we have $h(A) > 0$ when $A > A^{**}$ and $h(A) \leq 0$ when $A \leq A^{**}$, where A^{**} is the unique root to the following algebraic equation:

$$1-k-e^{-\eta(T-t)} - e^{-r\frac{A}{G}} \left\{ 1 - e^{-\eta\left[(T-t)-\frac{A}{G}\right]} \right\} = 0.$$

It is easily seen that $A^{**} < G(T-t)$. The threshold A^{**} is determined by the various competing factors related to the relative magnitudes of proportional penalty charge, loss of time value of the amount not withdrawn and insurance fee paid on the policy fund.

We deduce that both (i) $1-k-e^{-\eta(T-t)} > 0$ and (ii) $A > A^{**}$ are the necessary and sufficient conditions for the adoption of the optimal strategy of an immediate withdrawal of the finite amount $A - A^{**}$ followed by continuous withdrawal at the rate G. The first condition states that the net proportional amount received after paying the proportional penalty charge has to be above the insurance fee paid on the policy fund over the remaining life of the policy. The second condition arises since an immediate withdrawal would avoid larger loss on the time value when A is above some threshold value A^{**}. When both conditions are satisfied, we choose $\delta = A - A^{**}$ and the remaining guarantee amount A^{**} is withdrawn continuously at the rate G. By substituting $\delta = A - A^{**}$ into eq. (C.2), the resulting far field boundary condition for the value function $V(W, A, t)$ at $W \to \infty$ is then given by eq. (4.9a). However, when $A < A^{**}$, an immediate withdrawal becomes non-optimal and the guarantee amount A will be withdrawn continuously at the rate of G over the period A/G. The depletion of the guarantee occurs prior to maturity date since $A < A^{**} < G(T-t)$. By substituting $\delta = 0$ and $T^* - t = \frac{A}{G}$ in eq. (C.2), this leads to the far field boundary condition shown in eq. (4.9b). Lastly, an immediate withdrawal is always non-optimal when the first condition: $1-k-e^{-\eta(T-t)} > 0$ is not observed. The depletion of the guarantee account occurs over the period $\min\left(\frac{A}{G}, T-t\right)$. By substituting $\delta = 0$ and $T^* - t = \min\left(\frac{A}{G}, T-t\right)$, the corresponding far field boundary condition is given by eq. (4.9c).

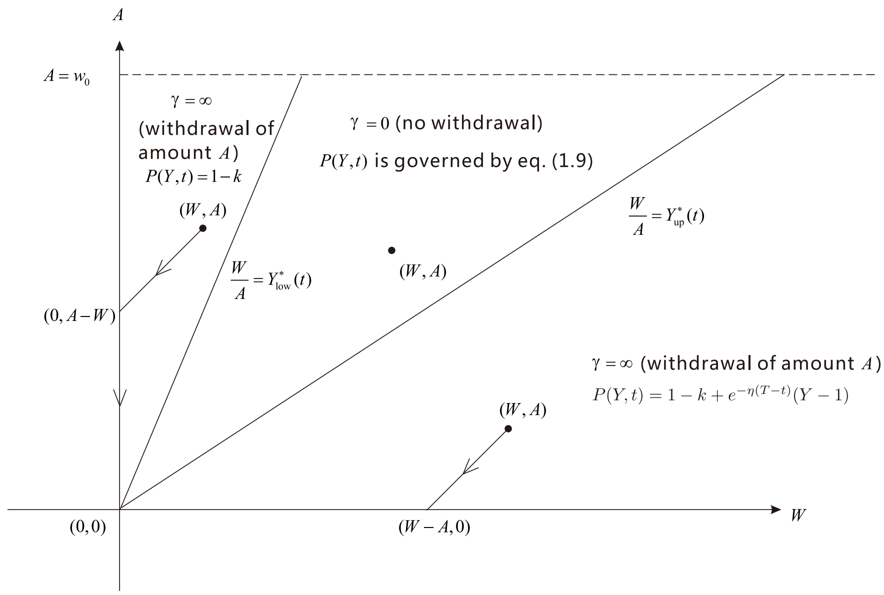

Figure 1. We illustrate the separation of the solution domain $\{(W, A)\colon W \geq 0 \text{ and } 0 \leq A \leq w_0\}$ of the pricing model under $G = 0$ into the two withdrawal regions of infinite withdrawal rate ($\gamma = \infty$) and continuation region ($\gamma = 0$). The separating boundaries are a pair of straight lines: (i) $\frac{W}{A} = Y^*_{\text{low}}(t)$, $Y^*_{\text{low}}(t) < 1$, and (ii) $\frac{W}{A} = Y^*_{\text{up}}(t)$, $Y^*_{\text{up}}(t) > 1$. When (W, A) falls within either one of the withdrawal regions, the whole guarantee amount A is depleted immediately (see the two arrows shown in the two regions where $\gamma = \infty$).

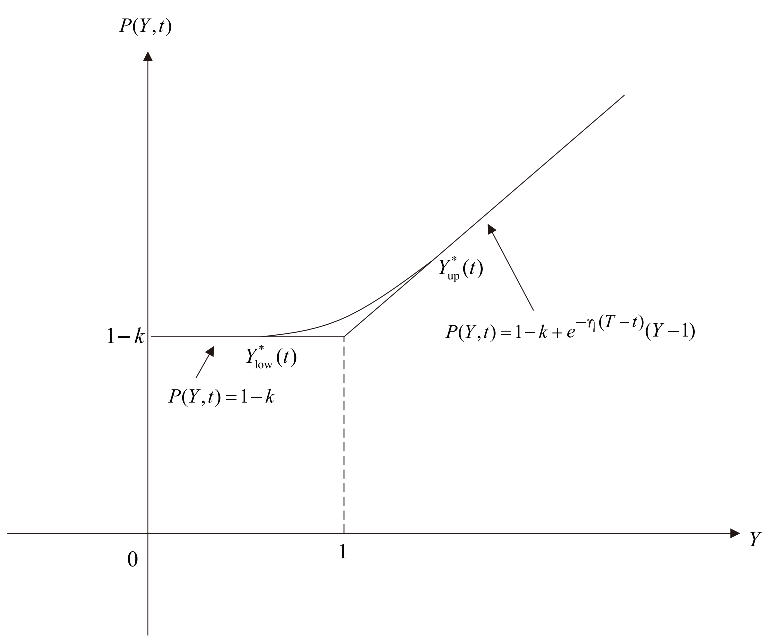

Figure 2. The plot of $P(Y, t)$ against Y and the obstacle function: $1 - k + \max(e^{-\eta(T-t)}(Y - 1), 0)$. In the continuation (no withdrawal) region: $Y^*_{\text{low}}(t) < Y < Y^*_{\text{up}}(t)$, $P(Y, t)$ is governed by eq. (3.6). In the two separate withdrawal regions: $Y \leq Y^*_{\text{low}}(t)$ and $Y \geq Y^*_{\text{up}}(t)$, $P(Y, t)$ assumes the same value as that of the obstacle function.

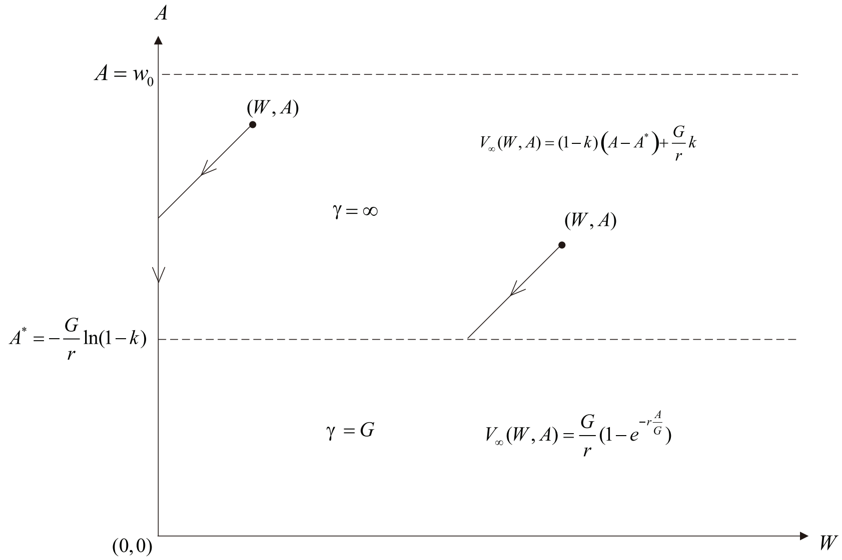

Figure 3. We illustrate the separation of the solution domain $\{(W, A)\colon W \geq 0 \text{ and } 0 \leq A \leq w_0\}$ of the pricing model into "$\gamma = \infty$" region and "$\gamma = G$" region under perpetuality. The separating boundary is the horizontal line: $A = A^*$, where $A^* = -\frac{G}{r}\ln(1-k)$. When (W, A) falls within "$\gamma = \infty$" region, the finite amount $A - A^*$ is withdrawn immediately, so A drops to A^* immediately (see the two arrows shown in the region where $\gamma = \infty$).

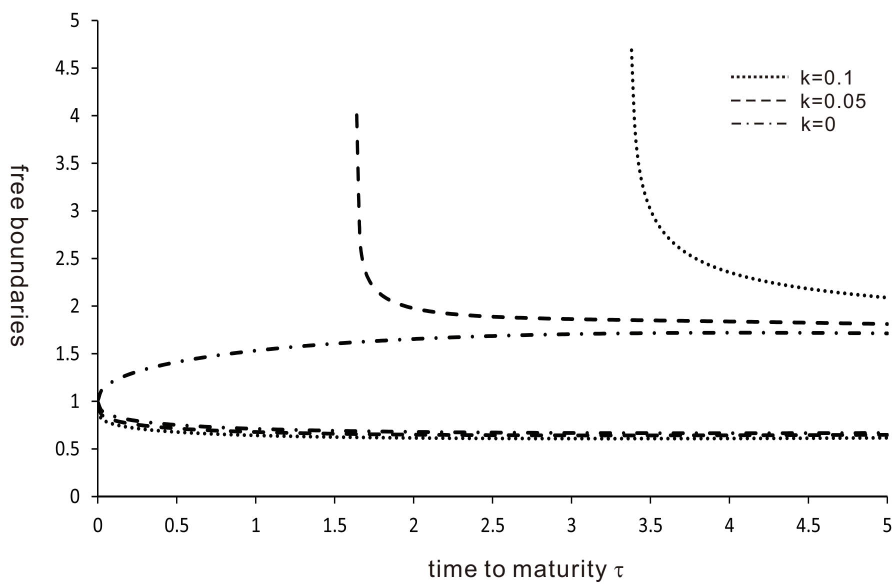

Figure 4. Numerical plots of the withdrawal boundaries $Y^*_{\text{up}}(\tau)$ and $Y^*_{\text{low}}(\tau)$ against time to maturity τ under $G = 0$ with varying values of k. When $k > 0$, $Y^*_{\text{up}}(\tau)$ is not defined for $\tau \leq \tau^*$, where $\tau^* = -\frac{\ln(1-k)}{\eta}$. The threshold value τ^* for $k = 0.1$ and $k = 0.05$ are 3.3677 and 1.6395, respectively. When $k = 0$, $Y^*_{\text{low}}(0) = 1 - k$ and $Y^*_{\text{up}}(\tau)$ is defined for all values of τ. We also observe that $Y^*_{\text{low}}(\tau)$ is not sensitive to change in value of k.

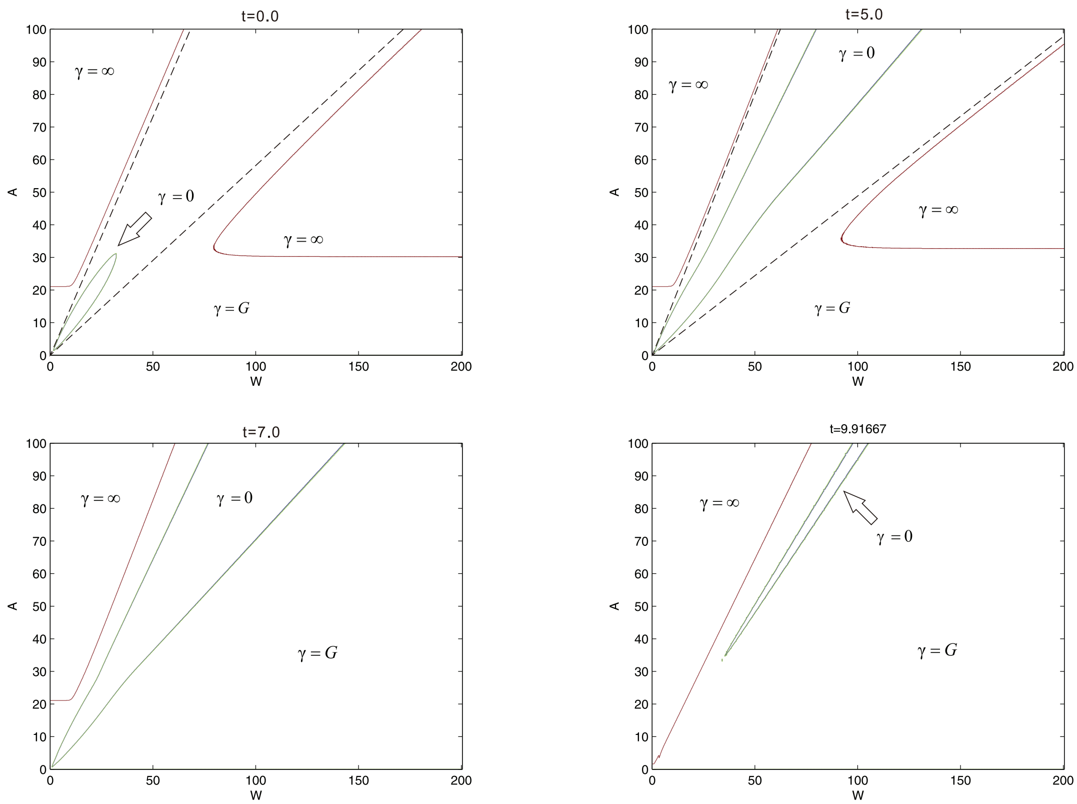

Figure 5: The numerical plots of the optimal withdrawal regions with penalty parameter $k = 0.1$ at varying values of the calendar time t. The horizontal asymptote: $A = A^{**}$ exists (shown in the plots for $t = 0$ and $t = 5.0$) when the calendar time is sufficiently far from expiry. In this set of sample calculations, we obtain $A^{**} = 30.2118$ and non-existence of the horizontal asymptote in the numerical plots occurs when time to expiry is shorter than 3.3677 (consistent with the absence of the horizontal asymptote for $t = 7.0$ and $t = 9.91667$). The dashed lines shown in the plots for $t = 0$ and $t = 5.0$ are the oblique asymptotes for the two "$\gamma = \infty$" regions at large value of A.